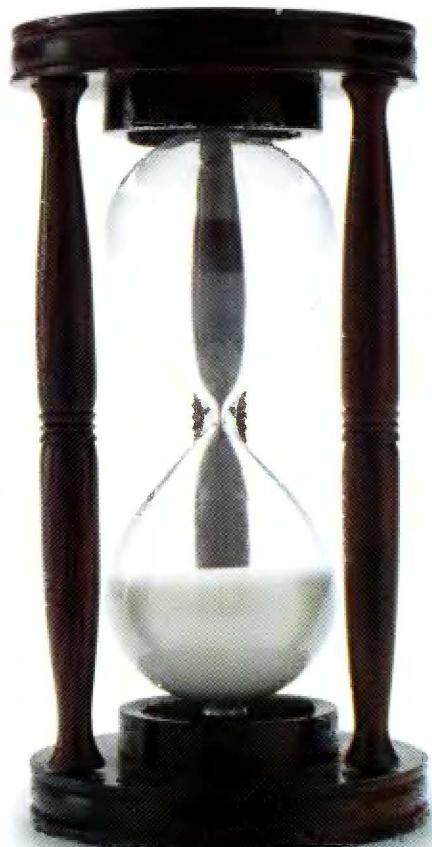


# MATHÉMATIQUES CONCRÈTES

Fondations pour l'informatique

2<sup>e</sup> édition

Ronald L. Graham, Donald E. Knuth et Oren Patashnik  
Traduction de Alain Denise



# **Mathématiques concrètes**

## **Fondations pour l'informatique**

2<sup>e</sup> édition

*En hommage à Leonhard Euler (1707-1783)*

Ronald L. GRAHAM, Donald E. KNUTH, Oren PATASHNIK

# Mathématiques concrètes

## Fondations pour l'informatique

2<sup>e</sup> édition

*Traduction de Alain Denise*



INTERNATIONAL THOMSON PUBLISHING FRANCE  
An International Thomson Publishing Company

E-mail : *contact@itp.fr*  
Listes de diffusion : *listserver@itp.fr*  
World-Wide Web : *http://www.itp.fr*

---

Paris • Albany • Belmont • Bonn • Boston • Cincinnati • Detroit • Johannesburg  
Londres • Madrid • Melbourne • Mexico • New York • Singapour • Toronto • Tokyo

# Préface

*“Audience, level,  
and treatment—a  
description of such  
matters is what pre-  
faces are supposed  
to be about.”*

—P. R. Halmos [173]

Ce livre est basé sur un cours donné à l’Université de Stanford tous les ans depuis 1970. Une cinquantaine d’étudiants le suivent chaque année— principalement des étudiants de troisième cycle, mais aussi d’autres plus jeunes— et certains d’entre eux, devenus enseignants à leur tour, ont commencé à propager cet enseignement en d’autres lieux. C’est pourquoi nous pensons que le moment est venu de le faire connaître à un plus large public.

Les Mathématiques Concètes ont vu le jour en des temps sombres et tumultueux. Durant cette époque troublée, bien des valeurs établies étaient remises en cause, particulièrement dans les campus universitaires, qui étaient des foyers de contestation. On s’attaquait même aux programmes universitaires, et les mathématiques n’échappaient pas à la règle. John Hammersley venait juste d’écrire un article provocateur, intitulé “De l’affaiblissement des capacités intellectuelles en mathématiques par l’usage des “Mathématiques Modernes” et autres sottises dans les écoles et les universités” [176]. D’autres mathématiciens [332] se demandaient même si “les mathématiques pouvaient être sauvées”. L’un des auteurs du présent ouvrage (DEK) avait commencé à écrire une série de trois livres, *The Art of Computer Programming*. En rédigeant le premier, il s’était aperçu qu’il ne disposait pas de tous les outils nécessaires ; pour vraiment comprendre, de façon approfondie, les programmes informatiques, il fallait des mathématiques tout à fait différentes de celles qu’il avait apprises à l’université. C’est pourquoi il décida de concevoir un nouveau cours dans lequel il enseignerait ce qu’il aurait voulu qu’on lui apprenne.

Dans l’esprit de son auteur, ce cours, intitulé “Mathématiques Concètes”, devait d’abord être un antidote aux “Mathématiques Abstraites”. Nous étions en effet sous le joug de ce qu’on appelait les “Maths Modernes”, une déferlante d’idées abstraites au nom desquelles tous les résultats classiques “concrets” avaient été éjectés des programmes universitaires. Les mathématiques abstraites constituent un sujet merveilleux et donnent lieu à des développements passionnantes et fort utiles ; mais leurs tenants étaient persuadés que les autres mathématiques leur étaient inférieures, qu’elles ne méritaient pas qu’on s’y intéresse. Ainsi, toute une génération de mathématiciens était obnubilée par un unique but : généraliser. Ils étaient incapables

*“People do acquire a little brief authority by equipping themselves with jargon: they can pontificate and air a superficial expertise. But what we should ask of educated mathematicians is not what they can speechify about, nor even what they know about the existing corpus of mathematical knowledge, but rather what can they now do with their learning and whether they can actually solve mathematical problems arising in practice. In short, we look for deeds not words.”*

—J. Hammersley [176]

de voir la beauté du détail, de ressentir le plaisir de résoudre des problèmes quantitatifs, de goûter la valeur d'une technique de calcul. Les mathématiques abstraites fonctionnaient en cercle fermé ; elles perdaient la réalité de vue. Pour guérir, l'enseignement des mathématiques avait besoin d'un contrepoids.

DEK commença son tout premier cours à l'Université de Stanford en expliquant pourquoi il avait choisi ce titre plutôt inhabituel. Il annonça que, contrairement aux attentes de certains de ses collègues, il n'allait pas enseigner la théorie des agrégats, ni le théorème de plongement de Stone, ni même la compactification de Stone–Čech. (A ce moment précis, un certain nombre d'étudiants du département de génie civil se levèrent et quittèrent calmement la salle).

Ce cours n'avait pas été créé uniquement en réaction contre certaines tendances ; il l'avait été aussi et surtout pour des raisons positives. Il fit d'ailleurs, alors que son contenu se "solidifiait" au fil des années, la preuve de son utilité dans bon nombre d'applications nouvelles. D'autre part, la validité du choix du nom se trouva confirmée lorsque Z. A. Melzak publia deux volumes intitulés *Companion to Concrete Mathematics* [267].

A première vue, les mathématiques concrètes peuvent sembler n'être qu'un ramassis disparate de "trucs" ; on réalise vite que c'est en fait un ensemble cohérent de techniques. De plus, il s'avère qu'elles séduisent beaucoup de ceux qui les ont pratiquées. Lorsque, en 1979, un autre des auteurs de ce livre (RLG) enseigna pour la première fois ce cours, les étudiants étaient si enthousiastes qu'ils décidèrent de se retrouver l'année suivante.

Au fait, que signifie exactement "mathématiques concrètes" ? C'est un mélange de mathématiques continues et de mathématiques DISCRÈTES. Plus concrètement, c'est l'utilisation d'un ensemble de techniques de manipulation de formules mathématiques en vue de résoudre des problèmes. Une fois que vous aurez appris ce qu'il y a dans ce livre, vous saurez calculer des sommes d'apparence très compliquée, résoudre des relations de récurrence complexes, découvrir des relations subtiles dans les données d'un problème ; vous serez tellement à l'aise en calcul algébrique que vous préferez chercher (et trouver) des résultats exacts plutôt que de vous contenter d'approximations. Pour cela, il vous suffira d'avoir une tête en état de marche, une grande feuille de papier et une écriture lisible.

Parmi les principaux thèmes traités dans ce livre, vous trouverez la manipulation de sommes, de récurrences, les bases de la théorie des nombres, les coefficients binomiaux, les fonctions génératrices, les probabilités discrètes et le calcul asymptotique. Nous insisterons plus sur les techniques de manipulation que sur les théorèmes d'existence ou le raisonnement combinatoire. Notre but est que chaque lecteur devienne aussi familier avec les

*Le titre original du livre, "Concrete Mathematics", peut se traduire en "Mathématiques concrètes", mais aussi en "Mathématiques en béton". Le "théorème de plongement de Stone" s'écrit, en anglais, "Stone's Embedding Theorem", ce qu'on peut traduire littéralement par "théorème de scellement de pierres" (N.D.T.).*

*"It is downright sinful to teach the abstract before the concrete."*

— Z. A. Melzak [267]

*"The heart of mathematics consists of concrete examples and concrete problems."*

— P. R. Halmos [172]

*Les mathématiques concrètes sont un pont vers les mathématiques abstraites.*

*"The advanced reader who skips parts that appear too elementary may miss more than the less advanced reader who skips parts that appear too complex."*

— G. Pólya [297]

(Nous n'avons  
pas osé prendre  
"Mathématiques  
Distinues").

"... a concrete  
life preserver thrown  
to students sinking  
in a sea of abstraction."

— W. Gottschalk

opérations discrètes (comme la fonction partie entière ou la sommation finie) qu'un étudiant en mathématiques classiques peut être familier avec les opérations continues (comme la fonction valeur absolue ou l'intégration).

Notez que cette liste de thèmes est tout à fait différente de ce qu'on peut trouver dans ce qu'on appelle maintenant les cours de "Mathématiques Discrètes". C'est pourquoi le titre de notre cours, "Mathématiques Concrètes", est différent aussi.

A l'origine, le texte de référence pour le cours de mathématiques concrètes de Stanford était la section "Préliminaires mathématiques" du livre *The Art of Computer Programming* [207]. L'un des auteurs (OP) eut l'idée de rédiger des notes additionnelles à cette présentation plutôt concise du cours (elle ne faisait que 110 pages). Ces notes ont donné naissance au présent ouvrage, qui est à la fois un prolongement et une introduction en douceur aux notions abordées dans les "Préliminaires mathématiques". Quelques notions, parmi les plus difficiles, ont été supprimées, et beaucoup d'autres ont été ajoutées.

Nous avons eu énormément de plaisir à rédiger ce livre ensemble. En fait, nous avons presque eu l'impression qu'il se faisait tout seul, tant son contenu semblait prendre vie devant nos yeux. Nous avions tous trois une conception un peu inhabituelle de la façon de faire des mathématiques, forgée par des années d'expérience ; nous nous sommes tellement bien entendus que nous ne pouvons pas nous empêcher de voir cet ouvrage comme une sorte de manifeste, dans lequel nous présentons les mathématiques telles que nous aimons les faire. Nous avons essayé de montrer dans ce livre la beauté et de la magie des mathématiques, et nous espérons que nos lecteurs ressentiront une part du plaisir que nous avons eu à l'écrire.

Considérant que ce cours a vu le jour dans un milieu universitaire, nous avons choisi d'adopter un style direct pour être plus proches de l'ambiance d'une salle de classe. Certains personnes croient que les mathématiques doivent être sérieuses, froides, rigides ; nous pensons au contraire qu'on peut s'amuser en faisant des mathématiques, et nous n'en avons pas honte. Pourquoi faudrait-il tracer une frontière entre le travail et le plaisir ? Il y a énormément de choses attrayantes dans les mathématiques concrètes ; les calculs ne sont pas toujours faciles, mais les résultats peuvent être surprenants et amusants. Vous trouverez dans ce livre les joies et les peines des mathématiques, tout simplement parce qu'elles font partie de la vie.

Il est bien connu que les étudiants en savent toujours plus que leurs enseignants. C'est pourquoi nous avons demandé aux premiers utilisateurs de ce cours de donner leur avis sous forme de "graffitis" dans les marges. Parmi ces notes, certaines sont faciles, d'autres sont profondes ; certaines permettent de lever des ambiguïtés ou d'éclaircir des notions obscures, d'autres reflètent la proverbiale sagesse des habitués du dernier rang ; certaines sont po-

Graffiti de maths :  
Libérez le groupe.  
 $N=1 \Rightarrow P=N.P$ .

sitives, d'autres sont négatives, d'autres encore sont nulles. Toutes peuvent véritablement aider à assimiler les notions présentées dans le livre. L'idée de ces notes de marge provient d'un fascicule distribué par l'université de Stanford aux nouveaux arrivants. La présentation officielle de l'université y est commentée par les étudiants. Par exemple, lorsqu'on lit "l'entité protéiforme que constitue l'université de Stanford", on trouve la note en marge suivante : "Protéiforme ? Qu'est-ce que c'est que ce truc ? Ça ne m'étonne pas, c'est typique du pseudo-intellectualisme qui règne ici" ; en face de "Il n'y a pas de limites au potentiel d'un groupe d'étudiants qui vivent ensemble", on peut lire "Les chambrées de Stanford sont des zoos sans gardien".

Il nous a semblé intéressant de présenter, en parallèle à certains développements, des citations d'illustres mathématiciens qui nous ont précédés. Les mathématiques sont un perpétuel chantier, sur lequel chaque mathématicien, d'hier ou d'aujourd'hui, apporte sa pierre. Ainsi, on trouvera en notes de marge les mots exacts avec lesquels Leibniz, Euler, Gauss et d'autres encore ont présenté quelques-uns de leurs résultats fondamentaux.

Cet ouvrage comporte plus de 500 exercices classés en six catégories :

- **Les échauffements** sont les exercices que CHAQUE LECTEUR devrait essayer de faire après avoir lu les notions correspondantes.
- **Les exercices de base** servent à aborder des notions qu'on comprend mieux en les découvrant par soi-même qu'en les lisant.
- **Les devoirs à la maison** sont destinés à approfondir la compréhension de certaines notions abordées dans le chapitre correspondant.
- **Les problèmes d'examen** font généralement appel à des notions provenant de plusieurs chapitres. Ils sont faits pour être résolus à la maison plutôt que lors d'un examen en temps limité.
- **Les questions subsidiaires** vont généralement au-delà de ce qu'on peut demander à un étudiant qui suit un cours basé sur ce livre ; elles constituent une manière intéressante de prolonger le cours.
- **Les problèmes de recherche** peuvent-ils humainement être résolus ? Nous l'ignorons, mais ceux-ci méritent qu'on y réfléchisse (sans être pressé par le temps).

On trouvera dans l'annexe A les solutions de tous les exercices, avec parfois des informations complémentaires (bien entendu, les "solutions" des problèmes de recherche sont incomplètes, mais on y présente des résultats partiels ou des indices qui pourraient s'avérer utiles). Il est conseillé de lire ces solutions, particulièrement celles des exercices d'échauffement, mais seulement APRÈS avoir sérieusement essayé de résoudre les problèmes.

Concevoir un exercice intéressant est un travail qui nécessite souvent une grande créativité, et parfois aussi un peu de chance. C'est pourquoi

*Tout cela ne m'intéresse que marginalement.*

*Ce cours est le plus intéressant que j'ai jamais suivi. Cependant, il serait bon de faire une pause de temps en temps pour faire un bilan des choses apprises.*

*Les devoirs à la maison étaient durs, mais ils m'ont permis de beaucoup apprendre. Cela vaut la peine d'y passer du temps.*

*Ces examens à la maison sont vitaux. Gardez les.*

*J'ai été surpris par la difficulté des examens par rapport à celle des devoirs à la maison.*

*On peut toujours tricher en copiant les réponses, mais ce serait tricher avec soi-même.*

nous avons essayé, dans l'annexe C, de citer la source de chacun des exercices de ce livre. Il est malheureusement d'usage, chez les mathématiciens, d'emprunter les exercices des autres sans songer à les remercier. Nous pensons que la tradition contraire, pratiquée par exemple dans les livres et les magazines d'échecs (les noms, dates et lieux d'origine des problèmes y sont régulièrement spécifiés), est bien meilleure. Hélas, nous avons été incapables de retrouver l'origine de tous les exercices de ce livre, loin s'en faut ; beaucoup font maintenant partie du folklore. Nous serons reconnaissants à tout lecteur qui pourra nous apprendre l'origine d'un exercice pour lequel il n'y a pas de citation, ou pour lequel celle-ci est inexacte. Nous en tiendrons compte dans les prochaines éditions de ce livre.

*Les examens sont trop difficiles pour les étudiants qui ont d'autres matières à réviser.*

*Je ne vois pas comment ce que j'ai appris pourrait m'être utile.*

*J'ai eu beaucoup de difficultés à assimiler ce cours, mais il m'a permis d'améliorer ma pratique des maths et ma façon de réfléchir.*

*Cher prof, merci pour (1) les jeux de mots, (2) les mathématiques.*

La police de caractères utilisée pour les formules mathématiques est due à Hermann Zapf [227] ; elle a été commandée par l'American Mathematical Society et développée avec l'aide d'un comité composé de B. Beeton, R. P. Boas, L. K. Durst, D. E. Knuth, P. Murdock, R. S. Palais, P. Renz, E. Swanson, S. B. Whidden et W. B. Woolf. Cette police a été conçue dans le but de représenter les mathématiques comme elles pourraient l'être par un mathématicien doué d'une très belle écriture. Dans ce domaine, l'écriture manuelle est préférable à l'écriture mécanique car c'est en général avec un crayon, un stylo ou une craie que les mathématiques se construisent (par exemple, dans cette police, le sommet du symbole "0" est légèrement pointu car c'est ce qui arrive naturellement lorsqu'on "ferme" le zéro qu'on trace à la main). Les lettres sont droites au lieu d'être en italiques, de sorte que les indices, les exposants et les accents s'accordent mieux avec les autres symboles. Cette nouvelle police s'appelle *AMS Euler*, en hommage au grand mathématicien suisse Leonhard Euler (1707–1783), qui a découvert une grande part des mathématiques que nous connaissons aujourd'hui. Elle comporte plusieurs alphabets : Euler Text (Aa Bb Cc ... Xx Yy Zz), Euler Fraktur (Aa Bb Cc ... Xx Yy Zz), Euler Script Capitals (A B C ... X Y Z), Euler Greek (Aα Bβ Γγ ... Xχ Ψψ Ωω) et d'autres symboles comme ϕ et Σ. Nous sommes particulièrement heureux de l'étrenner dans ce livre, car l'esprit d'Euler y est présent partout : les mathématiques concrètes sont des mathématiques eulériennes.

Nous exprimons notre profonde gratitude à Andrei Broder, Ernst Mayr, Andrew Yao et Frances Yao qui ont grandement contribué à ce livre en enseignant les Mathématiques Concrites à Stanford. Aux assistants qui ont, chaque année, rédigé des notes de cours et participé à la conception des examens, nous adressons 1024 mercis ; leurs noms sont cités dans l'annexe C. Ce livre, qui est en quelque sorte un abrégé de seize années de notes de cours, n'aurait pas pu exister sans leur excellent travail.

Beaucoup d'autres personnes ont participé à la réalisation de cet ouvrage. Les étudiants de Brown, de Columbia, de CUNY, de Princeton, de

x PRÉFACE

Rice et de Stanford y ont ajouté leurs graffitis et ont nous aidés à corriger les premiers brouillons. Les contacts avec notre éditeur, Addison-Wesley, ont été très fructueux ; nous remercions particulièrement Peter Gordon, Bette Aaronson, Roy Brown et Lyn Dupré. La National Science Foundation et l'Office of Naval Research nous ont offert un soutien sans faille. Cheryl Graham nous a apporté une aide formidablement efficace dans la préparation de l'index. Par dessus tout, nous remercions nos épouses (Fan, Jill et Amy) pour leur patience, leur soutien, leurs encouragements et leurs idées.

*Murray Hill, New Jersey  
et Stanford, Californie  
Mai 1988 et Octobre 1993*

*Conseil aux étudiants peu sérieux :  
ne vous aventurez pas dans ce cours.*

— RLG  
DEK  
OP

# Notes sur les notations

Certaines notations utilisées dans ce livre ne sont pas (pas encore ?) standard. Elles sont présentées dans la liste qui suit. Chacune d'entre elles est suivie du numéro de la page dans laquelle elle est définie. On trouvera dans l'index général, en fin d'ouvrage, des références à des notations plus standard.

<i>Notation</i>	<i>Nom</i>	<i>Page</i>
$\ln x$	<i>logarithme népérien</i> : $\log_e x$	293
$\lg x$	<i>logarithme en base 2</i> : $\log_2 x$	75
$\log x$	<i>logarithme vulgaire</i> : $\log_{10} x$	476
$\lfloor x \rfloor$	<i>partie entière inférieure</i> : $\max\{n \mid n \leq x, n \text{ entier}\}$	72
$\lceil x \rceil$	<i>partie entière supérieure</i> : $\min\{n \mid n \geq x, n \text{ entier}\}$	72
$x \bmod y$	<i>reste de la division entière</i> : $x - y\lfloor x/y \rfloor$	88
$\{x\}$	<i>partie fractionnaire</i> : $x \bmod 1$	75
$\sum f(x) \delta x$	<i>sommation indéfinie</i>	52
$\sum_a^b f(x) \delta x$	<i>sommation définie</i>	53
$x^n$	<i>puissance factorielle descendante</i> : $x!/(x-n)!$	51, 225
$x^{\overline{n}}$	<i>puissance factorielle montante</i> : $\Gamma(x+n)/\Gamma(x)$	51, 225
$n_j$	<i>sous-factorielle</i> : $n!/0! - n!/1! + \dots + (-1)^n n!/n!$	207
$\Re z$	<i>partie réelle</i> : $x$ , si $z = x + iy$	68
$\Im z$	<i>partie imaginaire</i> : $y$ , si $z = x + iy$	68
$H_n$	<i>nombre harmonique</i> : $1/1 + \dots + 1/n$	31
$H_n^{(x)}$	<i>nombre harmonique généralisé</i> : $1/1^x + \dots + 1/n^x$	294
$f^{(m)}(z)$	<i>mième dérivée de f en z</i>	497

xii NOTES SUR LES NOTATIONS

$\begin{bmatrix} n \\ m \end{bmatrix}$	nombre de Stirling de première espèce	275
$\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$	nombre de Stirling de deuxième espèce	274
$\left\langle \begin{matrix} n \\ m \end{matrix} \right\rangle$	nombre eulérien	283
$\left\langle\!\! \left\langle \begin{matrix} n \\ m \end{matrix} \right\rangle\!\! \right\rangle$	nombre eulérien du second ordre	286
$(a_m \dots a_0)_b$	notation en base $b$ de $\sum_{k=0}^m a_k b^k$	12
$K(a_1, \dots, a_n)$	continuant	321
$F\left(\begin{matrix} a, b \\ c \end{matrix} \middle  z\right)$	fonction hypergéométrique	218
$\#A$	cardinal : nombre d'éléments de l'ensemble $A$	42
$[z^n] f(z)$	coefficient de $z^n$ dans $f(z)$	211
$[\alpha \dots \beta]$	intervalle fermé : l'ensemble $\{x \mid \alpha \leq x \leq \beta\}$	79
$[m=n]$	1 si $m = n$ , sinon 0 *	26
$[m \backslash n]$	1 si $m$ divise $n$ , sinon 0 *	110
$[m \backslash\backslash n]$	1 si $m$ divise exactement $n$ , sinon 0 *	157
$[m \perp n]$	1 si $m$ est premier par rapport à $n$ , sinon 0 *	125

\*Plus généralement, si  $S$  est une proposition qui peut être vraie ou fausse,  $[S]$  est égal à 1 si  $S$  est vraie, à 0 sinon.

Dans ce livre, nous considérerons qu'une expression de la forme " $a/bc$ " est égale à " $a/(bc)$ ". D'autre part,  $\log x/\log y = (\log x)/(\log y)$  et  $2n! = 2(n!)$ .

# Table des matières

---

<b>1</b>	<b>Problèmes récurrents</b>	<b>1</b>
1.1	La tour de Hanoi	1
1.2	Droites dans le plan	5
1.3	Le problème de Josèphe	9
	Exercices	18
<b>2</b>	<b>Sommes</b>	<b>23</b>
2.1	Notations	23
2.2	Sommes et récurrences	27
2.3	Manipulation de sommes	32
2.4	Sommes multiples	36
2.5	Méthodes générales	45
2.6	Calcul fini et infini	50
2.7	Sommes infinies	60
	Exercices	66
<b>3</b>	<b>Fonctions entières</b>	<b>72</b>
3.1	Parties entières	72
3.2	Applications	75
3.3	Récurrences et parties entières	85
3.4	“mod” : l’opération binaire	88
3.5	Sommes de parties entières	92
	Exercices	102
<b>4</b>	<b>Théorie des nombres</b>	<b>110</b>
4.1	Divisibilité	110
4.2	Nombres premiers	114
4.3	Premiers exemples premiers	116
4.4	Facteurs factoriels	120
4.5	Primalité relative	124
4.6	“mod” : la congruence	133
4.7	Résidus indépendants	136
4.8	Autres applications	139
4.9	Phi et Mu	143
	Exercices	154
<b>5</b>	<b>Coefficients binomiaux</b>	<b>164</b>
5.1	Identités de base	164
5.2	Pratique de base	185
5.3	Trucs et astuces	199

xiv TABLE DES MATIÈRES

5.4 Fonctions génératrices	210
5.5 Fonctions hypergéométriques	218
5.6 Identités hypergéométriques	231
5.7 Sommes hypergéométriques finies	238
5.8 Sommation automatique	245
Exercices	258
<b>6 Nombres remarquables</b>	<b>273</b>
6.1 Nombres de Stirling	273
6.2 Nombres eulériens	283
6.3 Nombres harmoniques	289
6.4 Sommation harmonique	296
6.5 Nombres de Bernoulli	300
6.6 Nombres de Fibonacci	309
6.7 Continuants	320
Exercices	328
<b>7 Fonctions génératrices</b>	<b>340</b>
7.1 Théorie des dominos et monnaie	340
7.2 Manœuvres de base	351
7.3 Résolution de récurrences	357
7.4 Fonctions génératrices remarquables	371
7.5 Convolutions	374
7.6 Fonctions génératrices exponentielles	386
7.7 Fonctions génératrices de Dirichlet	392
Exercices	394
<b>8 Probabilités discrètes</b>	<b>405</b>
8.1 Définitions	405
8.2 Moyenne et variance	411
8.3 Fonctions génératrices de probabilité	419
8.4 Pile ou face	426
8.5 Hachage	435
Exercices	452
<b>9 Calcul asymptotique</b>	<b>465</b>
9.1 Une hiérarchie de fonctions	466
9.2 La notation $\mathcal{O}$	469
9.3 Manipulation de $\mathcal{O}$	476
9.4 Deux trucs asymptotiques	491
9.5 La formule de sommation d'Euler	497
9.6 Dernières Sommations	504
Exercices	518
<b>A Solutions des exercices</b>	<b>526</b>
<b>B Bibliographie</b>	<b>639</b>
<b>C Références pour les exercices</b>	<b>668</b>
<b>Index</b>	<b>673</b>
<b>Liste des Tables</b>	<b>688</b>

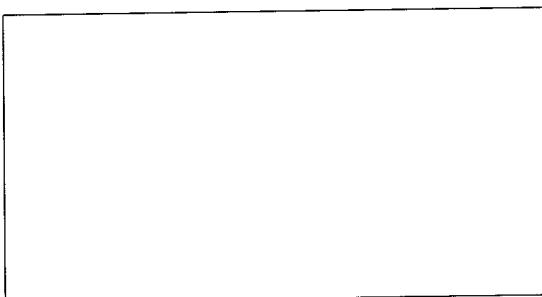
# 1

## Problèmes récurrents

DANS CE CHAPITRE, nous considérons trois problèmes particuliers qui vont nous donner un aperçu de ce qui suivra. Ces problèmes ont deux points communs : tous trois ont été maintes fois étudiés par les mathématiciens, et toutes leurs solutions utilisent l'idée de *récurrence*, selon laquelle la solution de chaque problème dépend des solutions d'instances plus petites du même problème.

### 1.1 LA TOUR DE HANOI

Commençons par étudier la Tour de Hanoi, un joli petit casse-tête inventé par le mathématicien français Edouard Lucas en 1883. Considérons une tour de huit disques, initialement empilés par ordre de tailles décroissantes sur un axe vertical, deux autres axes identiques restant libres :



Le but du jeu est de déplacer toute la tour sur l'un des axes libres en respectant les règles suivantes : il ne faut déplacer qu'un disque à la fois, et ne jamais poser un disque sur un disque plus petit.

Lucas [260] présentait son jeu en l'agrémentant d'une légende romanesque selon laquelle il existerait une tour bien plus grande, la Tour de Brahma, composée de 64 disques d'or pur reposant sur trois aiguilles de diamant. Au commencement des temps, disait-il, Dieu plaça ces disques

*Que tous ceux qui n'ont jamais vu ça lèvent le doigt.  
Les autres peuvent aller directement à l'équation (1.1).*

*Des disques d'or !  
Les nôtres sont plus concrets.*

## 2 PROBLÈMES RÉCURRENTS

d'or sur la première aiguille et ordonna qu'un groupe de prêtres le transférât sur la troisième selon les règles indiquées ci-dessus. Les prêtres, paraît-il, se consacrent nuit et jour à leur tâche. Quand ils auront terminé, la tour s'écroulera et notre monde arrivera à sa fin.

Il n'est pas évident à première vue qu'il existe une solution à ce casse-tête. Cependant, avec un peu de réflexion (ou si l'on a déjà rencontré le problème), on se convainc qu'il y en a une. On se pose alors la question suivante : quelle est la meilleure façon d'y arriver ? Pour parler plus précisément, combien de déplacements de disques sont nécessaires et suffisants pour arriver au but ?

La meilleure façon de s'attaquer à un problème de ce genre est de le généraliser un peu. La Tour de Brahma est composée de 64 disques, la Tour de Hanoi en a 8. Voyons ce qui se passe quand il y a  $n$  disques.

Un avantage de cette généralisation, c'est qu'on peut réduire encore la taille du problème. Nous verrons souvent dans ce livre qu'il vaut mieux d'abord CONSIDÉRER DES CAS DE PETITE TAILLE. On voit sans difficulté comment transférer une tour qui contient seulement un ou deux disques ; et au bout de quelques tâtonnements, on arrive à déplacer une tour de trois disques.

L'étape suivante dans la résolution du problème consiste à introduire une notation appropriée, suivant le précepte : NOMMER POUR RÉGNER. Appelons  $T_n$  le nombre minimum de déplacements nécessaires pour transférer  $n$  disques d'un axe à une autre en respectant les règles de Lucas.  $T_1$  est évidemment égal à 1, et  $T_2 = 3$ .

On peut aussi avoir une autre information pour le même prix, en considérant le plus petit cas possible :  $T_0$  est clairement égal à 0, puisque aucun mouvement n'est nécessaire pour déplacer une tour de  $n = 0$  disque ! Les mathématiciens malins n'ont pas honte de penser petit, parce qu'ils savent qu'on saisit plus facilement les problèmes généraux lorsqu'on a bien compris les cas extrêmes (même s'ils sont triviaux).

Maintenant, changeons de point de vue et essayons de penser grand. Comment déplacer une grande tour ? Lorsqu'il y a trois disques, la stratégie gagnante consiste à déplacer les deux disques du haut vers l'axe du milieu, puis déplacer le troisième disque, enfin empiler les deux autres dessus. Ceci nous donne un indice pour déplacer  $n$  disques : nous allons d'abord transférer les  $n - 1$  plus petits disques vers l'axe du milieu, ce qui nécessitera  $T_{n-1}$  déplacements ; puis transférer le plus grand disque ; enfin, ré-empiler les  $n - 1$  plus petits disques sur le grand, en faisant de nouveau  $T_{n-1}$  déplacements. Ainsi nous pouvons transférer  $n$  disques ( $n > 0$ ) en effectuant au plus  $2T_{n-1} + 1$  déplacements :

$$T_n \leq 2T_{n-1} + 1, \quad \text{pour } n > 0.$$

Dans cette formule, nous avons écrit “ $\leq$ ” et non “ $=$ ” parce que notre construction prouve seulement que  $2T_{n-1} + 1$  déplacements suffisent ; nous n'avons pas montré qu'ils sont nécessaires. Peut-être quelqu'un d'astucieux pourrait-il faire mieux.

*La plupart des “solutions” publiées de ce problème, comme celle d'Allardice et Fraser [7], ne peuvent expliquer pourquoi  $T_n$  doit être  $\geq 2T_{n-1} + 1$ .*

Existe-t-il en fait une meilleure méthode ? En réalité non. Il arrive un moment où on doit déplacer le plus grand disque. Pour ce faire, il faut que les  $n - 1$  autres soient sur un seul axe, et il a fallu au moins  $T_{n-1}$  déplacements pour les y mettre. Si on n'est pas trop vigilant, il est possible que l'on déplace le plus grand disque plus d'une fois. Mais après son dernier déplacement, on doit derechef poser les  $n - 1$  plus petits (qui sont nécessairement sur un seul axe) sur le plus grand. Ceci nécessite encore  $T_{n-1}$  déplacements. En conséquence,

$$T_n \geq 2T_{n-1} + 1, \quad \text{pour } n > 0.$$

De ces deux inégalités, et de la solution triviale pour  $n = 0$ , on déduit que

$$\begin{aligned} T_0 &= 0; \\ T_n &= 2T_{n-1} + 1, \quad \text{pour } n > 0. \end{aligned} \tag{1.1}$$

(Remarquez que ces formules sont cohérentes avec les valeurs déjà connues  $T_1 = 1$  et  $T_2 = 3$ . Nos expériences sur des petits cas ne nous ont pas seulement aidés à trouver une formule générale ; elles nous ont aussi fourni un moyen pratique de vérifier que nous n'avons pas commis une erreur idiote. De telles vérifications seront particulièrement précieuses lorsque nous effectuerons des manœuvres plus compliquées dans les chapitres suivants).

*Ouais ouais...  
j'ai déjà vu ce mot  
quelque part.*

Un ensemble d'égalités comme (1.1) est appelé une *récurrence* (ou une relation de récurrence). Elle se compose d'une valeur initiale et d'une équation générale permettant de calculer une valeur en fonction des précédentes. Parfois l'équation générale seule est appelée récurrence, bien que techniquement elle doive être accompagnée d'une valeur initiale pour être complète.

La récurrence nous permet de calculer  $T_n$  pour n'importe quelle valeur de  $n$ . Mais en vérité personne n'a envie d'utiliser une récurrence pour un ce genre de calcul : ça dure trop longtemps si  $n$  est grand. La récurrence nous donne seulement une information indirecte, ou “locale”. Ce qui nous plairait bien davantage, ce serait une *solution de la récurrence*, c'est-à-dire une gentille “formule close” pour le nombre  $T_n$ , qui nous permettrait de le calculer rapidement même si  $n$  est grand.

Mais comment résoudre une récurrence ? Une méthode possible consiste à deviner la solution, puis prouver que notre prévision est correcte. Et pour deviner la solution, notre meilleur espoir réside (encore) dans les petits cas. Calculons donc successivement  $T_3 = 2 \cdot 3 + 1 = 7$  ;  $T_4 = 2 \cdot 7 + 1 = 15$  ;

## 4 PROBLÈMES RÉCURRENTS

$T_5 = 2 \cdot 15 + 1 = 31$  ;  $T_6 = 2 \cdot 31 + 1 = 63$ . Ah ah ! On dirait bien que

$$T_n = 2^n - 1, \quad \text{pour } n \geq 0. \quad (1.2)$$

En tout cas c'est vrai pour  $n \leq 6$ .

*L'induction mathématique* est une méthode générale permettant de prouver qu'une assertion concernant un entier  $n$  est vraie pour tout  $n \geq n_0$ . D'abord on prouve qu'elle est vraie lorsque  $n$  prend sa plus petite valeur,  $n_0$  ; c'est ce qu'on appelle la *base*. Puis on démontre qu'elle est vraie pour  $n > n_0$ , en supposant qu'elle a déjà été prouvée pour toutes les valeurs comprises entre  $n_0$  et  $n - 1$  ; c'est ce qu'on appelle l'*induction*. Ce genre de preuve permet d'obtenir une infinité de résultats avec une quantité de travail finie.

Les récurrences sont des sujets idéaux pour appliquer l'*induction mathématique*. Dans notre cas, par exemple, (1.2) se déduit aisément de (1.1) : la base est triviale car  $T_0 = 2^0 - 1 = 0$  ; et l'*induction* s'ensuit pour  $n > 0$  si on suppose que (1.2) est vraie lorsqu'on remplace  $n$  par  $n - 1$  :

$$T_n = 2T_{n-1} + 1 = 2(2^{n-1} - 1) + 1 = 2^n - 1.$$

Ainsi (1.2) est vraie aussi pour  $n$ . Bon ! Nos recherches sur  $T_n$  s'achèvent sur un succès.

Bien entendu, les prêtres ne sont pas près d'avoir mené à bien leur tâche. Ils en ont encore pour un bon moment car, pour  $n = 64$ , il faut déplacer  $2^{64} - 1$  disques (soit à peu près 18 trillions). Même en faisant un mouvement par microseconde, ce qui est bien sûr impossible, il leur faudrait plus de 5000 siècles pour déplacer la Tour de Brahma. Le casse-tête original de Lucas est plus raisonnable : il nécessite  $2^8 - 1 = 255$  mouvements, ce qui demande à peu près quatre minutes à quelqu'un de rapide.

La récurrence de la Tour de Hanoi est un exemple typique des problèmes que l'on rencontre dans des applications de toutes sortes. Pour trouver une expression explicite d'une quantité donnée comme  $T_n$ , on procède en trois étapes :

- 1 Etudier les petits cas. Cela nous permet d'avoir un aperçu du problème ; de plus cela nous sera utile dans les étapes suivantes.
- 2 Trouver et prouver une expression mathématique pour la quantité en question. Pour le problème de la Tour de Hanoi, c'est la récurrence (1.1), qui nous donne la possibilité de calculer  $T_n$  pour tout  $n$ .
- 3 Trouver et prouver une formule close pour l'expression mathématique de l'étape qui précède. Pour la Tour de Hanoi, la formule close est la formule (1.2), solution de la récurrence.

Cette troisième étape est celle sur laquelle nous concentrerons particulièrement notre attention tout au long de ce livre. Il nous arrivera même

*L'induction mathématique montre qu'on peut grimper aussi haut qu'on veut sur une échelle, pourvu qu'on sache monter sur le premier barreau (base) et qu'à partir chaque barreau on puisse se hisser jusqu'au suivant (induction).*

souvent de sauter les deux premières parce que notre point de départ sera déjà une expression mathématique. Nous pourrons cependant être amenés à résoudre des sous-problèmes qui, eux, nécessiteront un passage par les trois étapes.

Notre analyse du problème de la Tour de Hanoi nous a conduits à la bonne réponse. Cependant elle a nécessité un “saut inductif” : nous avons d’abord deviné la solution, et notre raisonnement s’est appuyé sur cette heureuse prédiction. Un des principaux objectifs de ce livre est d’expliquer comment on peut résoudre des récurrences sans être extra-lucide. Par exemple, nous verrons que la récurrence (1.1) peut se simplifier en ajoutant 1 des deux côtés des équations :

$$\begin{aligned} T_0 + 1 &= 1; \\ T_n + 1 &= 2T_{n-1} + 2, \quad \text{pour } n > 0. \end{aligned}$$

*Intéressant : on se débarrasse du +1, non pas en soustrayant, mais en additionnant.*

Maintenant, si on pose  $U_n = T_n + 1$ , on obtient

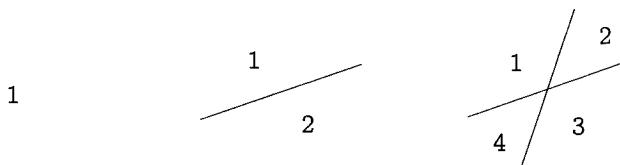
$$\begin{aligned} U_0 &= 1; \\ U_n &= 2U_{n-1}, \quad \text{pour } n > 0. \end{aligned} \tag{1.3}$$

Pas besoin d’être un génie pour découvrir que la solution de *cette* récurrence est tout simplement  $U_n = 2^n$  ; donc  $T_n = 2^n - 1$ . Même un ordinateur pourrait voir ça.

## 1.2 DROITES DANS LE PLAN

Notre deuxième exemple est d’essence plus géométrique : combien peut-on couper de parts de pizza en la coupant  $n$  fois en ligne droite avec un couteau ? Ou bien, pour parler plus formellement : quel est le nombre maximum de régions définies par  $n$  droites dans le plan ? Ce problème a été résolu en 1826 par le mathématicien suisse Jacob Steiner [338].

Une fois de plus, regardons d’abord les petits cas, en n’oubliant pas de commencer par le plus petit d’entre eux. Le plan, sans aucune droite, comprend une région ; avec une droite, il a deux régions ; et avec deux droites il a quatre régions :



$$L_0 = 1$$

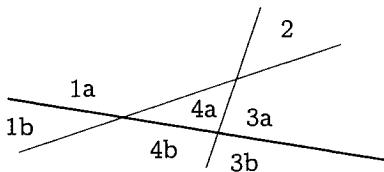
$$L_1 = 2$$

$$L_2 = 4$$

(Chaque droite se prolonge infiniment dans les deux directions).

## 6 PROBLÈMES RÉCURRENTS

Il est évident que  $L_n = 2^n$ , pensons-nous ! En ajoutant une droite, on double tout simplement le nombre de régions. C'est malheureusement faux. On pourrait y arriver si la  $n$ ième droite partageait toutes les anciennes régions en deux. Il est indubitable qu'elle partage chacune d'entre elles en au plus deux parties, puisqu'elles sont toutes convexes. (Une droite peut partager une région convexe en au plus deux régions, qui sont elles mêmes convexes). Cependant, quand on ajoute la troisième droite — en gras dans la figure ci-dessous — on s'aperçoit rapidement qu'elle ne peut pas traverser plus de trois régions, quelles que soient les positions des deux autres droites :



*Une région est convexe si elle contient tout segment joignant deux quelconques de ses points (ce n'est pas ce qui est écrit dans mon dictionnaire, mais c'est ce que pensent les mathématiciens).*

Ainsi, on ne peut faire mieux que  $L_3 = 4 + 3 = 7$ .

En réfléchissant un peu, on trouve la bonne façon de généraliser. La  $n$ ième droite (pour  $n > 0$ ) ajoute  $k$  régions si et seulement si elle traverse  $k$  anciennes régions ; et elle traverse  $k$  anciennes régions si et seulement si elle croise les droites précédentes en  $k - 1$  endroits différents. Comme deux droites distinctes ne peuvent se couper qu'en un point au plus, la nouvelle droite peut croiser les  $n - 1$  anciennes qu'en  $n - 1$  points distincts au plus ; donc  $k \leq n$ . Nous avons ainsi établi la borne supérieure.

$$L_n \leq L_{n-1} + n, \quad \text{pour } n > 0.$$

De plus, on montre facilement par induction que l'inégalité de cette formule peut être remplacée par une égalité. Il suffit de positionner la  $n$ ième droite de sorte qu'elle ne soit parallèle à aucune des autres (donc elle les coupe toutes), et qu'elle ne passe par aucun des points d'intersection existants (ainsi elle coupe les autres droites en des points distincts). Voici donc la récurrence :

$$\begin{aligned} L_0 &= 1; \\ L_n &= L_{n-1} + n, \quad \text{pour } n > 0. \end{aligned} \tag{1.4}$$

Les valeurs de  $L_1$ ,  $L_2$  et  $L_3$  que nous avons déjà calculées collent parfaitement ; la récurrence est donc adoptée.

Il nous faut maintenant une formule close pour la solution. On pourrait de nouveau jouer aux devinettes, mais la suite 1, 2, 4, 7, 11, 16, ... ne nous est pas franchement familière. Essayons donc autre chose. Il est souvent possible de comprendre une récurrence en la "développant" ou en la

“dépliant” entièrement comme ceci :

$$\begin{aligned}
 L_n &= L_{n-1} + n \\
 &= L_{n-2} + (n-1) + n \\
 &= L_{n-3} + (n-2) + (n-1) + n \\
 &\quad \vdots \\
 &= L_0 + 1 + 2 + \cdots + (n-2) + (n-1) + n \\
 &= 1 + S_n, \quad \text{où } S_n = 1 + 2 + 3 + \cdots + (n-1) + n.
 \end{aligned}$$

En d’autres termes,  $L_n$  est égal à 1 plus  $S_n$ , la somme des  $n$  premiers entiers strictement positifs.

Comme nous rencontrerons plusieurs fois cette quantité  $S_n$  dans ce livre, profitons-en pour faire un tableau contenant ses premières valeurs. Ainsi nous reconnaîtrons plus facilement ces nombres la prochaine fois que nous les verrons :

$n$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$S_n$	1	3	6	10	15	21	28	36	45	55	66	78	91	105

Ces nombres sont aussi appelés  *nombres triangulaires*, car  $S_n$  compte le nombre de quilles de bowling comprises dans un triangle de hauteur  $n$ . Par exemple, le triangle de hauteur 4 contient  $S_4 = 10$  quilles.

Pour calculer  $S_n$ , on peut utiliser l’astuce que Gauss trouva, dit-on, en 1786, alors qu’il avait neuf ans [88] (voir aussi Euler [114, partie 1, §415]) :

$$\begin{aligned}
 S_n &= 1 + 2 + 3 + \cdots + (n-1) + n \\
 + S_n &= n + (n-1) + (n-2) + \cdots + 2 + 1 \\
 \hline
 2S_n &= (n+1) + (n+1) + (n+1) + \cdots + (n+1) + (n+1)
 \end{aligned}$$

On attribue beaucoup de choses à Gauss. Soit il était vraiment futé, soit il avait un excellent attaché de presse.

Ou alors peut-être qu’il émanait de lui un magnétisme irrésistible.

On dit souvent que Gauss est le plus grand mathématicien de tous les temps. Alors ça fait du bien de comprendre au moins un de ses résultats.

Il suffit d’ajouter  $S_n$  à son image-miroir, de sorte que la somme de chacune des  $n$  colonnes de droite fasse  $n+1$ . En simplifiant, on obtient

$$S_n = \frac{n(n+1)}{2}, \quad \text{pour } n \geq 0. \tag{1.5}$$

Parfait, nous avons notre solution :

$$L_n = \frac{n(n+1)}{2} + 1, \quad \text{pour } n \geq 0. \tag{1.6}$$

En tant qu’experts, nous pourrions nous satisfaire de ce calcul et l’accepter comme preuve, même si le dépliage et l’image-miroir ont été décrits un peu vaguement. Cependant, les étudiants en mathématiques pourraient

## 8 PROBLÈMES RÉCURRENTS

être confrontés à de plus strictes exigences. Faisons donc une preuve rigoureuse par induction, ça ne pourra pas faire de mal. Voici le point clé de l'induction :

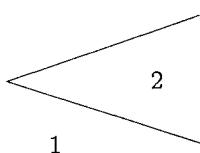
$$L_n = L_{n-1} + n = \left(\frac{1}{2}(n-1)n+1\right) + n = \frac{1}{2}n(n+1) + 1.$$

Maintenant, on ne peut plus avoir de doute sur la formule close (1.6).

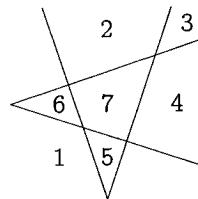
A ce propos, nous avons parlé plusieurs fois de "formule close" sans avoir explicitement défini ce que c'était. En général, c'est assez clair. Les récurrences, comme (1.1) et (1.4), ne sont pas des formules closes, car elles expriment une quantité en fonction d'elle-même. Par contre, les formules (1.2) et (1.6) en sont. Les sommes comme  $1 + 2 + \dots + n$  ne sont pas des formules closes, car elles trichent en utilisant ' $\dots$ ', mais les expressions du genre de  $n(n+1)/2$  en sont. On pourrait grossièrement définir une formule close comme ceci : une expression d'une quantité  $f(n)$  est une formule close si on peut la calculer en utilisant au plus un nombre fixé d'opérations standard "bien connues", ce nombre étant indépendant de  $n$ . Par exemple,  $2^n - 1$  et  $n(n+1)/2$  sont des formules closes car elles utilisent, de façon explicite, uniquement l'addition, la soustraction, la multiplication, la division et l'exponentielle.

Il existe des récurrences pour lesquelles il n'existe pas de formule close simple. Si une telle récurrence s'avère importante, du fait qu'on la voit souvent apparaître, on peut se permettre d'ajouter de nouvelles opérations à la liste des opérations "standard". On peut ainsi considérablement agrandir l'ensemble des problèmes dont la solution est une formule close "simple". Par exemple, le produit des  $n$  premiers entiers non nuls,  $n!$ , s'est avéré tellement important qu'on le considère maintenant comme une opération de base. Ainsi l'expression  $n!$  est une formule close, bien que son équivalent ' $1 \cdot 2 \cdot \dots \cdot n$ ' n'en constitue pas une.

Voyons maintenant brièvement une variante du problème des droites dans le plan : supposons qu'à la place de lignes droites on trace des lignes brisées, chacune d'entre elles contenant une seule "brisure". Quel est le nombre maximum  $Z_n$  de régions déterminées par  $n$  lignes brisées de ce type tracées dans le plan ? On pourrait s'attendre à ce que  $Z_n$  soit à peu près deux fois plus grand que  $L_n$ , ou peut-être trois fois plus. Voyons voir :



$$Z_1 = 2$$



$$Z_2 = 7$$

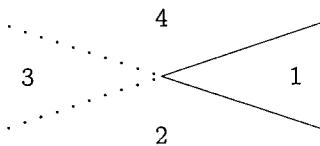
A partir de ces trois petits cas, en réfléchissant un peu, on réalise qu'une

*En cas de doute, observez bien les mots. Pourquoi est-elle "close", par opposition à "ouverte" ?*

*Réponse : l'équation est "close", en ce sens qu'elle ne réfère pas à elle-même, contrairement à une récurrence. L'affaire est "close", on n'en parlera plus. La clé réside dans les métaphores.*

*... et un peu plus encore...*

ligne brisée peut être considérée comme deux droites qui se coupent, sauf que les régions situées “après” leur point d’intersection s’unissent pour n’en former qu’une.



Les régions 2, 3 et 4, qui seraient distinctes si on considérait deux droites, constituent une seule région dans le cas d’une ligne brisée. On perd donc deux régions. Deux régions par ligne, c’est tout ce qu’on perdra dans tous les cas si on s’y prend correctement, c’est-à-dire si la brisure se situe “après” les intersections avec toutes les lignes tracées auparavant. Ainsi,

Voir les détails dans l’exercice 18.

$$\begin{aligned} Z_n &= L_{2n} - 2n = 2n(2n+1)/2 + 1 - 2n \\ &= 2n^2 - n + 1, \quad \text{pour } n \geq 0. \end{aligned} \quad (1.7)$$

En comparant les formules closes (1.6) et (1.7), on trouve que, lorsque  $n$  est grand,

$$\begin{aligned} L_n &\sim \frac{1}{2}n^2, \\ Z_n &\sim 2n^2; \end{aligned}$$

donc il y a à peu près quatre fois plus de régions avec des lignes brisées qu’avec des lignes droites. (Nous verrons dans d’autres chapitres comment connaître de façon approchée le comportement d’une fonction entière lorsque  $n$  est grand. Le symbole ‘~’ est défini dans la section 9.1).

### 1.3 LE PROBLÈME DE JOSÈPHE

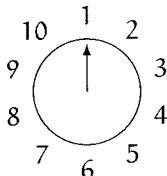
(Ahrens [5, vol. 2] et Herstein et Kaplansky [187] rapportent l’intéressante histoire de ce problème. Josèphe lui-même [197] reste assez vague à ce sujet).

... grâce à quoi nous pouvons lire cette histoire.

Notre dernier exemple introductif est une variante d’un problème ancien, à qui on a donné le nom de Flavius Josèphe, célèbre historien du premier siècle. D’après la légende, les fameux talents de mathématiciens de Josèphe lui auraient sauvé la vie. Pendant la guerre entre Rome et les Juifs, il faisait partie d’un groupe de 41 rebelles juifs piégés dans une cave par les Romains. Choisissant de mourir plutôt que d’être capturés, les rebelles décidèrent de former un cercle puis, en suivant sa circonférence, de tuer un homme sur trois comptés parmi les survivants, jusqu’à ce qu’il ne reste plus personne. Josèphe ne voulait pas ce suicide absurde. Il en était de même pour un autre conspirateur, resté anonyme. Alors Josèphe calcula rapidement où son ami et lui devaient se placer dans le cercle vicieux.

## 10 PROBLÈMES RÉCURRENTS

Dans notre variante, on commence avec  $n$  personnes numérotées de 1 à  $n$  sur un cercle, puis on élimine une personne sur *deux* jusqu'à ce qu'il n'y ait plus qu'un survivant. Voici par exemple la configuration de départ pour  $n = 10$  :



On élimine dans l'ordre 2, puis 4, 6, 8, 10, 3, 7, 1, 9 ; le survivant est 5. Le problème posé en général est le suivant : déterminer le numéro  $J(n)$  du survivant.

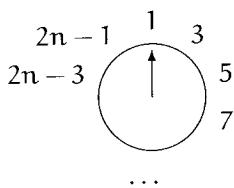
Nous venons de voir que  $J(10) = 5$ . On peut conjecturer que  $J(n) = n/2$  lorsque  $n$  est pair. Le cas  $n = 2$  satisfait cette conjecture :  $J(2) = 1$ . Cependant, l'étude de quelques autres cas nous dissuade vite : la conjecture est fausse pour  $n = 4$  et  $n = 6$ .

*Voici un cas où poser  $n = 0$  n'a aucun sens.*

$n$	1	2	3	4	5	6
$J(n)$	1	1	3	1	3	5

Remettons-nous au travail pour trouver une meilleure conjecture. Voyons voyons... on dirait que  $J(n)$  est toujours impair. Il y a en effet une bonne raison à cela : les nombres pairs sont tous éliminés pendant le premier tour du cercle. De plus, si le nombre  $n$  de personnes est lui-même pair, on se retrouve à la fin du premier tour dans une configuration semblable à celle du début, à deux différences près : il reste deux fois moins de participants, et leurs numéros ont changé.

Supposons donc qu'il y a  $2n$  personnes au départ. Après le premier tour, on se retrouve avec



*Un mauvaise conjecture n'est pas forcément du temps perdu : cela permet de s'impliquer plus encore dans le problème.*

*C'est là la ruse : on a*  
 $J(2n) =$   
*nouveau( $J(n)$ )*,  
*où*  
*nouveau( $k$ ) =*  
 $2k - 1$ .

et le prochain candidat choisi sera le 3. C'est exactement comme si on commençait avec  $n$  personnes, en ayant toutefois multiplié chaque numéro par 2, puis retranché 1. En d'autres termes,

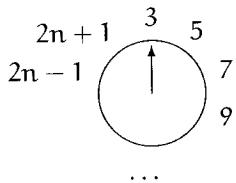
$$J(2n) = 2J(n) - 1, \quad \text{pour } n \geq 1.$$

On peut maintenant calculer rapidement  $J(n)$  pour des plus grandes valeurs de  $n$ . Nous savons par exemple que  $J(10) = 5$ , donc

$$J(20) = 2J(10) - 1 = 2 \cdot 5 - 1 = 9.$$

De même,  $J(40) = 17$ , et on en déduit aisément que  $J(5 \cdot 2^m) = 2^{m+1} + 1$ .

Mais que se passe-t-il dans le cas impair ? S'il y a  $2n+1$  personnes, la personne numéro 1 est éliminée juste après la personne numéro  $2n$ , et on se retrouve dans la situation suivante :



On se retrouve encore une fois dans la situation de départ avec  $n$  personnes, mais cette fois on a *ajouté 1* à chaque nombre après l'avoir multiplié par deux. Donc

$$J(2n+1) = 2J(n) + 1, \quad \text{pour } n \geq 1.$$

En ajoutant  $J(1) = 1$  à ces équations, on obtient une récurrence qui définit  $J$  dans tous les cas :

$$\begin{aligned} J(1) &= 1; \\ J(2n) &= 2J(n) - 1, \quad \text{pour } n \geq 1; \\ J(2n+1) &= 2J(n) + 1, \quad \text{pour } n \geq 1. \end{aligned} \tag{1.8}$$

Cette récurrence est beaucoup plus efficace qu'une relation qui donnerait  $J(n)$  en fonction de  $J(n-1)$ , car elle divise  $n$  par deux au moins chaque fois qu'on l'applique. Ainsi, on peut par exemple calculer  $J(1000000)$  en n'appliquant que 19 fois (1.8). Cependant nous allons quand même chercher une formule close, parce qu'elle sera plus rapide encore, et plus instructive aussi. N'oublions pas que c'est une question de vie ou de mort !

Avec notre récurrence, on peut écrire très rapidement un tableau des petites valeurs de  $J(n)$ . Il pourra peut-être nous aider à deviner la réponse.

$n$	1	2 3	4 5 6 7	8 9 10 11 12 13 14 15	16
$J(n)$	1	1 3	1 3 5 7	1 3 5 7 9 11 13 15	1

Ça y est ! On dirait bien qu'on peut grouper les nombres par puissances de 2, comme indiqué dans le tableau par des traits verticaux.  $J(n)$  est toujours égal à 1 au début de chaque groupe, puis il augmente de 2 en 2 dans le même groupe. Ainsi, si l'on pose  $n = 2^m + l$ , où  $2^m$  est la plus

## 12 PROBLÈMES RÉCURRENTS

grande puissance de 2 plus petite que  $n$  et  $l$  représente le reste, il semblerait que la solution de notre récurrence soit

$$J(2^m + l) = 2l + 1, \quad \text{pour } m \geq 0 \text{ et } 0 \leq l < 2^m. \quad (1.9)$$

(Notons que si  $2^m \leq n < 2^{m+1}$ , le reste  $l = n - 2^m$  satisfait  $0 \leq l < 2^{m+1} - 2^m = 2^m$ ).

Il nous faut maintenant prouver (1.9). Nous allons, comme précédemment, utiliser l'induction ; mais cette fois elle portera sur  $m$ . Si  $m = 0$ , alors nécessairement  $l = 0$  ; ainsi la base de (1.9) est  $J(1) = 1$ , ce qui est correct. L'étape d'induction comprend deux parties, selon que  $l$  est pair ou impair. Si  $m > 0$  et  $2^m + l = 2n$ , alors  $l$  est pair et

$$\begin{aligned} J(2^m + l) &= 2J(2^{m-1} + l/2) - 1 \\ &= 2(2l/2 + 1) - 1 = 2l + 1, \end{aligned}$$

de par (1.8) et l'hypothèse d'induction ; c'est exactement ce que nous demandions. Une preuve similaire permet de résoudre le cas impair, lorsque  $2^m + l = 2n + 1$ . Notons en passant que (1.8) entraîne la relation

$$J(2n + 1) - J(2n) = 2.$$

Quoi qu'il en soit, l'induction est complète et (1.9) établie.

Pour illustrer cette solution, calculons  $J(100)$  : comme  $100 = 2^6 + 36$ , on obtient  $J(100) = 2 \cdot 36 + 1 = 73$ .

Maintenant que nous avons fait le plus difficile (la solution du problème), voyons quelque chose de plus facile : toute solution d'un problème peut être généralisée pour s'appliquer à une classe plus large de problèmes. Une fois qu'on a appris une technique, il s'avère instructif de l'étudier de près et de voir jusqu'où on peut aller avec. Dans la suite de cette section, nous allons donc examiner la solution (1.9) et explorer quelques généralisations de la récurrence (1.8). Nous allons découvrir grâce à cela la structure qui sous-tend tous les problèmes de ce genre.

Les puissances de 2 ont joué un rôle important dans notre raisonnement. Il est donc naturel de se pencher sur les représentations en base 2 de  $n$  et  $J(n)$ . Supposons que  $n$  s'écrit en base 2 comme suit :

$$n = (b_m b_{m-1} \dots b_1 b_0)_2.$$

Cela signifie que

$$n = b_m 2^m + b_{m-1} 2^{m-1} + \dots + b_1 2 + b_0,$$

*Mais il y une façon plus simple de procéder ! La clé réside dans le fait que  $J(2^m) = 1$  pour tout  $m$ , ce qui se déduit immédiatement de notre première équation,  $J(2n) = 2J(n) - 1$ .*

*Nous savons donc que la première personne survit chaque fois que  $n$  est une puissance de 2.*

*Dans le cas général, quand  $n = 2^m + l$ , le nombre de personnes devient une puissance de 2 dès qu'il y a eu l exécutions. La première personne, parmi celles qui restent à ce moment-là, porte le numéro  $2l + 1$ .*

où chaque  $b_i$  a pour valeur 0 ou 1, et le premier bit  $b_m$  est égal à 1. sachant que  $n = 2^m + 1$ , on a, successivement,

$$\begin{aligned} n &= (1 b_{m-1} b_{m-2} \dots b_1 b_0)_2, \\ l &= (0 b_{m-1} b_{m-2} \dots b_1 b_0)_2, \\ 2l &= (b_{m-1} b_{m-2} \dots b_1 b_0 0)_2, \\ 2l+1 &= (b_{m-1} b_{m-2} \dots b_1 b_0 1)_2, \\ J(n) &= (b_{m-1} b_{m-2} \dots b_1 b_0 b_m)_2. \end{aligned}$$

(La dernière ligne provient du fait que  $J(n) = 2l + 1$  et  $b_m = 1$ ). Nous avons donc prouvé que

$$J((b_m b_{m-1} \dots b_1 b_0)_2) = (b_{m-1} \dots b_1 b_0 b_m)_2. \quad (1.10)$$

Un informaticien dirait qu'on trouve  $J(n)$  à partir de  $n$  en effectuant un décalage circulaire d'un bit vers la gauche. C'est magique ! Par exemple, si  $n = 100 = (1100100)_2$ , alors  $J(n) = J((1100100)_2) = (1001001)_2$ , ce qui fait  $64 + 8 + 1 = 73$ . Si nous avions travaillé en représentation binaire depuis le début, nous aurions probablement immédiatement remarqué ce fait.

Si on itère la fonction  $J$   $m + 1$  fois en partant de  $n$ , on effectue  $m + 1$  décalages circulaires d'un bit. Puisque  $n$  s'écrit avec  $(m+1)$  bits, on devrait donc s'attendre à trouver  $n$  ; mais ça ne marche pas. Par exemple, si  $n = 13$ , on a  $J((1101)_2) = (1011)_2$  ; mais alors  $J((1011)_2) = (111)_2$  et le processus se détraque : le 0 disparaît lorsqu'il arrive en première position. En fait, comme  $J(n)$  est le numéro du survivant, il est  $\leq n$  par définition ; donc, si  $J(n) < n$ , on ne peut jamais retrouver  $n$  en continuant à itérer.

En appliquant  $J$  de façon répétée, on obtient une suite décroissante de valeurs qui finit par atteindre un "point fixe" pour lequel  $J(n) = n$ . On voit facilement ce que sera ce point fixe, grâce à une propriété évidente de notre décalage circulaire : en itérant la fonction un nombre suffisant de fois, on obtient toujours un nombre représenté par une suite de 1 ; ce nombre a pour valeur  $2^{\nu(n)} - 1$ , où  $\nu(n)$  est le nombre de chiffres 1 dans la représentation binaire de  $n$ . Ainsi, comme  $\nu(13) = 3$ , on obtient

au moins deux )  

$$\overbrace{J(J(\dots J(13) \dots))}^{au \ moins \ deux} = 2^3 - 1 = 7;$$

de la même façon,

au moins 8  

$$\overbrace{J(J(\dots J((101101101101011)_2) \dots))}^{au \ moins \ 8} = 2^{10} - 1 = 1023.$$

Incroyable mais vrai.

*(Ici, "itérer" signifie appliquer une fonction sur elle-même).*

*Curieusement, si  $M$  est une  $n$ -variété  $C^\infty$  compacte ( $n > 1$ ), il existe un plongement différentiable de  $M$  dans  $\mathbf{R}^{2n-\nu(n)}$ , mais pas nécessairement dans  $\mathbf{R}^{2n-\nu(n)-1}$ .*  
*Je me demande si Josèphe n'était pas un topographe ignoré.*

## 14 PROBLÈMES RÉCURRENTS

Revenons brièvement à notre première conjecture :  $J(n) = n/2$  lorsque  $n$  est pair. C'est évidemment faux en général, mais nous pouvons maintenant déterminer dans quels cas c'est vrai :  $J(n) = n/2$ ,  $2l + 1 = (2^m + l)/2$ ,  $l = \frac{1}{3}(2^m - 2)$ . Si le nombre  $l = \frac{1}{3}(2^m - 2)$  est entier, alors  $n = 2^m + l$  est une solution car  $l$  est plus petit que  $2^m$ . Il est facile de vérifier que  $2^m - 2$  est un multiple de 3 lorsque  $m$  est impair, mais pas quand il est pair (nous étudierons des problèmes de ce genre au chapitre 4). Il y a donc une infinité de solutions à l'équation  $J(n) = n/2$ . Voici les premières :

$m$	$l$	$n = 2^m + l$	$J(n) = 2l + 1 = n/2$	$n$ (en base 2)
1	0	2	1	10
3	2	10	5	1010
5	10	42	21	101010
7	42	170	85	10101010

Remarquez que la dernière colonne est constituée des nombres binaires pour lesquels un décalage circulaire d'un chiffre vers la gauche produit le même effet qu'un décalage ordinaire d'un chiffre vers la droite (division par 2).

Maintenant que nous connaissons extrêmement bien la fonction  $J$ , il s'agit de la généraliser. Que se serait-il passé si nous avions dû résoudre une récurrence du genre de (1.8), mais avec des constantes différentes ? Nous n'aurions probablement pas pu deviner la solution, car elle doit être plutôt étrange. Voyons cela de plus près en introduisant des constantes  $\alpha$ ,  $\beta$  et  $\gamma$  et en cherchant une forme close pour la récurrence plus générale

*Pour moi, c'est comme si c'était du grec.*

$$\begin{aligned} f(1) &= \alpha; \\ f(2n) &= 2f(n) + \beta, \quad \text{pour } n \geq 1; \\ f(2n+1) &= 2f(n) + \gamma, \quad \text{pour } n \geq 1. \end{aligned} \tag{1.11}$$

(Dans la récurrence d'origine, nous avions  $\alpha = 1$ ,  $\beta = -1$  et  $\gamma = 1$ ). Avec cette récurrence, on peut construire le tableau suivant :

$n$	$f(n)$	
1	$\alpha$	
2	$2\alpha + \beta$	
3	$2\alpha + \beta + \gamma$	
4	$4\alpha + 3\beta$	
5	$4\alpha + 2\beta + \gamma$	
6	$4\alpha + \beta + 2\gamma$	
7	$4\alpha + \beta + 3\gamma$	
8	$8\alpha + 7\beta$	
9	$8\alpha + 6\beta + \gamma$	

Les coefficients de  $\alpha$  semblent bien être les plus grandes puissances de 2 inférieures ou égales à  $n$ . De plus, entre deux puissances de 2 consécutives, les coefficients de  $\beta$  décroissent de 1 en 1 jusqu'à 0, tandis que ceux de  $\gamma$  croissent de 1 en 1 à partir de 0. Donc, si on écrit  $f(n)$  sous la forme

$$f(n) = A(n)\alpha + B(n)\beta + C(n)\gamma \quad (1.13)$$

pour distinguer ce qui dépend de  $\alpha$ ,  $\beta$  et  $\gamma$ , il semble que

$$\begin{aligned} A(n) &= 2^m; \\ B(n) &= 2^m - 1 - l; \\ C(n) &= l, \end{aligned} \quad (1.14)$$

où, comme d'habitude,  $n = 2^m + l$  et  $0 \leq l < 2^m$ , pour  $n \geq 1$ .

*Accrochez-vous, ce qui va suivre est nouveau.*

Il n'est pas extrêmement difficile de prouver (1.13) et (1.14) par induction, mais les calculs ne sont pas très propres et ne nous apprennent rien. Il y a heureusement une meilleure façon de procéder : il s'agit de choisir des valeurs particulières et de les combiner entre elles. Illustrons ceci en considérant le cas particulier  $\alpha = 1$ ,  $\beta = \gamma = 0$ . La fonction  $f(n)$  est donc supposée égale à  $A(n)$ , et la récurrence (1.11) devient

$$\begin{aligned} A(1) &= 1; \\ A(2n) &= 2A(n), \quad \text{pour } n \geq 1; \\ A(2n+1) &= 2A(n), \quad \text{pour } n \geq 1. \end{aligned}$$

On en déduit facilement, par induction sur  $m$ , que  $A(2^m + l) = 2^m$ .

*Bonne idée !*

Maintenant, nous allons utiliser la récurrence (1.11) et la solution (1.13) à l'envers, en partant d'une fonction  $f(n)$  et en regardant s'il existe des constantes  $(\alpha, \beta, \gamma)$  qui permettent de la définir. En injectant la fonction constante  $f(n) = 1$  dans (1.11), on trouve

$$\begin{aligned} 1 &= \alpha; \\ 1 &= 2 \cdot 1 + \beta; \\ 1 &= 2 \cdot 1 + \gamma. \end{aligned}$$

La solution de ce système est le triplet  $(\alpha, \beta, \gamma) = (1, -1, -1)$  ; ainsi,  $A(n) - B(n) - C(n) = f(n) = 1$ . On peut de la même manière injecter  $f(n) = n$  dans (1.11) :

$$\begin{aligned} 1 &= \alpha; \\ 2n &= 2 \cdot n + \beta; \\ 2n+1 &= 2 \cdot n + \gamma; \end{aligned}$$

Ces équations sont vérifiées pour tout  $n$  si  $\alpha = 1$ ,  $\beta = 0$  et  $\gamma = 1$  ; inutile donc de prouver par induction que, avec ces valeurs pour les paramètres  $\alpha$ ,

## 16 PROBLÈMES RÉCURRENTS

$\beta$  et  $\gamma$ ,  $f(n)$  sera égal à  $n$ . Nous savons déjà que la solution sera  $f(n) = n$  dans ce cas, car la récurrence (1.11) définit  $f(n)$  de façon unique pour toute valeur de  $n$ .

C'est quasiment terminé ! Nous avons montré que les fonctions  $A(n)$ ,  $B(n)$  et  $C(n)$  de (1.13), qui donnent la solution générale de (1.11), satisfont les équations

$$\begin{aligned} A(n) &= 2^m, \quad \text{où } n = 2^m + l \text{ et } 0 \leq l < 2^m; \\ A(n) - B(n) - C(n) &= 1; \\ A(n) + C(n) &= n. \end{aligned}$$

Nos conjectures de (1.14) s'ensuivent immédiatement, du fait qu'en résolvant ces équations on trouve  $C(n) = n - A(n) = l$  et  $B(n) = A(n) - 1 - C(n) = 2^m - 1 - l$ .

Cette approche est une illustration d'une méthode étonnamment utile pour résoudre des récurrences, la *méthode du répertoire*. Il s'agit d'abord de trouver des valeurs des paramètres pour lesquelles la solution est connue. On constitue ainsi un répertoire de cas particuliers que l'on sait résoudre. On résout alors le cas général en combinant ces cas particuliers. Il faut utiliser autant de cas particuliers indépendants que le nombre de paramètres du cas général (ici trois, pour  $\alpha$ ,  $\beta$  and  $\gamma$ ). Les exercices 16 et 20 donnent d'autres exemples de la méthode du répertoire.

Revenons à la récurrence d'origine. Nous savons qu'elle a une solution magique en base 2 :

$$J((b_m b_{m-1} \dots b_1 b_0)_2) = (b_{m-1} \dots b_1 b_0 b_m)_2, \quad \text{où } b_m = 1.$$

La récurrence de Josèphe généralisée possède-t-elle aussi cette magie ?

Bien sûr ! Pourquoi pas ? On peut réécrire la récurrence généralisée (1.11) ainsi :

$$\begin{aligned} f(1) &= \alpha; \\ f(2n+j) &= 2f(n) + \beta_j, \quad \text{pour } j = 0, 1 \quad \text{et} \quad n \geq 1, \end{aligned} \tag{1.15}$$

en posant  $\beta_0 = \beta$  et  $\beta_1 = \gamma$ . En développant cette récurrence en mode binaire, on obtient

$$\begin{aligned} f((b_m b_{m-1} \dots b_1 b_0)_2) &= 2f((b_m b_{m-1} \dots b_1)_2) + \beta_{b_0} \\ &= 4f((b_m b_{m-1} \dots b_2)_2) + 2\beta_{b_1} + \beta_{b_0} \\ &\quad \vdots \\ &= 2^m f((b_m)_2) + 2^{m-1} \beta_{b_{m-1}} + \dots + 2\beta_{b_1} + \beta_{b_0} \\ &= 2^m \alpha + 2^{m-1} \beta_{b_{m-1}} + \dots + 2\beta_{b_1} + \beta_{b_0}. \end{aligned}$$

Attention : les auteurs espèrent nous donner une idée de la méthode du répertoire en nous la présentant sur des cas d'école au lieu de nous la décrire de façon structurée. Cette méthode marche le mieux avec les récurrences "linéaires", c'est-à-dire les récurrences dont les solutions peuvent s'exprimer comme une somme de produits de certains paramètres par des fonctions de  $n$ , comme dans l'équation (1.13). La formule-clé est (1.13).

(“relâcher” =  
“détruire”)

Supposons qu'on relâche les contraintes de l'écriture en base 2 pour autoriser n'importe quel chiffre, au lieu de 0 et 1 seulement. On déduit du calcul précédent que

$$f((b_m b_{m-1} \dots b_1 b_0)_2) = (\alpha \beta_{b_{m-1}} \beta_{b_{m-2}} \dots \beta_{b_1} \beta_{b_0})_2. \quad (1.16)$$

Bien. Nous aurions pu voir cela plus tôt si nous avions écrit (1.12) d'une autre manière :

n	f(n)
1	$\alpha$
2	$2\alpha + \beta$
3	$2\alpha + \gamma$
4	$4\alpha + 2\beta + \beta$
5	$4\alpha + 2\beta + \gamma$
6	$4\alpha + 2\gamma + \beta$
7	$4\alpha + 2\gamma + \gamma$

*Je crois que j'ai compris : les représentations binaires de A(n), B(n) et C(n) ont leurs 1 dans des positions différentes.*

Par exemple, lorsque  $n = 100 = (1100100)_2$ , en prenant les valeurs d'origine  $\alpha = 1$ ,  $\beta = -1$  et  $\gamma = 1$ , on obtient

$$\begin{aligned} n &= (1 & 1 & 0 & 0 & 1 & 0 & 0)_2 = 100 \\ f(n) &= (1 & 1 & -1 & -1 & 1 & -1 & -1)_2 \\ &= +64 & +32 & -16 & -8 & +4 & -2 & -1 = 73 \end{aligned}$$

comme auparavant. La propriété du décalage circulaire se déduit du fait que chaque bloc de chiffres binaires  $(10 \dots 00)_2$  de la représentation de  $n$  se transforme en

$$(1-1 \dots -1-1)_2 = (00 \dots 01)_2.$$

*"There are two kinds of generalizations. One is cheap and the other is valuable. It is easy to generalize by diluting a little idea with a big terminology. It is much more difficult to prepare a refined and condensed extract from several good ingredients."*

— G. Pólya [297]

Ainsi, en changeant de notation nous avons trouvé la solution compacte (1.16) de la récurrence générale (1.15). On peut même se permettre de généraliser plus encore. La récurrence

$$\begin{aligned} f(j) &= \alpha_j, & \text{pour } 1 \leq j < d; \\ f(dn + j) &= cf(n) + \beta_j, & \text{pour } 0 \leq j < d \text{ et } n \geq 1, \end{aligned} \quad (1.17)$$

est la même que la précédente, sauf qu'on part de nombres en base  $d$  pour produire des nombres en base  $c$ . Ainsi, la solution, avec changement de base, est

$$f((b_m b_{m-1} \dots b_1 b_0)_d) = (\alpha_{b_m} \beta_{b_{m-1}} \beta_{b_{m-2}} \dots \beta_{b_1} \beta_{b_0})_c. \quad (1.18)$$

## 18 PROBLÈMES RÉCURRENTS

Supposons par exemple que, par chance, on nous donne la récurrence

*Ou peut-être par malchance.*

$$\begin{aligned}f(1) &= 34, \\f(2) &= 5, \\f(3n) &= 10f(n) + 76, \quad \text{pour } n \geq 1, \\f(3n+1) &= 10f(n) - 2, \quad \text{pour } n \geq 1, \\f(3n+2) &= 10f(n) + 8, \quad \text{pour } n \geq 1,\end{aligned}$$

pour calculer  $f(19)$ . Ici, on a  $d = 3$  et  $c = 10$ . Nous savons que  $19 = (201)_3$ , et la solution avec changement de base nous indique qu'il faut effectuer un remplacement chiffre par chiffre, en passant de la base 3 vers la base 10. Ainsi, le 2 devient un 5, et les 0 et 1 se transforment respectivement en 76 et  $-2$ , pour donner la réponse :

$$f(19) = f((201)_3) = (5\ 76\ -2)_{10} = 1258,$$

Ainsi, Josèphe et la guerre entre Rome et les Juifs nous ont amenés à rencontrer quelques récurrences générales intéressantes.

*Je suis contre les guerres récurrentes.*

## Exercices

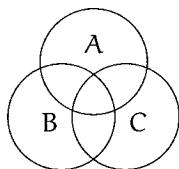
### Echauffements

- 1 Tous les chevaux ont la même couleur ; on peut le prouver par induction sur le nombre de chevaux d'un ensemble donné. Voici comment : "s'il n'y a qu'un cheval, il est de la même couleur que lui-même, donc la base est triviale. Pour l'étape d'induction, supposons qu'il y a  $n$  chevaux numérotés de 1 à  $n$ . De par l'hypothèse d'induction, les chevaux 1 à  $n-1$  sont de la même couleur et les chevaux de 2 à  $n$  sont de la même couleur. Cependant, les chevaux numérotés de 2 à  $n-1$  ne peuvent pas changer de couleur en changeant de groupe ; ce sont des chevaux, non des caméléons. Donc, par transitivité, les chevaux de 1 à  $n$  doivent tous être de la même couleur. Ainsi les  $n$  chevaux ont tous la même couleur. CQFD." Qu'est-ce qui cloche dans ce raisonnement ?
- 2 Trouver la plus courte suite de mouvements pour transférer une tour de  $n$  disques de l'axe de gauche A vers l'axe de droite B, si on interdit les transferts directs entre A et B (on ne peut effectuer un mouvement que depuis ou vers l'axe du milieu). Comme d'habitude, on ne peut pas poser un disque sur un disque plus petit.
- 3 Montrer que, au cours du transfert d'une tour sous les hypothèses de l'exercice précédent, on va rencontrer toutes les configurations possibles de  $n$  disques convenablement empilés sur trois axes.

*Vous êtes priés de faire tous les échauffements de tous les chapitres !*

— La Directio

- 4 Existe-t-il des configurations de départ et d'arrivée de  $n$  disques sur trois axes qui sont espacées de plus de  $2^n - 1$  mouvements, si on respecte les règles originales de Lucas ?
- 5 On utilise souvent un “diagramme de Venn”, constitué de trois cercles qui se recouvrent partiellement, pour illustrer les huit sous-ensembles que l'on peut associer à trois ensembles donnés :



Peut-on illustrer les seize possibilités offertes par quatre ensembles en traçant, de façon similaire, quatre cercles ?

- 6 Parmi les régions définies par  $n$  droites dans le plan, certaines sont infinies, d'autres bornées. Combien peut-il y avoir de régions bornées au maximum ?
- 7 Soit  $H(n) = J(n+1) - J(n)$ . L'équation (1.8) nous indique que  $H(2n) = 2$  et  $H(2n+1) = J(2n+2) - J(2n+1) = (2J(n+1) - 1) - (2J(n) + 1) = 2H(n) - 2$ , pour tout  $n \geq 1$ . Il semble donc possible de prouver, par induction sur  $n$ , que  $H(n) = 2$  pour tout  $n$ . Qu'est-ce qui cloche ?

### **Devoirs à la maison**

- 8 Résoudre la récurrence

$$\begin{aligned} Q_0 &= \alpha; & Q_1 &= \beta; \\ Q_n &= (1 + Q_{n-1})/Q_{n-2}, & \text{pour } n > 1. \end{aligned}$$

On suppose que  $Q_n \neq 0$  pour tout  $n \geq 0$ . *Suggestion* :  $Q_4 = (1+\alpha)/\beta$ .

... maintenant, il y a un cheval de couleur différente.

- 9 On peut parfois utiliser l'induction à l'envers, en prouvant des choses en allant de  $n$  à  $n - 1$  au lieu du contraire. Considérons par exemple la proposition

$$P(n) : \quad x_1 \dots x_n \leq \left( \frac{x_1 + \dots + x_n}{n} \right)^n, \quad \text{si } x_1, \dots, x_n \geq 0.$$

Elle est vraie lorsque  $n = 2$ , car  $(x_1 + x_2)^2 - 4x_1 x_2 = (x_1 - x_2)^2 \geq 0$ .

- a En posant  $x_n = (x_1 + \dots + x_{n-1})/(n-1)$ , prouver que  $P(n)$  implique  $P(n-1)$  lorsque  $n > 1$ .
- b Montrer que  $P(n)$  et  $P(2)$  impliquent  $P(2n)$ .
- c Expliquer pourquoi cela entraîne que  $P(n)$  est vraie pour tout  $n$ .

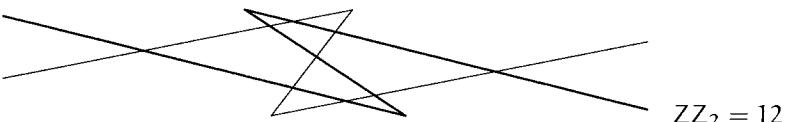
## 20 PROBLÈMES RÉCURRENTS

- 10 Soit  $Q_n$  le nombre minimum de mouvements nécessaires pour transférer une tour de  $n$  disques de A à B, en imposant que tous les mouvements se fassent *dans les sens des aiguilles d'une montre*, c'est-à-dire soit de A vers B, soit de B vers l'autre axe, soit de l'autre axe vers A. D'autre part, soit  $R_n$  le nombre minimum de mouvements requis pour transférer la tour de B vers A sous les mêmes restrictions. Montrer que

$$Q_n = \begin{cases} 0, & \text{si } n = 0; \\ 2R_{n-1} + 1, & \text{si } n > 0; \end{cases} \quad R_n = \begin{cases} 0, & \text{si } n = 0; \\ Q_n + Q_{n-1} + 1, & \text{si } n > 0. \end{cases}$$

(Pas besoin de résoudre ces récurrences pour l'instant ; nous verrons comment le faire au chapitre 7).

- 11 On appelle Tour de Hanoi double une tour qui contient  $2n$  disques de  $n$  tailles différentes, deux de chaque taille. Comme d'habitude, on ne peut déplacer qu'un disque à la fois, et on ne peut poser un disque que sur un disque plus grand.
- a On suppose que deux disques de même taille sont indiscernables l'un de l'autre. Combien de mouvements sont nécessaires pour transférer une tour double d'un axe à un autre ?
  - b Qu'en est-il si les disques de même taille doivent être disposés dans le même ordre dans la tour d'arrivée que dans la tour de départ ? [Suggestion : ce n'est pas facile, c'est en réalité une "question subsidiaire".]
- 12 Généralisons encore l'exercice 11a, en supposant qu'il y a  $n$  tailles de disques différentes, et exactement  $m_k$  disques de taille  $k$ . Déterminer  $A(m_1, \dots, m_n)$ , le nombre minimum de mouvements requis pour déplacer une tour, en supposant que les disques de même taille sont indistinguables les uns des autres.
- 13 Quel est le nombre maximum de régions que l'on peut obtenir avec  $n$  lignes brisées ainsi faites,



chacune d'entre elles étant composée de deux demi-droites infinies, jointes par un segment ?

- 14 En combien de parts peut-on partager un fromage épais si on le coupe cinq fois avec un couteau, sachant que le fromage doit rester dans sa position d'origine pendant toute l'opération et qu'une coupe correspond à un plan dans l'espace à trois dimensions. Trouver une relation de récurrence pour  $P_n$ , le nombre maximum de régions tri-dimensionnelles qui peuvent être définies par  $n$  plans distincts.

*Bon courage pour le laisser dans sa position d'origine !*

- 15 Josèphe avait un ami qui a été sauvé en se plaçant à l'avant-dernière position. Quelle est la valeur de  $I(n)$ , le numéro de l'avant-dernier survivant, quand on exécute une personne sur deux dans le cercle ?
- 16 Utiliser la méthode du répertoire pour résoudre la récurrence générale à quatre paramètres

$$\begin{aligned} g(1) &= \alpha; \\ g(2n+j) &= 3g(n) + \gamma n + \beta_j, \quad \text{pour } j = 0, 1 \quad \text{et} \quad n \geq 1. \end{aligned}$$

*Suggestion :* essayez la fonction  $g(n) = n$ .

### Problèmes d'examen

- 17 Soit  $W_n$  est le nombre minimum de mouvements nécessaires pour transférer une tour de  $n$  disques d'un axe à un autre, en supposant qu'on dispose de quatre axes au lieu de trois. Montrer que

$$W_{n(n+1)/2} \leq 2W_{n(n-1)/2} + T_n, \quad \text{pour } n > 0,$$

où  $T_n = 2^n - 1$  désigne le nombre de mouvements dans le cas où il y a trois axes. En déduire une formule close  $f(n)$  telle que  $W_{n(n+1)/2} \leq f(n)$  pour tout  $n \geq 0$ .

- 18 Montrer que l'ensemble de  $n$  lignes brisées décrit ci-après délimite  $Z_n$  régions, où  $Z_n$  est défini par (1.7) : la  $j$ ème ligne brisée, pour  $1 \leq j \leq n$ , a une brisure au point de coordonnées  $(n^{2j}, 0)$ , puis continue en passant par les points de coordonnées  $(n^{2j} - n^j, 1)$  et  $(n^{2j} - n^j - n^{-n}, 1)$ .
- 19 Peut-on obtenir  $Z_n$  régions avec  $n$  lignes brisées si l'angle de chaque brisure est de  $30^\circ$  ?
- 20 Utiliser la méthode du répertoire pour résoudre la récurrence à cinq paramètres suivante :

$$\begin{aligned} h(1) &= \alpha; \\ h(2n+j) &= 4h(n) + \gamma_j n + \beta_j, \quad \text{pour } j = 0, 1 \quad \text{et} \quad n \geq 1. \end{aligned}$$

*Suggestion :* essayez les fonctions  $h(n) = n$  et  $h(n) = n^2$ .

- 21 Supposons que  $2n$  personnes forment un cercle ; les  $n$  premières sont "gentilles", tandis que les  $n$  dernières sont "méchantes". Montrer qu'on peut toujours trouver un entier  $m$  (dépendant de  $n$ ) tel que, si on parcourt le cercle en exécutant chaque  $m$ ième personne, les méchants sont les premiers à disparaître. (Par exemple, si  $n = 3$ , on peut prendre  $m = 5$  ; si  $n = 4$ , on peut prendre  $m = 30$ ).

## 22 PROBLÈMES RÉCURRENTS

### Questions subsidiaires

- 22 Montrer qu'on peut construire un Diagramme de Venn pour tous les  $2^n$  sous-ensembles possibles de  $n$  ensembles donnés, en utilisant  $n$  polygones identiques, positionnés par rotation autour d'un même axe.
- 23 Supposons que Josèphe se trouve à une place imposée  $n$ , mais a la possibilité de choisir le nombre  $q$  qui sera le paramètre d'élimination : que chaque  $q$ ième personne sera exécutée. Peut-il sauver sa vie dans tous les cas ?

### Sujets de recherche

- 24 Trouver toutes les relations de récurrence de la forme

$$X_n = \frac{1 + a_1 X_{n-1} + \cdots + a_k X_{n-k}}{b_1 X_{n-1} + \cdots + b_k X_{n-k}}$$

dont la solution est périodique quelque soient les valeurs initiales  $X_0, \dots, X_{k-1}$ .

- 25 Résoudre une infinité de variantes du problème de la Tour de Hanoi en montrant qu'il y a égalité dans la relation de l'exercice 17.
- 26 Pour généraliser l'exercice 23, convenons qu'un *sous-ensemble de Josèphe* de  $\{1, 2, \dots, n\}$  est un ensemble de  $k$  nombres tel qu'il existe un  $q$  pour lequel les  $n - k$  autres numéros seront éliminés les premiers (ce sont donc les  $k$  places des "gentils" que Josèphe veut sauver). Il se trouve que, lorsque  $n = 9$ , trois parmi les  $2^9$  sous-ensembles possibles ne sont pas des sous-ensembles de Josèphe :  $\{1, 2, 5, 8, 9\}$ ,  $\{2, 3, 4, 5, 8\}$  et  $\{2, 5, 6, 7, 8\}$ . Il y en a 13 dans ce cas lorsque  $n = 12$ , et aucun pour toute autre valeur de  $n \leq 12$ . Lorsque  $n$  est grand, les sous-ensembles qui ne sont pas des sous-ensembles de Josèphe sont-ils rares ?

Oui, et bravo si vous les trouvez.

# 2

## Sommes

---

DANS TOUS LES DOMAINES des mathématiques, on se trouve confronté à des sommes. Il nous faut donc absolument des outils pour les manipuler. Ce chapitre présente donc des notations et des techniques qui rendent plus faciles les manipulations de sommes.

### 2.1 NOTATIONS

Nous avons déjà rencontré, dans le chapitre 1, la somme des  $n$  premiers entiers, et nous l'avons notée  $1 + 2 + 3 + \cdots + (n - 1) + n$ . Dans une formule de ce genre, les points de suspension indiquent que l'expression doit être complétée en fonction des premiers et des derniers termes de la somme. Bien entendu, il n'est pas facile de compléter des sommes du style de  $1 + 7 + \cdots + 41,7$ , qui n'ont aucun sens en dehors d'un contexte donné. D'un autre côté, les termes  $3$  et  $(n - 1)$  de la première somme sont superflus ; l'expression  $1 + 2 + \cdots + n$  est aussi explicite. Si on osait, on pourrait même écrire tout simplement  $1 + \cdots + n$ .

Nous allons considérer des sommes de la forme

$$a_1 + a_2 + \cdots + a_n , \tag{2.1}$$

où chaque  $a_k$  est un nombre que l'on suppose défini. L'avantage de cette notation, c'est qu'avec un peu d'imagination on peut "voir" la somme comme si elle avait été écrite entièrement.

Un élément  $a_k$  d'une somme est appelé un *terme*. Lorsque les termes sont spécifiés par des formules, ce qui est souvent le cas, il faut faire en sorte que leur signification soit perçue sans ambiguïté. C'est pourquoi il est parfois nécessaire de les écrire sous une forme développée. Par exemple, si

$$1 + 2 + \cdots + 2^{n-1}$$

est supposée être une somme de  $n$  termes plutôt que de  $2^{n-1}$  termes, il

*On n'est pas encore arrivé au terme de ce cours.*

vaudrait mieux l'écrire plus explicitement

$$2^0 + 2^1 + \cdots + 2^{n-1}.$$

Les points de suspension sont très utiles mais peuvent, d'une part être cause d'ambiguïtés, d'autre part alourdir les expressions. D'autres notations existent, la plus utilisée étant la "notation Sigma"

$$\sum_{k=1}^n a_k, \quad (2.2)$$

*"Le signe  $\sum_{i=1}^{i=\infty}$  indique que l'on doit donner au nombre entier i toutes ses valeurs 1, 2, 3, ..., et prendre la somme des termes."*

— J. Fourier [127]

ainsi appelée parce qu'elle est représentée par la lettre grecque  $\Sigma$  (sigma majuscule). Cette notation signifie qu'il faut inclure dans la somme exactement tous les termes  $a_k$  dont l'indice  $k$  est un entier compris entre les deux limites 1 et  $n$ . On dit que c'est une "somme pour  $k$  allant de 1 à  $n$ ". C'est Joseph Fourier qui a introduit en 1820 cette notation qui a rapidement bouleversé le monde des mathématiques.

La quantité écrite après  $\Sigma$  ( $a_k$  dans notre cas) s'appelle le *terme générique*. La variable d'indice  $k$  est dite *liée* au signe  $\Sigma$  dans (2.2), car le  $k$  de  $a_k$  est indépendant d'éventuelles occurrences de cette même lettre  $k$  en dehors de la notation Sigma. On pourrait remplacer  $k$  par  $n'$ importe quelle autre lettre sans changer le sens de (2.2). La lettre  $i$  est souvent utilisée pour cela (probablement pour "indice"), mais nous sommerons plutôt sur  $k$  dans ce livre car, traditionnellement,  $i$  désigne aussi la valeur  $\sqrt{-1}$ .

Il existe aussi une notation Sigma généralisée, plus utile encore que sa forme délimitée. On écrit simplement une ou plusieurs conditions sous le  $\Sigma$ , pour spécifier l'ensemble des indices sur lesquels on veut sommer. Par exemple, les sommes (2.1) et (2.2) peuvent aussi s'écrire

$$\sum_{\substack{1 \leq k \leq n}} a_k. \quad (2.3)$$

*Je ne prendrais quand même pas a ou n comme variable d'indice à la place de k, car ces lettres sont des variables "libres" qui ont ici un sens en dehors du  $\Sigma$ .*

Dans cet exemple particulier, il n'y a pas beaucoup de différence entre la nouvelle forme et celle de (2.2). Cependant la forme généralisée permet de sommer sur des ensemble d'indices qui ne sont pas forcément des entiers consécutifs. Par exemple, on peut écrire la somme des carrés de tous les entiers positifs impairs strictement inférieurs à 100 :

$$\sum_{\substack{1 \leq k < 100 \\ k \text{ impair}}} k^2.$$

L'équivalent délimité de cette somme,

$$\sum_{k=0}^{49} (2k+1)^2,$$

est plus lourd et bien moins clair. De façon similaire, la somme des inverses de tous les nombres premiers entre 1 et N s'écrit

$$\sum_{\substack{p \leq N \\ p \text{ premier}}} \frac{1}{p}.$$

Si nous voulions utiliser la forme délimitée, il faudrait écrire

$$\sum_{k=1}^{\pi(N)} \frac{1}{p_k},$$

où  $p_k$  est le  $k$ ème entier et  $\pi(N)$  est le nombre de nombres premiers inférieurs ou égaux à  $N$ . (Notons en passant que cette somme donne une approximation du nombre moyen de facteurs premiers distincts d'un entier aléatoire proche de  $N$ , car la probabilité que cet entier soit divisible par  $p$  est à peu près égale à  $1/p$ . La valeur de la somme lorsque  $N$  est grand est proche de  $\ln \ln N + M$ , où  $M \approx 0.2614972128476427837554268386086958590515666$  est la constante de Mertens [271],  $\ln x$  est le logarithme népérien (ou naturel) de  $x$  et  $\ln \ln x$  signifie  $\ln(\ln x)$ ).

L'avantage majeur de la notation Sigma généralisée est qu'elle peut être manipulée plus aisément que la forme délimitée. Par exemple, supposons que nous voulions changer la variable d'indice  $k$  en  $k+1$ . Si nous utilisons la forme générale, nous avons

$$\sum_{1 \leq k \leq n} a_k = \sum_{1 \leq k+1 \leq n} a_{k+1}.$$

On voit facilement ce qui se passe, et on peut faire cette substitution presque sans réfléchir. En revanche, avec la forme délimitée, on a

$$\sum_{k=1}^n a_k = \sum_{k=0}^{n-1} a_{k+1}.$$

Il est plus difficile de voir ce qui se passe, et donc plus facile de se tromper.

Malgré tout, la forme délimitée n'est pas totalement inutile. Elle est esthétique, explicite, et peut s'écrire vite, car (2.2) ne comprend que sept symboles tandis que (2.3) en contient huit. Par conséquent, pour poser un problème ou présenter des résultats, nous utiliserons souvent la forme délimitée ; en revanche, lorsqu'il nous faudra manipuler une somme en modifiant ses variables d'indice, nous utiliserons plutôt la forme générale.

Etant donné que le signe  $\sum$  apparaît plus de 1000 fois dans ce livre, nous devons être tout à fait sûrs de bien comprendre sa signification.

*Le symbole de sommation ressemble à un pacman tordu.*

*Une coquette somme, en somme.*

*Ce n'est rien. Essayez de voir combien il y a de  $\Sigma$  dans l'Iliade.*

Formellement,

$$\sum_{P(k)} a_k \quad (2.4)$$

est une abréviation pour la somme de tous les termes  $a_k$  tels que  $k$  est un entier satisfaisant une propriété donnée  $P(k)$  (une “propriété  $P(k)$ ” est une assertion sur  $k$ , qui peut être soit vraie, soit fausse). Pour le moment, nous supposerons que, parmi les entiers  $k$  qui satisfont  $P(k)$ , ceux pour lesquels  $a_k \neq 0$  sont en nombre fini. Dans le cas contraire, on additionne un nombre infini de nombres non nuls, ce qui peut poser des problèmes délicats. A l'autre extrême, si  $P(k)$  est faux pour tout entier  $k$ , la somme est “vide” et sa valeur est nulle par définition.

Si une somme doit apparaître dans le texte d'un paragraphe plutôt que dans une équation, nous utiliserons une version légèrement différente de (2.4) : nous écrirons “ $\sum_{P(k)} a_k$ ”, pour que la formule ne soit pas trop haute. De manière similaire, dans un texte, nous écrirons “ $\sum_{k=1}^n a_k$ ” à la place de (2.2).

On est souvent tenté d'écrire

$$\sum_{k=2}^{n-1} k(k-1)(n-k) \quad \text{au lieu de} \quad \sum_{k=0}^n k(k-1)(n-k)$$

puisque les termes correspondant à  $k = 0, 1$  et  $n$  sont nuls. A première vue, il semble en effet plus efficace d'additionner  $n - 2$  termes au lieu de  $n + 1$ . Cependant il ne faut pas céder à ce genre de tentation : un calcul efficace n'a rien à voir avec une compréhension efficace ! Il vaut mieux que les bornes d'une somme soient les plus simples possibles, car ainsi on peut manipuler la somme bien plus aisément. De plus, une forme comme  $\sum_{k=2}^{n-1}$  peut être dangereusement ambiguë, car sa signification n'est pas claire du tout lorsque  $n = 0$  ou  $n = 1$  (voir l'exercice 1). Les termes nuls ne causent aucun tort et permettent souvent d'éviter des problèmes.

Les notations que nous avons utilisées jusqu'ici sont tout à fait standard. Maintenant, nous allons sévèrement bousculer la tradition. Kenneth E. Iverson a eu une merveilleuse idée dans son langage de programmation APL [191, page 11; voir aussi 220], et nous allons voir qu'elle permet de bien simplifier beaucoup de ce que nous voulons faire dans ce livre. L'idée consiste simplement à mettre une proposition (vraie ou fausse) entre crochets, et dire que le résultat est 1 si la proposition est vraie, 0 sinon. Par exemple,

$$[p \text{ premier}] = \begin{cases} 1 & \text{si } p \text{ est un nombre premier;} \\ 0 & \text{sinon.} \end{cases}$$

La convention d'Iverson nous permet d'exprimer des sommes sans aucune contrainte sur l'indice de sommation. En effet, nous pouvons réécrire (2.4)

*En fait, le “symbole de Kronecker”, que j'ai vu dans d'autres livres ( $\delta_{kn} = 1$  si  $k = n$ , 0 sinon) est tout simplement un cas particulier de la convention d'Iverson : on peut écrire  $[k = n]$  à la place de  $\delta_{kn}$ .*

de la façon suivante :

$$\sum_k a_k [P(k)]. \quad (2.5)$$

Si  $P(k)$  est faux, alors le terme  $a_k [P(k)]$  est nul, donc nous pouvons l'inclure sans crainte dans les termes de la somme. Ceci permet de manipuler aisément l'index de sommation car on n'a pas à s'inquiéter des conditions aux bornes.

*"I am often surprised by new, important applications [of this notation]."*  
— B. de Finetti [123]

Il reste à mentionner un petit point technique :  $a_k$  n'est pas toujours défini pour tout entier  $k$ . On contourne cette difficulté en déclarant que  $[P(k)]$  est "très fortement nul" quand  $P(k)$  est faux : il est tellement nul que  $a_k [P(k)]$  est égal à zéro même si  $a_k$  n'est pas défini. Par exemple, si on utilise la convention d'Iverson pour écrire la somme des inverses des nombres premiers  $\leq N$

$$\sum_p [p \text{ premier}] [p \leq N] / p,$$

il n'y a aucun problème de division par zéro quand  $p = 0$  car, selon la convention que nous nous sommes donnée,  $[0 \text{ premier}] [0 \leq N] / 0 = 0$ .

Résumons ce que nous avons appris jusqu'à présent sur les sommes. Il y a deux bonnes façons d'écrire une somme de termes : l'une utilise "..." tandis que l'autre utilise " $\sum$ ". La première forme peut souvent suggérer des manipulations utiles, par exemple combiner des termes adjacents, car il est plus facile de repérer une simplification possible lorsqu'on a l'intégralité de la somme devant les yeux. Cependant, il peut aussi arriver qu'une trop grande abondance de détails nuise. La notation Sigma est compacte, impressionne la famille et les amis, et peut souvent suggérer des manipulations qui ne seraient pas évidentes sous la forme "points de suspension". Quand on travaille avec la notation Sigma, les termes nuls sont rarement nuisibles ; en fait, il arrive souvent qu'ils facilitent les manipulations.

## 2.2 SOMMES ET RÉCURRENCES

Nous connaissons donc un certain nombre de notations différentes pour exprimer des sommes. Maintenant, la question est : comment calculer la valeur d'une somme ? On peut remarquer qu'il existe une étroite relation entre les sommes et les récurrences. La somme

$$S_n = \sum_{k=0}^n a_k$$

*...et offre moins de risques d'être collé à un examen pour "manque de rigueur".*

est équivalente à la récurrence

$$\begin{aligned} S_0 &= a_0; \\ S_n &= S_{n-1} + a_n, \quad \text{pour } n > 0. \end{aligned} \tag{2.6}$$

*Pensez à  $S_n$ , non comme un simple nombre, mais comme une suite définie pour tout  $n \geq 0$ .*

On peut donc trouver une expression exacte d'une somme en utilisant les méthodes du chapitre 1.

Par exemple, si  $a_n$  est égal à une constante plus un multiple de  $n$ , la récurrence (2.6) prend la forme générale

$$\begin{aligned} R_0 &= \alpha; \\ R_n &= R_{n-1} + \beta + \gamma n, \quad \text{pour } n > 0. \end{aligned} \tag{2.7}$$

En procédant comme au chapitre 1, on trouve  $R_1 = \alpha + \beta + \gamma$ ,  $R_2 = \alpha + 2\beta + 3\gamma$  etc. La solution générale s'écrit

$$R_n = A(n)\alpha + B(n)\beta + C(n)\gamma, \tag{2.8}$$

où  $A(n)$ ,  $B(n)$  et  $C(n)$  sont des coefficients dépendant des paramètres  $\alpha$ ,  $\beta$  et  $\gamma$ .

La méthode du répertoire nous suggère d'essayer de remplacer  $R_n$  par des fonctions simples de  $n$ , en espérant trouver des paramètres  $\alpha$ ,  $\beta$  et  $\gamma$  pour lesquels la solution est particulièrement simple. Si  $R_n = 1$ , alors  $\alpha = 1$ ,  $\beta = 0$  et  $\gamma = 0$  ; donc

$$A(n) = 1.$$

En posant  $R_n = n$ , on a  $\alpha = 0$ ,  $\beta = 1$  et  $\gamma = 0$  ; donc

$$B(n) = n.$$

Enfin, si  $R_n = n^2$ , alors  $\alpha = 0$ ,  $\beta = -1$  et  $\gamma = 2$  ; donc

$$2C(n) - B(n) = n^2$$

et on obtient  $C(n) = (n^2 + n)/2$ . C'est du gâteau.

Donc, si l'on veut calculer

$$\sum_{k=0}^n (a + bk),$$

*En voici une plus facile :  $\pi = \sum_{n \geq 0} \frac{8}{(4n+1)(4n+3)}$ .*

la récurrence (2.6) équivaut à (2.7) avec  $\alpha = \beta = a$ ,  $\gamma = b$  et le résultat est  $aA(n) + aB(n) + bC(n) = a(n+1) + b(n+1)n/2$ .

Inversement, beaucoup de récurrences peuvent se réduire à des sommes. Par conséquent, les méthodes spécifiques de calcul de sommes que nous verrons plus loin dans ce chapitre pourront nous aider à résoudre des récurrences qui, sans cela, auraient été difficiles. La récurrence de la Tour de Hanoi en est un bon exemple :

$$\begin{aligned} T_0 &= 0; \\ T_n &= 2T_{n-1} + 1, \quad \text{pour } n > 0. \end{aligned}$$

En divisant les deux membres de chaque égalité par  $2^n$ , on peut écrire cette récurrence sous la forme (2.6) :

$$\begin{aligned} T_0/2^0 &= 0; \\ T_n/2^n &= T_{n-1}/2^{n-1} + 1/2^n, \quad \text{pour } n > 0. \end{aligned}$$

Posons maintenant  $S_n = T_n/2^n$  pour obtenir

$$\begin{aligned} S_0 &= 0; \\ S_n &= S_{n-1} + 2^{-n}, \quad \text{pour } n > 0. \end{aligned}$$

Il s'ensuit que

$$S_n = \sum_{k=1}^n 2^{-k}.$$

(Remarquez que nous avons ôté de cette somme le terme correspondant à  $k = 0$ ). Nous verrons un peu plus loin que la somme de la série géométrique  $2^{-1} + 2^{-2} + \dots + 2^{-n} = (\frac{1}{2})^1 + (\frac{1}{2})^2 + \dots + (\frac{1}{2})^n$  est égale à  $1 - (\frac{1}{2})^n$ . Il s'ensuit que  $T_n = 2^n S_n = 2^n - 1$ .

Au cours de ce calcul, nous sommes passés de  $S_n$  à  $T_n$  en remarquant que la récurrence pouvait être divisée par  $2^n$ . Cette astuce est un cas particulier d'une technique générale permettant de transformer toute récurrence de la forme

$$a_n T_n = b_n T_{n-1} + c_n \tag{2.9}$$

en une somme. L'idée consiste à multiplier les deux membres par un *facteur de sommation*  $s_n$  :

$$s_n a_n T_n = s_n b_n T_{n-1} + s_n c_n.$$

Ce facteur  $s_n$  doit être judicieusement choisi pour que

$$s_n b_n = s_{n-1} a_{n-1}.$$

### 30 SOMMES

Alors, si l'on écrit  $S_n = s_n a_n T_n$ , on obtient la récurrence

$$S_n = S_{n-1} + s_n c_n.$$

Donc

$$S_n = s_0 a_0 T_0 + \sum_{k=1}^n s_k c_k = s_1 b_1 T_0 + \sum_{k=1}^n s_k c_k,$$

et la solution de la récurrence de départ (2.9) est

$$T_n = \frac{1}{s_n a_n} \left( s_1 b_1 T_0 + \sum_{k=1}^n s_k c_k \right). \quad (2.10)$$

Par exemple, lorsque  $n = 1$  on obtient  $T_1 = (s_1 b_1 T_0 + s_1 c_1) / s_1 a_1 = (b_1 T_0 + c_1) / a_1$ .

Au fait, comment trouver le bon  $s_n$ ? Pas de problème : il suffit de développer la relation  $s_n = s_{n-1} a_{n-1} / b_n$  de sorte que la fraction

$$s_n = \frac{a_{n-1} a_{n-2} \dots a_1}{b_n b_{n-1} \dots b_2}, \quad (2.11)$$

$s_1$  disparaît, donc il peut prendre n'importe quelle valeur sauf zéro.

ou un quelconque multiple de cette valeur, soit un facteur de sommation adéquat. Par exemple, dans la récurrence de la Tour de Hanoi, nous avons  $a_n = 1$  et  $b_n = 2$ . D'après la méthode générale que nous venons de décrire, il serait bon de multiplier par  $s_n = 2^{-n}$  pour réduire la récurrence à une somme. Nul besoin d'une brillante inspiration pour trouver ce facteur.

Il faut cependant faire attention, comme toujours, de ne pas effectuer de division par zéro. La méthode du facteur de sommation marche pourvu que les  $a_k$  et les  $b_k$  soient tous différents de zéro.

Appliquons maintenant ces idées à une récurrence qui apparaît dans l'étude du "tri rapide" (quicksort), l'une des plus importantes méthodes de tri de données à l'aide d'un ordinateur. Le nombre moyen de comparaisons effectuées par le tri rapide lorsqu'il est appliqué à un ensemble de  $n$  objets ordonnés au hasard satisfait la récurrence

(Le tri rapide a été développé par Hoare en 1962 [189].)

$$C_0 = 0;$$

$$C_n = n + 1 + \frac{2}{n} \sum_{k=0}^{n-1} C_k, \quad \text{pour } n > 0. \quad (2.12)$$

Hum ! A première vue, c'est bien plus épouvantable que les récurrences que nous avons vues jusqu'à présent ; il y a là une somme de toutes les valeurs précédentes, plus une division par  $n$ . Si on prend des petits cas, on trouve bien des résultats ( $C_1 = 2$ ,  $C_2 = 5$ ,  $C_3 = \frac{26}{3}$ ), mais ils ne sont pas franchement rassurants.

Malgré tout, on peut diminuer systématiquement la complexité de (2.12), d'abord en se débarrassant de la division, puis du signe  $\sum$ . L'idée est de multiplier les deux membres par  $n$  pour obtenir la relation

$$nC_n = n^2 + n + 2 \sum_{k=0}^{n-1} C_k, \quad \text{pour } n > 0,$$

puis, en remplaçant  $n$  par  $n - 1$ ,

$$(n-1)C_{n-1} = (n-1)^2 + (n-1) + 2 \sum_{k=0}^{n-2} C_k, \quad \text{pour } n-1 > 0.$$

Maintenant on peut soustraire la seconde équation de la première, et ainsi faire disparaître le signe  $\sum$  :

$$nC_n - (n-1)C_{n-1} = 2n + 2C_{n-1}, \quad \text{pour } n > 1.$$

On remarque que cette relation est valable aussi pour  $n = 1$ , car  $C_1 = 2$ . On arrive ainsi à une récurrence pour  $C_n$  bien plus simple que celle d'origine :

$$\begin{aligned} C_0 &= 0; \\ nC_n &= (n+1)C_{n-1} + 2n, \quad \text{pour } n > 0. \end{aligned}$$

C'est déjà un progrès. Nous pouvons maintenant utiliser un facteur de sommation, car cette récurrence est de la même forme que (2.9), avec  $a_n = n$ ,  $b_n = n+1$  et  $c_n = 2n$ . La méthode décrite plus haut nous indique qu'il faut multiplier la récurrence par un multiple de

$$s_n = \frac{a_{n-1}a_{n-2}\dots a_1}{b_nb_{n-1}\dots b_2} = \frac{(n-1)\cdot(n-2)\cdot\dots\cdot 1}{(n+1)\cdot n\cdot\dots\cdot 3} = \frac{2}{(n+1)n}.$$

*Nous sommes partis avec un  $\sum$  dans une récurrence, et nous avons travaillé dur pour pour nous en débarrasser. Et après avoir utilisé un facteur de sommation, on retrouve un  $\sum$ . Alors, on veut s'en débarrasser ou pas ? Il faudrait savoir !*

Voici donc la solution, donnée par (2.10) :

$$C_n = 2(n+1) \sum_{k=1}^n \frac{1}{k+1}.$$

La somme qui reste ressemble beaucoup à une quantité que l'on rencontre souvent dans des applications diverses. Elle apparaît même si souvent qu'on lui a donné un nom et une notation spécifiques :

$$H_n = 1 + \frac{1}{2} + \dots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k}. \tag{2.13}$$

La lettre  $H$  signifie "harmonique" ;  $H_n$  est un *nombre harmonique*, ainsi appelé car la  $k$ ième harmonique produite par une corde de violon est aussi

la note fondamentale produite par une corde  $k$  fois plus courte (donc dont la longueur est  $1/k$  fois celle de la première corde).

Pour terminer notre étude de la récurrence du tri rapide (2.12), nous allons trouver une formule close pour  $C_n$ . Ce sera fait si nous arrivons à exprimer  $C_n$  en fonction de  $H_n$ . Dans notre formule pour  $C_n$ , la somme s'écrit

$$\sum_{k=1}^n \frac{1}{k+1} = \sum_{1 \leq k \leq n} \frac{1}{k+1}.$$

On peut la lier à  $H_n$  sans trop de difficultés en remplaçant  $k$  par  $k+1$  et en modifiant les conditions limites :

$$\begin{aligned} \sum_{1 \leq k \leq n} \frac{1}{k+1} &= \sum_{1 \leq k-1 \leq n} \frac{1}{k} = \sum_{2 \leq k \leq n+1} \frac{1}{k} \\ &= \left( \sum_{1 \leq k \leq n} \frac{1}{k} \right) - \frac{1}{1} + \frac{1}{n+1} = H_n - \frac{n}{n+1}. \end{aligned}$$

Parfait ! Nous avons trouvé la solution de (2.12) : le nombre moyen de comparaisons effectuées par le tri rapide quand il est appliqué à un ensemble de  $n$  données réparties au hasard est

$$C_n = 2(n+1)H_n - 2n. \quad (2.14)$$

Vérifions, comme d'habitude, que les petits cas se conduisent correctement :  $C_0 = 0$ ,  $C_1 = 2$ ,  $C_2 = 5$ .

## 2.3 MANIPULATION DE SOMMES

*A ne pas confondre avec la finance.*

Avec les sommes, la clé du succès est la capacité de transformer un  $\sum$  en un autre plus simple, ou bien plus proche d'un certain but. Pour cela, il suffit d'apprendre quelques règles de transformation de base et de s'entraîner à les utiliser.

Soit  $K$  un ensemble fini d'entiers. Pour transformer les sommes sur les éléments de  $K$ , on peut utiliser trois règles simples :

$$\sum_{k \in K} c a_k = c \sum_{k \in K} a_k; \quad (\text{distributivité}) \quad (2.15)$$

$$\sum_{k \in K} (a_k + b_k) = \sum_{k \in K} a_k + \sum_{k \in K} b_k; \quad (\text{associativité}) \quad (2.16)$$

$$\sum_{k \in K} a_k = \sum_{p(k) \in K} a_{p(k)}. \quad (\text{commutativité}) \quad (2.17)$$

Pourquoi ne pas l'appeler "perméativité" plutôt que "commutativité" ?

La distributivité nous permet de déplacer les constantes à l'intérieur ou à l'extérieur d'un  $\sum$ . L'associativité autorise à partager un  $\sum$  en deux ou à joindre deux  $\sum$  en un seul. Quant à la commutativité, elle indique qu'on peut modifier l'ordre des termes comme on le désire ; ici,  $p(k)$  est une permutation quelconque de l'ensemble des entiers. Par exemple, si  $K = \{-1, 0, +1\}$  et  $p(k) = -k$ , ces trois règles donnent respectivement

$$\begin{aligned} ca_{-1} + ca_0 + ca_1 &= c(a_{-1} + a_0 + a_1); && \text{(distributivité)} \\ (a_{-1} + b_{-1}) + (a_0 + b_0) + (a_1 + b_1) &= (a_{-1} + a_0 + a_1) + (b_{-1} + b_0 + b_1); && \text{(associativité)} \\ a_{-1} + a_0 + a_1 &= a_1 + a_0 + a_{-1}. && \text{(commutativité)} \end{aligned}$$

L'astuce de Gauss du chapitre 1 peut être vue comme une application de ces trois règles de base. Supposons que nous voulons calculer la somme générale d'une *progression arithmétique*,

$$S = \sum_{0 \leq k \leq n} (a + bk).$$

C'est comme un changement de variable dans une intégrale, en plus facile.

La commutativité nous autorise à remplacer  $k$  par  $n - k$  pour obtenir

$$S = \sum_{0 \leq n-k \leq n} (a + b(n - k)) = \sum_{0 \leq k \leq n} (a + bn - bk).$$

On peut additionner ces deux équations en utilisant l'associativité :

$$2S = \sum_{0 \leq k \leq n} ((a + bk) + (a + bn - bk)) = \sum_{0 \leq k \leq n} (2a + bn).$$

Maintenant, on applique la distributivité et on calcule une somme triviale :

"What's one and one?"

"I don't know," said Alice.

"I lost count."

"She can't do Addition."

— Lewis Carroll [50]

$$2S = (2a + bn) \sum_{0 \leq k \leq n} 1 = (2a + bn)(n + 1).$$

En divisant par 2, nous démontrons que

$$\sum_{k=0}^n (a + bk) = (a + \frac{1}{2}bn)(n + 1). \quad (2.18)$$

Remarquons que le membre droit est égal à la moyenne du premier et du dernier terme,  $\frac{1}{2}(a + (a + bn))$ , multipliée par le nombre de termes, soit  $(n + 1)$ . C'est un bon moyen de s'en souvenir.

Il est important de garder en tête le fait que la fonction  $p(k)$  de la définition de la commutativité (2.17) est une permutation de tout l'ensemble

### 34 SOMMES

des entiers. En d'autres termes, pour tout entier  $n$  il existe exactement un entier  $k$  tel que  $p(k) = n$ . Autrement, la commutativité serait prise en défaut ; l'exercice 3 illustre bien cela. Les transformations comme  $p(k) = k + c$  ou  $p(k) = c - k$ , où  $c$  est une constante entière, qui sont toujours des permutations, marchent sans aucun problème.

D'un autre côté, on peut alléger un peu la restriction concernant la permutation : il suffit en fait qu'il existe exactement un entier  $k$  tel que  $p(k) = n$  lorsque  $n$  appartient à l'ensemble d'indices  $K$ . Si  $n \notin K$  (si  $n$  n'appartient pas à  $K$ ), le nombre d'entiers  $k$  tels que  $p(k) = n$  n'a aucune importance puisque ces entiers n'interviennent pas dans la somme. Ainsi, par exemple, nous pouvons soutenir que

$$\sum_{\substack{k \in K \\ k \text{ pair}}} a_k = \sum_{\substack{n \in K \\ n \text{ pair}}} a_n = \sum_{\substack{2k \in K \\ 2k \text{ pair}}} a_{2k} = \sum_{2k \in K} a_{2k}, \quad (2.19)$$

car il existe un unique  $k$  tel que  $2k = n$  quand  $n \in K$  et  $n$  est pair.

La convention d'Iverson, qui nous permet d'obtenir les valeurs 0 ou 1 en mettant des propositions logiques entre crochets, peut être combinée avec les propriétés de distributivité, d'associativité et de commutativité pour déduire de nouvelles propriétés. Voici par exemple une règle importante pour combiner différents ensembles d'indices. Si  $K$  et  $K'$  sont des ensembles d'entiers, alors

$$\sum_{k \in K} a_k + \sum_{k \in K'} a_k = \sum_{k \in K \cap K'} a_k + \sum_{k \in K \cup K'} a_k. \quad (2.20)$$

Cela découle des formules générales

$$\sum_{k \in K} a_k = \sum_k a_k [k \in K] \quad (2.21)$$

et

$$[k \in K] + [k \in K'] = [k \in K \cap K'] + [k \in K \cup K']. \quad (2.22)$$

En général, on utilise la règle (2.20), soit pour combiner deux ensemble d'indices presque disjoints, comme ceci :

$$\sum_{k=1}^m a_k + \sum_{k=m}^n a_k = a_m + \sum_{k=1}^n a_k, \quad \text{pour } 1 \leq m \leq n;$$

soit pour détacher un unique terme de la somme :

$$\sum_{0 \leq k \leq n} a_k = a_0 + \sum_{1 \leq k \leq n} a_k, \quad \text{pour } n \geq 0. \quad (2.23) \quad \begin{array}{l} \text{(Ici, on a permute} \\ \text{les deux membres de} \\ \text{(2.20).)} \end{array}$$

Cette opération de détachement d'un terme est la base d'une *méthode de perturbation* qui permet souvent de trouver une formule close pour une somme. L'idée consiste à partir d'une somme inconnue et de l'appeler  $S_n$  :

$$S_n = \sum_{0 \leq k \leq n} a_k.$$

(Nommer pour régner). Puis on réécrit  $S_{n+1}$  de deux façons différentes, en détachant son dernier et son premier termes :

$$\begin{aligned} S_n + a_{n+1} &= \sum_{0 \leq k \leq n+1} a_k = a_0 + \sum_{1 \leq k \leq n+1} a_k \\ &= a_0 + \sum_{1 \leq k+1 \leq n+1} a_{k+1} \\ &= a_0 + \sum_{0 \leq k \leq n} a_{k+1}. \end{aligned} \tag{2.24}$$

Nous pouvons maintenant travailler sur cette dernière somme pour essayer de l'exprimer en fonction de  $S_n$ . Si nous y arrivons, nous obtiendrons une équation dont la solution sera la somme que nous cherchons.

Par exemple, utilisons cette approche pour trouver la somme d'une progression géométrique générale,

*Si c'est géométrique,  
on doit pouvoir  
trouver une preuve  
géométrique.*

$$S_n = \sum_{0 \leq k \leq n} ax^k.$$

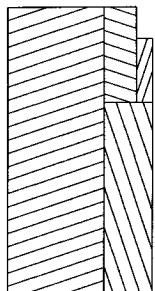
La règle de perturbation (2.24) nous indique que

$$S_n + ax^{n+1} = ax^0 + \sum_{0 \leq k \leq n} ax^{k+1},$$

et, en appliquant la distributivité, on trouve que le terme de droite est  $x \sum_{0 \leq k \leq n} ax^k = xS_n$ . Donc  $S_n + ax^{n+1} = a + xS_n$ , et en résolvant cette équation en  $S_n$  on obtient

$$\sum_{k=0}^n ax^k = \frac{a - ax^{n+1}}{1 - x}, \quad \text{pour } x \neq 1. \tag{2.25}$$

*Ah oui, j'ai appris  
cette formule au  
lycée.*



(Si  $x = 1$ , la somme est évidemment égale à  $(n+1)a$ ). On peut mémoriser le membre droit en remarquant que c'est la différence entre le premier terme inclus dans la somme et le premier terme qui en est exclu (le terme juste après le dernier), divisée par 1 moins la raison de la progression.

C'était presque trop facile. Essayons d'utiliser cette même technique de perturbation sur une somme un peu plus difficile,

$$S_n = \sum_{0 \leq k \leq n} k 2^k.$$

Dans ce cas, on a  $S_0 = 0$ ,  $S_1 = 2$ ,  $S_2 = 10$ ,  $S_3 = 34$ ,  $S_4 = 98$ . Quelle est la formule générale ? D'après (2.24) on a

$$S_n + (n+1)2^{n+1} = \sum_{0 \leq k \leq n} (k+1)2^{k+1},$$

et il nous faut exprimer le membre droit en fonction de  $S_n$ . En utilisant l'associativité, on peut toujours le séparer en deux sommes,

$$\sum_{0 \leq k \leq n} k2^{k+1} + \sum_{0 \leq k \leq n} 2^{k+1},$$

dont la première est égale à  $2S_n$ . L'autre somme est une progression géométrique, qui est égale à  $(2 - 2^{n+2})/(1 - 2) = 2^{n+2} - 2$  d'après (2.25). Par conséquent,  $S_n + (n+1)2^{n+1} = 2S_n + 2^{n+2} - 2$ , et on obtient

$$\sum_{0 \leq k \leq n} k2^k = (n-1)2^{n+1} + 2.$$

Maintenant on comprend pourquoi  $S_3 = 34$  : c'est  $32 + 2$ , et non pas  $2 \cdot 17$ .

En effectuant le même travail avec  $x$  à la place de 2, on trouverait l'équation  $S_n + (n+1)x^{n+1} = xS_n + (x - x^{n+2})/(1-x)$ . On en déduit donc que

$$\sum_{k=0}^n kx^k = \frac{x - (n+1)x^{n+1} + nx^{n+2}}{(1-x)^2}, \quad \text{pour } x \neq 1. \quad (2.26)$$

Il est intéressant de remarquer qu'on aurait pu obtenir cette formule close de manière complètement différente, avec des techniques élémentaires de calcul différentiel. En partant de l'équation

$$\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x}$$

et en dérivant des deux côtés par rapport à  $x$ , on obtient

$$\sum_{k=0}^n kx^{k-1} = \frac{(1-x)(-(n+1)x^n) + 1 - x^{n+1}}{(1-x)^2} = \frac{1 - (n+1)x^n + nx^{n+1}}{(1-x)^2},$$

du fait que la dérivée d'une somme est la somme des dérivées de ses termes. Nous verrons dans les chapitres suivants qu'il existe de nombreux liens entre le calcul infinitésimal et les mathématiques discrètes.

## 2.4 SOMMES MULTIPLES

Il est possible de spécifier les termes d'une somme par deux ou plusieurs indices, au lieu d'un seulement. Voici par exemple une somme

double de neuf termes, indicés par deux variables  $j$  et  $k$  :

$$\sum_{\substack{1 \leq j, k \leq 3}} a_j b_k = a_1 b_1 + a_1 b_2 + a_1 b_3 \\ + a_2 b_1 + a_2 b_2 + a_2 b_3 \\ + a_3 b_1 + a_3 b_2 + a_3 b_3.$$

Notez qu'on ne somme pas sur tout  $j \geq 1$  et tout  $k \leq 3$ .

On utilise pour ce genre de sommes les mêmes notations et méthodes que pour les sommes à un seul indice. Ainsi, si  $P(j, k)$  est une propriété de  $j$  et  $k$ , la somme de tous les termes  $a_{j,k}$  tels que  $P(j, k)$  est vraie peut s'écrire de deux façons. L'une d'elles utilise la convention d'Iverson et somme sur toutes les paires d'entiers  $j$  et  $k$  :

$$\sum_{P(j,k)} a_{j,k} = \sum_{j,k} a_{j,k} [P(j, k)].$$

Un seul signe  $\sum$  est nécessaire, bien qu'il y ait plus d'un indice de sommation.  $\sum$  désigne une somme sur toutes les combinaisons possibles des indices.

On utilise deux  $\sum$  lorsque on est en présence d'une somme de sommes. Par exemple,

$$\sum_j \sum_k a_{j,k} [P(j, k)]$$

est une abréviation de

$$\sum_j \left( \sum_k a_{j,k} [P(j, k)] \right),$$

Les  $\Sigma$  doivent être évalués de droite à gauche (à l'envers).

qui désigne la somme, sur tous les entiers  $j$ , de  $\sum_k a_{j,k} [P(j, k)]$ , cette dernière somme étant la somme, sur tous les entiers  $k$ , des termes  $a_{j,k}$  pour lesquels  $P(j, k)$  est vraie. Dans de tels cas, on dit que la somme double est "sommée d'abord sur  $k$ ". Une somme qui dépend de plus d'un indice peut être sommée d'abord sur n'importe lequel de ses indices.

Pour ce faire, il existe une loi de base, appelée *changement de l'ordre de sommation*, qui généralise la règle d'associativité (2.16) vue précédemment :

$$\sum_j \sum_k a_{j,k} [P(j, k)] = \sum_{P(j,k)} a_{j,k} = \sum_k \sum_j a_{j,k} [P(j, k)]. \quad (2.27)$$

Le membre du milieu est une somme sur deux indices. A gauche,  $\sum_j \sum_k$  signifie que l'on somme d'abord sur  $k$ , puis sur  $j$ . A l'inverse, dans le membre de droite,  $\sum_k \sum_j$  doit être sommé d'abord sur  $j$ , puis sur  $k$ . En pratique, quand on cherche une formule close pour une somme double, il y

a généralement un indice préférable à l'autre pour débuter la sommation ; il faut donc choisir cet indice avec soin.

Il n'y a pas de raison de paniquer devant une somme de sommes, bien qu'il soit naturel que les débutants soient quelque peu déroutés. Regardons donc encore quelques exemples. Nous allons illustrer les manipulations de sommes doubles en travaillant sur la somme de neuf termes présentée plus haut. Elle constitue un bon exemple, car elle peut effectivement être simplifiée, et le processus de simplification est typique de ce qu'on peut faire avec les  $\sum \sum$  :

$$\begin{aligned}
\sum_{1 \leq j, k \leq 3} a_j b_k &= \sum_{j, k} a_j b_k [1 \leq j, k \leq 3] = \sum_{j, k} a_j b_k [1 \leq j \leq 3][1 \leq k \leq 3] \\
&= \sum_j \sum_k a_j b_k [1 \leq j \leq 3][1 \leq k \leq 3] \\
&= \sum_j a_j [1 \leq j \leq 3] \sum_k b_k [1 \leq k \leq 3] \\
&= \sum_j a_j [1 \leq j \leq 3] \left( \sum_k b_k [1 \leq k \leq 3] \right) \\
&= \left( \sum_j a_j [1 \leq j \leq 3] \right) \left( \sum_k b_k [1 \leq k \leq 3] \right) \\
&= \left( \sum_{j=1}^3 a_j \right) \left( \sum_{k=1}^3 b_k \right).
\end{aligned}$$

Paniquer ? Je trouve cette règle assez évidente par rapport à certains passages du chapitre 1.

Ici, la première ligne désigne une somme de neuf termes sans ordre particulier. La seconde ligne les groupe par trois,  $(a_1 b_1 + a_1 b_2 + a_1 b_3) + (a_2 b_1 + a_2 b_2 + a_2 b_3) + (a_3 b_1 + a_3 b_2 + a_3 b_3)$ . Dans la troisième ligne, on utilise la distributivité pour factoriser les  $a$ , car  $a_j$  et  $[1 \leq j \leq 3]$  ne dépendent pas de  $k$  ; on obtient  $a_1(b_1 + b_2 + b_3) + a_2(b_1 + b_2 + b_3) + a_3(b_1 + b_2 + b_3)$ . La quatrième ligne est semblable à la troisième, mais on y a ajouté des parenthèses redondantes pour que la cinquième ligne ne paraisse pas trop mystérieuse. La cinquième ligne factorise les  $(b_1 + b_2 + b_3)$  qui apparaissent pour chaque valeur de  $j$  :  $(a_1 + a_2 + a_3)(b_1 + b_2 + b_3)$ . La dernière ligne est juste une façon différente d'écrire la précédente. On peut utiliser cette méthode de transformation pour prouver une *règle générale de distributivité*,

$$\sum_{\substack{j \in J \\ k \in K}} a_j b_k = \left( \sum_{j \in J} a_j \right) \left( \sum_{k \in K} b_k \right), \tag{2.28}$$

valable quels que soient les ensembles d'indices  $J$  et  $K$ .

Il existe beaucoup de variantes de la loi de base (2.27) permettant d'échanger l'ordre d'une sommation. On les utilise lorsqu'on veut restreindre la portée des indices pour éviter de sommer sur tous les entiers

$j$  et  $k$ . Ces variantes existent en deux versions : la simple et la moins simple. Voyons d'abord la première :

$$\sum_{j \in J} \sum_{k \in K} a_{j,k} = \sum_{\substack{j \in J \\ k \in K}} a_{j,k} = \sum_{k \in K} \sum_{j \in J} a_{j,k}. \quad (2.29)$$

C'est tout simplement une façon différente d'écrire (2.27), car l'Iversonien  $[j \in J, k \in K]$  se factorise en  $[j \in J][k \in K]$ . Cette règle peut être appliquée chaque fois que les portées de  $j$  et de  $k$  sont mutuellement indépendantes.

La formule moins simple est un peu plus technique. On peut l'appliquer quand la portée d'une somme intérieure dépend de la variable d'indice de la somme extérieure :

$$\sum_{j \in J} \sum_{k \in K(j)} a_{j,k} = \sum_{k \in K'} \sum_{j \in J'(k)} a_{j,k}. \quad (2.30)$$

Ici, les ensembles  $J$ ,  $K(j)$ ,  $K'$  et  $J'(k)$  sont liés de sorte que

$$[j \in J][k \in K(j)] = [k \in K'][j \in J'(k)].$$

Une factorisation comme celle-ci est toujours possible en principe, car on peut considérer que  $J = K'$  est l'ensemble de tous les entiers et  $K(j) = J'(k)$  est la propriété  $P(j, k)$  qui gouverne une somme double. Cependant, il existe des cas spécifiques importants pour lesquels  $J$ ,  $K(j)$ ,  $K'$  et  $J'(k)$  s'expriment simplement. Cela arrive dans beaucoup d'applications. Voici par exemple une factorisation particulièrement utile :

$$[1 \leq j \leq n][j \leq k \leq n] = [1 \leq j \leq k \leq n] = [1 \leq k \leq n][1 \leq j \leq k]. \quad (2.31)$$

D'après cette "équation iversonienne", on peut écrire

$$\sum_{j=1}^n \sum_{k=j}^n a_{j,k} = \sum_{1 \leq j \leq k \leq n} a_{j,k} = \sum_{k=1}^n \sum_{j=1}^k a_{j,k}. \quad (2.32)$$

Généralement, l'une de ces deux sommes doubles est plus facile à calculer que l'autre. Avec (2.32), on peut choisir celle qui convient le mieux.

Appliquons ces idées à un exemple concret. Considérons le tableau de  $n^2$  produits

$$\begin{bmatrix} a_1 a_1 & a_1 a_2 & a_1 a_3 & \dots & a_1 a_n \\ a_2 a_1 & a_2 a_2 & a_2 a_3 & \dots & a_2 a_n \\ a_3 a_1 & a_3 a_2 & a_3 a_3 & \dots & a_3 a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n a_1 & a_n a_2 & a_n a_3 & \dots & a_n a_n \end{bmatrix}.$$

(C'est le moment de faire les exercices d'échauffement 4 et 6).

(Ou d'aller voir s'il reste quelque chose à grignoter dans le frigo).

Notre but est de trouver une formule simple pour

$$S_{\nabla} = \sum_{1 \leq j \leq k \leq n} a_j a_k ,$$

la somme de tous les éléments situés dans ou au dessus de la diagonale principale du tableau. Comme  $a_j a_k = a_k a_j$ , le tableau est symétrique par rapport à sa première diagonale ; donc  $S_{\nabla}$  doit être à peu près égal à la moitié de la somme de *tous* les éléments (sauf un petit quelque chose à supprimer pour prendre en compte la diagonale principale).

Ces considérations suggèrent les manipulations suivantes. On a

$$S_{\nabla} = \sum_{1 \leq j \leq k \leq n} a_j a_k = \sum_{1 \leq k \leq j \leq n} a_k a_j = \sum_{1 \leq k \leq j \leq n} a_j a_k = S_{\Delta} ,$$

car on peut remplacer  $(j, k)$  par  $(k, j)$ . De plus, comme

$$[1 \leq j \leq k \leq n] + [1 \leq k \leq j \leq n] = [1 \leq j, k \leq n] + [1 \leq j = k \leq n] ,$$

on a

$$2S_{\nabla} = S_{\nabla} + S_{\Delta} = \sum_{1 \leq j, k \leq n} a_j a_k + \sum_{1 \leq j = k \leq n} a_j a_k .$$

D'après la règle de distributivité (2.28), la première somme est égale à  $(\sum_{j=1}^n a_j)(\sum_{k=1}^n a_k) = (\sum_{k=1}^n a_k)^2$ . La seconde est égale à  $\sum_{k=1}^n a_k^2$ . On obtient donc une expression de la somme du triangle supérieur en fonction de sommes plus simples :

$$S_{\nabla} = \sum_{1 \leq j \leq k \leq n} a_j a_k = \frac{1}{2} \left( \left( \sum_{k=1}^n a_k \right)^2 + \sum_{k=1}^n a_k^2 \right) , \quad (2.33)$$

Encouragés par ce succès, regardons une autre somme double :

$$S = \sum_{1 \leq j < k \leq n} (a_k - a_j)(b_k - b_j) .$$

Il y a de nouveau une symétrie lorsqu'on permute  $j$  et  $k$  :

$$S = \sum_{1 \leq k < j \leq n} (a_j - a_k)(b_j - b_k) = \sum_{1 \leq k < j \leq n} (a_k - a_j)(b_k - b_j) .$$

Ainsi, en utilisant l'identité

$$[1 \leq j < k \leq n] + [1 \leq k < j \leq n] = [1 \leq j, k \leq n] - [1 \leq j = k \leq n]$$

on peut ajouter  $S$  à elle-même pour conclure que

$$2S = \sum_{1 \leq j, k \leq n} (a_j - a_k)(b_j - b_k) - \sum_{1 \leq j=k \leq n} (a_j - a_k)(b_j - b_k).$$

La seconde somme est nulle ; qu'en est-il de la première ? Elle peut se développer en quatre sommes séparées :

$$\begin{aligned} & \sum_{1 \leq j, k \leq n} a_j b_j - \sum_{1 \leq j, k \leq n} a_j b_k - \sum_{1 \leq j, k \leq n} a_k b_j + \sum_{1 \leq j, k \leq n} a_k b_k \\ &= 2 \sum_{1 \leq j, k \leq n} a_k b_k - 2 \sum_{1 \leq j, k \leq n} a_j b_k \\ &= 2n \sum_{1 \leq k \leq n} a_k b_k - 2 \left( \sum_{k=1}^n a_k \right) \left( \sum_{k=1}^n b_k \right). \end{aligned}$$

Dans la dernière étape, on a simplifié les deux sommes selon la règle de distributivité (2.28). La manipulation de la première somme peut paraître mystérieuse ; voici la même au ralenti :

$$\begin{aligned} 2 \sum_{1 \leq j, k \leq n} a_k b_k &= 2 \sum_{1 \leq k \leq n} \sum_{1 \leq j \leq n} a_k b_k \\ &= 2 \sum_{1 \leq k \leq n} a_k b_k \sum_{1 \leq j \leq n} 1 \\ &= 2 \sum_{1 \leq k \leq n} a_k b_k n = 2n \sum_{1 \leq k \leq n} a_k b_k. \end{aligned}$$

On peut aisément éliminer une variable qui n'apparaît pas dans le terme générique (ici  $j$ ) en multipliant ce qui est à gauche par la taille de l'ensemble des indices (ici  $n$ ).

Revenons à nos moutons. Maintenant on peut tout diviser par 2 et réorganiser les choses pour obtenir une formule intéressante :

$$\left( \sum_{k=1}^n a_k \right) \left( \sum_{k=1}^n b_k \right) = n \sum_{k=1}^n a_k b_k - \sum_{1 \leq j < k \leq n} (a_k - a_j)(b_k - b_j) \quad (2.34)$$

(En réalité, le résultat de Tchebychev [58] concerne les intégrales, non les sommes :

$\left( \int_a^b f(x) dx \right) \cdot \left( \int_a^b g(x) dx \right) \leq (b-a) \cdot \left( \int_a^b f(x)g(x) dx \right)$ , si  $f(x)$  et  $g(x)$  sont deux fonctions monotones non décroissantes.

Cette identité généralise les *inégalités monotones de Tchebychev* :

$$\begin{aligned} \left( \sum_{k=1}^n a_k \right) \left( \sum_{k=1}^n b_k \right) &\leq n \sum_{k=1}^n a_k b_k, \quad \text{si } a_1 \leq \dots \leq a_n \text{ et } b_1 \leq \dots \leq b_n; \\ \left( \sum_{k=1}^n a_k \right) \left( \sum_{k=1}^n b_k \right) &\geq n \sum_{k=1}^n a_k b_k, \quad \text{si } a_1 \leq \dots \leq a_n \text{ et } b_1 \geq \dots \geq b_n. \end{aligned}$$

(Si  $a_1 \leq \dots \leq a_n$  et si  $p$  est une permutation de  $\{1, \dots, n\}$ , on peut prouver sans difficulté que l'expression  $\sum_{k=1}^n a_k b_{p(k)}$  atteint sa valeur maximale lorsque  $b_{p(1)} \leq \dots \leq b_{p(n)}$ , et sa valeur minimale quand  $b_{p(1)} \geq \dots \geq b_{p(n)}$ ).

Il existe une relation intéressante entre la sommation multiple et l'opération de changement d'indice de sommation dans les sommes *simples*. Nous savons qu'en vertu de la règle de commutativité, si  $p(k)$  est une permutation sur l'ensemble des entiers, alors

$$\sum_{k \in K} a_k = \sum_{p(k) \in K} a_{p(k)}.$$

Mais que se passe-t-il si l'on remplace  $k$  par  $f(j)$ , où  $f$  est une fonction arbitrairement choisie

$$f: J \rightarrow K$$

qui fait correspondre à tout entier  $j \in J$  un entier  $f(j) \in K$ ? La formule générale du changement d'indice est

$$\sum_{j \in J} a_{f(j)} = \sum_{k \in K} a_k \#f^-(k), \quad (2.35)$$

où  $\#f^-(k)$  représente le nombre d'éléments de l'ensemble

$$f^-(k) = \{ j \mid f(j) = k \},$$

c'est-à-dire le nombre de valeurs  $j \in J$  telles que  $f(j)$  est égal à  $k$ .

On peut aisément démontrer (2.35) en modifiant l'ordre de sommation,

$$\sum_{j \in J} a_{f(j)} = \sum_{\substack{j \in J \\ k \in K}} a_k [f(j)=k] = \sum_{k \in K} a_k \sum_{j \in J} [f(j)=k],$$

du fait que  $\sum_{j \in J} [f(j)=k] = \#f^-(k)$ . Dans le cas particulier où  $f$  est une correspondance terme à terme entre  $J$  et  $K$ , on a  $\#f^-(k) = 1$  pour tout  $k$ , et la formule générale (2.35) se réduit en

$$\sum_{j \in J} a_{f(j)} = \sum_{f(j) \in K} a_{f(j)} = \sum_{k \in K} a_k.$$

C'est une version un peu déguisée de la règle de commutativité que nous avions auparavant (2.17).

Les exemples de sommes multiples que nous avons vus jusqu'à présent mettaient en jeu des termes généraux, comme  $a_k$  ou  $b_k$ . Comme ce livre

*Mon autre prof de math appelle ça une "bijection"; peut-être qu'un jour je m'y habituerai.*

est sensé être concret, jetons quand même un coup d'œil sur une somme multiple faisant intervenir de vrais nombres :

*Méfiez-vous, les auteurs ont l'air de croire que  $j$ ,  $k$ , et  $n$  sont des "vrais nombres".*

$$S_n = \sum_{1 \leq j < k \leq n} \frac{1}{k-j}.$$

Par exemple,  $S_1 = 0$  ;  $S_2 = 1$  ;  $S_3 = \frac{1}{2-1} + \frac{1}{3-1} + \frac{1}{3-2} = \frac{5}{2}$ .

Pour évaluer une somme double, il faut sommer soit d'abord sur  $j$ , soit d'abord sur  $k$ . Essayons les deux alternatives.

$$\begin{aligned} S_n &= \sum_{1 \leq k \leq n} \sum_{1 \leq j < k} \frac{1}{k-j} && \text{on somme d'abord sur } j \\ &= \sum_{1 \leq k \leq n} \sum_{1 \leq k-j < k} \frac{1}{j} && \text{on remplace } j \text{ par } k-j \\ &= \sum_{1 \leq k \leq n} \sum_{0 < j \leq k-1} \frac{1}{j} && \text{on simplifie les bornes de } j \\ &= \sum_{1 \leq k \leq n} H_{k-1} && \text{d'après (2.13), la définition de } H_{k-1} \\ &= \sum_{1 \leq k+1 \leq n} H_k && \text{on remplace } k \text{ par } k+1 \\ &= \sum_{0 \leq k < n} H_k. && \text{on simplifie les bornes de } k \end{aligned}$$

Malheur ! Nous n'arrivons pas à trouver une formule close pour la somme des nombres harmoniques.

*Au cachot !*

Si on essaie de sommer de l'autre manière, on obtient

$$\begin{aligned} S_n &= \sum_{1 \leq j \leq n} \sum_{j < k \leq n} \frac{1}{k-j} && \text{on somme d'abord sur } k \\ &= \sum_{1 \leq j \leq n} \sum_{j < k+j \leq n} \frac{1}{k} && \text{on remplace } k \text{ par } k+j \\ &= \sum_{1 \leq j \leq n} \sum_{0 < k \leq n-j} \frac{1}{k} && \text{on simplifie les bornes de } k \\ &= \sum_{1 \leq j \leq n} H_{n-j} && \text{d'après (2.13), la définition de } H_{n-j} \\ &= \sum_{1 \leq n-j \leq n} H_j && \text{on remplace } j \text{ par } n-j \\ &= \sum_{0 \leq j < n} H_j. && \text{on simplifie les bornes de } j \end{aligned}$$

Nous sommes encore dans l'impassée.

## 44 SOMMES

Il y a cependant une *autre* façon de procéder, qui consiste à remplacer  $k$  par  $k + j$  *avant* de transformer  $S_n$  en une somme de sommes :

$$\begin{aligned}
 S_n &= \sum_{1 \leq j < k \leq n} \frac{1}{k-j} && \text{on recopie la somme donnée} \\
 &= \sum_{1 \leq j < k+j \leq n} \frac{1}{k} && \text{on remplace } k \text{ par } k+j \\
 &= \sum_{1 \leq k \leq n} \sum_{1 \leq j \leq n-k} \frac{1}{k} && \text{on somme d'abord sur } j \\
 &= \sum_{1 \leq k \leq n} \frac{n-k}{k} && \text{la somme sur } j \text{ est triviale} \\
 &= \sum_{1 \leq k \leq n} \frac{n}{k} - \sum_{1 \leq k \leq n} 1 && \text{d'après la règle d'associativité} \\
 &= n \left( \sum_{1 \leq k \leq n} \frac{1}{k} \right) - n && \text{d'après nous} \\
 &= nH_n - n. && \text{d'après (2.13), la définition de } H_n
 \end{aligned}$$

*Plutôt fûté de mettre  $k \leq n$  au lieu de  $k \leq n-1$  dans ce calcul. En prenant des bornes simples, on ne gaspille pas d'énergie.*

Enfin ! Nous avons trouvé  $S_n$ . En combinant ce résultat avec le faux départ de tout à l'heure, on gagne même une identité supplémentaire :

$$\sum_{0 \leq k < n} H_k = nH_n - n. \quad (2.36)$$

Il y a deux façons de comprendre le truc qui a fonctionné ici : l'une est algébrique, l'autre géométrique. (1) Algébriquement, si on est en présence d'une somme double dont les termes font appel à  $k + f(j)$  où  $f$  est une fonction, cet exemple nous dit que c'est une bonne idée d'essayer de remplacer  $k$  par  $k - f(j)$  et de sommer sur  $j$ . (2) Géométriquement, on peut considérer cette somme particulière  $S_n$  de la façon suivante, si  $n = 4$  par exemple :

$$\begin{array}{cccc}
 k = 1 & k = 2 & k = 3 & k = 4 \\
 j = 1 & \frac{1}{1} + \frac{1}{2} + \frac{1}{3} \\
 j = 2 & \frac{1}{1} + \frac{1}{2} \\
 j = 3 & \frac{1}{1} \\
 j = 4 &
 \end{array}$$

Lors de nos premiers essais, nous avons sommé sur  $j$  (par colonnes) ou sur  $k$  (par lignes) ; cela donnait  $H_1 + H_2 + H_3 = H_3 + H_2 + H_1$ . La stratégie gagnante a consisté en fait à sommer par diagonales pour trouver  $\frac{3}{1} + \frac{2}{2} + \frac{1}{3}$ .

## 2.5 MÉTHODES GÉNÉRALES

Nous allons maintenant consolider ce que nous avons appris en étudiant un même exemple sous différents angles. Dans les quelques pages qui suivent, nous allons essayer de trouver une formule close pour la somme des  $n$  premiers carrés, que nous allons appeler  $\square_n$  :

$$\square_n = \sum_{0 \leq k \leq n} k^2, \quad \text{pour } n \geq 0. \quad (2.37)$$

Nous verrons qu'il existe au moins sept façons différentes de résoudre ce problème. Ce faisant, nous découvrirons des stratégies bien utiles pour s'attaquer aux sommes en général.

Comme d'habitude, regardons d'abord quelques petits cas.

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12
$n^2$	0	1	4	9	16	25	36	49	64	81	100	121	144
$\square_n$	0	1	5	14	30	55	91	140	204	285	385	506	650

A première vue, il n'y a pas de formule close évidente. Quand nous en aurons trouvé une, ce tableau pourra quand même nous servir à vérifier ses premières valeurs.

### Méthode 0 : chercher la solution ailleurs.

Le problème de la somme des  $n$  premiers carrés a probablement déjà été résolu ailleurs. On trouvera donc très vraisemblablement la solution dans un bon ouvrage de référence. En effet, la réponse se trouve bien évidemment en page 36 de *CRC Standard Mathematical Tables* [28] :

$$\square_n = \frac{n(n+1)(2n+1)}{6}, \quad \text{pour } n \geq 0. \quad (2.38)$$

Juste pour être sûrs de l'avoir bien lue, vérifions que cette formule donne bien  $\square_5 = 5 \cdot 6 \cdot 11/6 = 55$ . Notons en passant que, la page 36 des *CRC Tables* donne aussi des informations sur les sommes des cubes, puissances quatrièmes etc, jusqu'aux puissances dixièmes.

La référence incontournable pour les formules mathématiques est le *Handbook of Mathematical Functions*, édité par Abramowitz et Stegun [2]. En pages 813 et 814, on y trouve les valeurs de  $\square_n$  pour  $n \leq 100$ . De plus, les pages 804 et 809 contiennent des formules équivalentes à (2.38), ainsi que des formules analogues pour les sommes des cubes etc, jusqu'aux puissances quinzièmes, avec ou sans signes alternants.

Toutefois, la meilleure source pour répondre à des questions sur des suites de nombres est un livre étonnant, *The Encyclopedia of Integer Sequences* de Sloane et Plouffe [330] (nouvelle édition du *Handbook of Integer Sequences* de Sloane, qui catalogue plusieurs milliers de suites d'entiers,

(On peut trouver des formules pour des sommes plus compliquées dans les tables très complètes de Hansen [178].)

chacune d'entre elles étant représentée par ses premières valeurs. Si vous avez trouvé une récurrence et que vous soupçonnez qu'elle a déjà été étudiée, tout ce que vous avez à faire est d'en calculer assez de termes pour la distinguer d'autres récurrences connues. Il y a alors de bonnes chances pour que vous trouviez dans cette Encyclopédie un pointeur sur une bonne référence. Par exemple, la suite 1, 5, 14, 30, ... porte le numéro M3844 dans l'ouvrage en question. On l'appelle la suite des "nombres pyramidaux carrés" (car il y a  $\square_n$  boules dans une pyramide dont la base est un carré de  $n^2$  boules). Sloane et Plouffe indiquent trois références, l'une étant le manuel d'Abrahamowitz et Stegun mentionné plus haut.

Un autre moyen d'explorer la réserve mondiale de savoir mathématique est de recourir à un logiciel (comme Axiom, MACSYMA, Maple ou Mathematica) permettant d'effectuer des calculs symboliques. De tels logiciels sont indispensables, notamment pour les gens qui doivent manipuler des formules de grande taille.

Il est bon de connaître les sources d'information standard, car elles peuvent être extrêmement utiles. Cependant la méthode 0 n'est pas vraiment en accord avec l'esprit de ce livre, car notre but est d'apprendre à trouver les réponses par nous-mêmes. Elle ne peut résoudre que des problèmes qui ont déjà été considérés par d'autres.

### **Méthode 1: Deviner la réponse et la prouver par induction.**

Imaginons qu'un petit oiseau nous ait soufflé la réponse, ou qu'on soit arrivé à une formule close par quelque autre moyen pas vraiment rigoureux. Il ne reste alors plus qu'à prouver que la formule est correcte.

On pourrait par exemple avoir remarqué que les valeurs de  $\square_n$  ont des facteurs premiers plutôt petits, et trouver ainsi que la formule (2.38) convient pour les petites valeurs de  $n$ . On aurait pu aussi conjecturer la formule équivalente

$$\square_n = \frac{n(n + \frac{1}{2})(n + 1)}{3}, \quad \text{pour } n \geq 0, \quad (2.39)$$

qui est plus agréable car plus facile à retenir. Il est quasiment évident que la formule (2.39) est correcte ; cependant, comme toute conjecture, elle doit être prouvée pour qu'aucun doute raisonnable ne subsiste. C'est pour cela qu'on a inventé l'induction mathématique.

"Bien. Votre Honneur, nous savons que  $\square_0 = 0 = 0(0 + \frac{1}{2})(0 + 1)/3$ , ce qui nous donne la base. En ce qui concerne l'induction, supposons que  $n > 0$  et que (2.39) est vérifiée si l'on remplace  $n$  par  $n - 1$ . Comme

$$\square_n = \square_{n-1} + n^2,$$

*Ou des problèmes ayant les mêmes réponses que des problèmes déjà résolus par d'autres.*

on a

$$\begin{aligned} 3\Box_n &= (n-1)(n-\frac{1}{2})(n) + 3n^2 \\ &= (n^3 - \frac{3}{2}n^2 + \frac{1}{2}n) + 3n^2 \\ &= (n^3 + \frac{3}{2}n^2 + \frac{1}{2}n) \\ &= n(n+\frac{1}{2})(n+1). \end{aligned}$$

Donc, sans aucun doute raisonnable, (2.39) est vraie pour tout  $n \geq 0$ .” Dans son infinie sagesse, le juge Wapner acquiesce.

Utiliser l’induction dans un problème de ce genre est une attitude tout à fait défendable ; c’est quand même un peu mieux que de chercher la réponse dans la littérature. Cependant ce n’est pas encore tout à fait ce qu’il nous faut. Jusque ici, nous avons réussi à calculer chacune des autres sommes de ce chapitre sans faire appel à l’induction ; nous devrions également pouvoir calculer une somme comme  $\Box_n$  en partant de rien, c’est-à-dire sans attendre un éclair d’inspiration subite. Nous devrions pouvoir évaluer des sommes même dans les moments où notre imagination nous fait défaut.

### *Method 2: Perturber la somme.*

Revenons donc à la méthode de perturbation qui a si bien marché pour la progression géométrique (2.25). Extrayons le premier et le dernier terme de  $\Box_{n+1}$  pour obtenir une équation pour  $\Box_n$  :

$$\begin{aligned} \Box_n + (n+1)^2 &= \sum_{0 \leq k \leq n} (k+1)^2 = \sum_{0 \leq k \leq n} (k^2 + 2k + 1) \\ &= \sum_{0 \leq k \leq n} k^2 + 2 \sum_{0 \leq k \leq n} k + \sum_{0 \leq k \leq n} 1 \\ &= \Box_n + 2 \sum_{0 \leq k \leq n} k + (n+1). \end{aligned}$$

*Match nul.*

Aïe ! Les  $\Box_n$  se neutralisent mutuellement. Il arrive parfois que, malgré tous nos efforts, la méthode de perturbation aboutisse à quelque chose comme  $\Box_n = \Box_n$  ; dans ce cas, on a raté notre coup.

D’un autre côté, on n’a pas tout perdu dans cette affaire. Ce calcul nous révèle un moyen de trouver une formule close pour la somme des  $n$  premiers entiers,

$$2 \sum_{0 \leq k \leq n} k = (n+1)^2 - (n+1),$$

alors que notre objectif initial était de sommer les carrés des premiers entiers. Se pourrait-il qu’en partant de la somme des cubes des entiers, qu’on pourrait appeler  $\boxtimes_n$ , on obtienne une expression pour la somme des carrés ?

Essayons.

$$\begin{aligned}\square_n + (n+1)^3 &= \sum_{0 \leq k \leq n} (k+1)^3 = \sum_{0 \leq k \leq n} (k^3 + 3k^2 + 3k + 1) \\ &= \square_n + 3\square_n + 3\frac{(n+1)n}{2} + (n+1).\end{aligned}$$

Effectivement, les  $\square_n$  disparaissent et nous avons maintenant suffisamment d'information pour déterminer  $\square_n$  sans faire appel à l'induction :

*Méthode 2' : perturber votre chargé de TD.*

$$\begin{aligned}3\square_n &= (n+1)^3 - 3(n+1)n/2 - (n+1) \\ &= (n+1)(n^2 + 2n + 1 - \frac{3}{2}n - 1) = (n+1)(n + \frac{1}{2})n.\end{aligned}$$

### Méthode 3: Construire un répertoire.

On peut aussi se contenter de généraliser un petit peu la récurrence (2.7). La solution de

$$\begin{aligned}R_0 &= \alpha; \\ R_n &= R_{n-1} + \beta + \gamma n + \delta n^2, \quad \text{pour } n > 0,\end{aligned}\tag{2.40}$$

sera de la forme

$$R_n = A(n)\alpha + B(n)\beta + C(n)\gamma + D(n)\delta;\tag{2.41}$$

où  $A(n)$ ,  $B(n)$  et  $C(n)$  sont déjà connus, car l'équation (2.40) est identique à (2.7) si  $\delta = 0$ . Maintenant, si on pose  $R_n = n^3$ , on trouve que  $n^3$  est solution quand  $\alpha = 0$ ,  $\beta = 1$ ,  $\gamma = -3$  et  $\delta = 3$ . Donc

$$3D(n) - 3C(n) + B(n) = n^3;$$

ce qui détermine  $D(n)$ .

C'est la somme  $\square_n$  qui nous intéresse ; elle est égale à  $\square_{n-1} + n^2$ . Ainsi  $\square_n = R_n$  si  $\alpha = \beta = \gamma = 0$  et  $\delta = 1$  dans (2.41). Par conséquent,  $\square_n = D(n)$ . Pas besoin de calculer  $D(n)$  en fonction de  $B(n)$  et  $C(n)$ , puisque nous connaissons déjà la réponse. Rassurons quand même les plus sceptiques en écrivant

$$3D(n) = n^3 + 3C(n) - B(n) = n^3 + 3\frac{(n+1)n}{2} - n = n(n + \frac{1}{2})(n + 1).$$

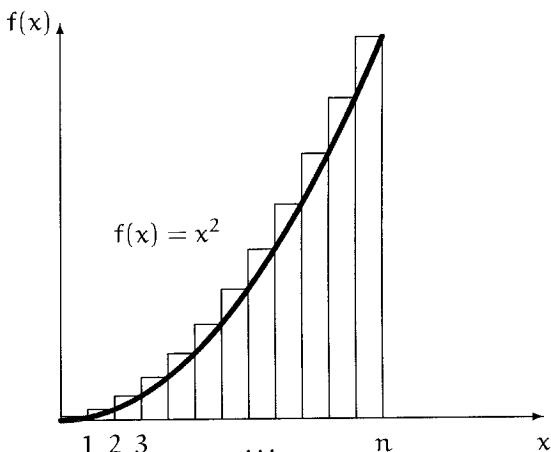
### Méthode 4 : remplacer les sommes par des intégrales.

Les personnes élevées au calcul infinitésimal plutôt qu'aux mathématiques discrètes sont généralement plus familières avec  $\int$  qu'avec  $\sum$ , et essaient donc naturellement de remplacer  $\sum$  par  $\int$ . Un des buts de ce livre est de nous permettre d'être tellement à l'aise avec  $\sum$  que nous trouverons  $\int$  plus difficile que  $\sum$  (au moins pour des calculs exacts). Cependant, c'est

une bonne idée de se pencher sur les relations qui existent entre  $\sum$  et  $\int$ , car la sommation et l'intégration sont basées sur des idées très similaires.

En calcul infinitésimal, une intégrale peut être vue comme l'aire sous une courbe. On peut approximer cette aire en additionnant les aires d'un certain nombre de rectangles verticaux allongés qui touchent la courbe. On peut aussi faire l'inverse pour approximer l'aire d'un ensemble de rectangles verticaux allongés donnés : comme  $\square_n$  est la somme des aires de rectangles de dimensions  $1 \times 1, 1 \times 4, \dots, 1 \times n^2$ , elle est à peu près égale à l'aire sous la courbe  $f(x) = x^2$  entre 0 et  $n$ .

*L'échelle horizontale est dix fois plus grande que l'échelle verticale.*



L'aire sous cette courbe est  $\int_0^n x^2 dx = n^3/3$  ; nous en déduisons donc que  $\square_n$  est approximativement égal à  $\frac{1}{3}n^3$ .

On peut utiliser ce fait pour calculer l'erreur d'approximation,  $E_n = \square_n - \frac{1}{3}n^3$ . Comme  $\square_n$  satisfait la récurrence  $\square_n = \square_{n-1} + n^2$ , on trouve que  $E_n$  satisfait une récurrence plus simple :

$$\begin{aligned} E_n &= \square_n - \frac{1}{3}n^3 = \square_{n-1} + n^2 - \frac{1}{3}n^3 = E_{n-1} + \frac{1}{3}(n-1)^3 + n^2 - \frac{1}{3}n^3 \\ &= E_{n-1} + n - \frac{1}{3}. \end{aligned}$$

Un autre façon d'utiliser l'approche par intégrale consiste à trouver une formule pour  $E_n$  en sommant les aires des termes de l'erreur, aires des "coins" des rectangles :

*Ca, c'est pour les accros du calcul infinitésimal.*

$$\begin{aligned} \square_n - \int_0^n x^2 dx &= \sum_{k=1}^n \left( k^2 - \int_{k-1}^k x^2 dx \right) \\ &= \sum_{k=1}^n \left( k^2 - \frac{k^3 - (k-1)^3}{3} \right) = \sum_{k=1}^n \left( k - \frac{1}{3} \right). \end{aligned}$$

D'une façon ou d'une autre, on trouvera  $E_n$  et donc  $\square_n$ .

**Méthode 5 : développer et simplifier.**

Voici encore un moyen de trouver une forme close pour  $\square_n$  : il s'agit de remplacer la somme de départ par un double somme, apparemment plus compliquée, mais que nous pourrons simplifier si nous la traitons convenablement :

$$\begin{aligned}\square_n &= \sum_{1 \leq k \leq n} k^2 = \sum_{1 \leq j \leq k \leq n} k \\ &= \sum_{1 \leq j \leq n} \sum_{j \leq k \leq n} k \\ &= \sum_{1 \leq j \leq n} \left( \frac{j+n}{2} \right) (n-j+1) \\ &= \frac{1}{2} \sum_{1 \leq j \leq n} (n(n+1) + j - j^2) \\ &= \frac{1}{2} n^2(n+1) + \frac{1}{4} n(n+1) - \frac{1}{2} \square_n = \frac{1}{2} n(n+\frac{1}{2})(n+1) - \frac{1}{2} \square_n.\end{aligned}$$

(Cette dernière étape ressemble un peu à la dernière étape de la méthode de perturbation : dans les deux cas, on obtient une équation contenant l'inconnue dans les deux membres).

A priori, passer d'une somme simple à une somme double peut passer pour une régression. C'est en réalité un progrès, car ainsi apparaissent des sommes plus faciles à manipuler. On ne peut pas espérer résoudre tout problème si on ne fait que simplifier, simplifier et encore simplifier : on ne peut pas atteindre les plus hauts pics d'une montagne en montant continuellement !

**Méthode 6 : utiliser le calcul fini.****Méthode 7 : utiliser les fonctions génératrices.**

Restez avec nous pour ne pas rater des calculs encore plus passionnants de  $\square_n = \sum_{k=0}^n k^2$ , avec les techniques que nous allons voir dans les sections et chapitres suivants.

## 2.6 CALCUL FINI ET INFINI

Nous avons vu un certain nombre de méthodes pour la manipulation directe des sommes. Il est temps maintenant d'acquérir une vision plus large des choses, en regardant le problème de la sommation d'un point de vue plus élevé. Avec le "calcul fini", que les mathématiciens ont développé par analogie avec le calcul infinitésimal, il est possible d'approcher les problèmes de sommation de façon agréable et systématique.

Le calcul infinitésimal est basé sur les propriétés de l'opérateur de *dérivation* D, défini par

$$Df(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Le calcul fini, quant à lui, est basé sur les propriétés de l'opérateur de *différence*  $\Delta$ , défini par

$$\Delta f(x) = f(x+1) - f(x). \quad (2.42)$$

C'est l'analogie fini de la dérivation, où on se restreint à des valeurs entières strictement positives de  $h$ . De ce fait, la valeur  $h = 1$  est la plus proche possible de la "limite"  $h \rightarrow 0$ , et  $\Delta f(x)$  est égale à  $(f(x+h) - f(x))/h$  pour  $h = 1$ .

Les symboles  $D$  et  $\Delta$  sont appelés des *opérateurs* car ils opèrent sur des fonctions pour donner de nouvelles fonctions ; ce sont des fonctions de fonctions qui produisent des fonctions. Si  $f$  est une fonction suffisamment lisse des nombres réels vers les nombres réels, alors  $Df$  est aussi une fonction des réels vers les réels. Si  $f$  est une *quelconque* fonction des réels vers les réels,  $\Delta f$  l'est aussi. Les valeurs des fonctions  $Df$  et  $\Delta f$  en un point  $x$  sont données par les définitions ci-dessus.

On apprend vite en mathématiques comment  $D$  opère sur les puissances  $f(x) = x^m$  : dans ce cas  $Df(x) = mx^{m-1}$ . On peut l'écrire moins formellement en oubliant  $f$  :

$$D(x^m) = mx^{m-1}.$$

Ce serait bien si l'opérateur  $\Delta$  produisait une résultat si élégant ; hélas ce n'est pas le cas. On a par exemple

$$\Delta(x^3) = (x+1)^3 - x^3 = 3x^2 + 3x + 1.$$

*Quelle puissance dans les maths !*

Toutefois, il existe un type de "puissance mième" qui se comporte très agréablement quand on lui applique  $\Delta$ , et c'est précisément ce qui rend le calcul fini intéressant. Ces puissances mièmes nouvelle mouture sont définies par la règle suivante :

$$x^{\underline{m}} = \overbrace{x(x-1)\dots(x-m+1)}^{\text{m facteurs}}, \quad m \text{ entier}, m \geq 0. \quad (2.43)$$

Remarquez que  $m$  est souligné ; cela signifie que les  $m$  facteurs sont supposés diminuer pas à pas. Il existe aussi une définition pour les facteurs qui augmentent :

$$x^{\overline{m}} = \overbrace{x(x+1)\dots(x+m-1)}^{\text{m facteurs}}, \quad m \text{ entier}, m \geq 0. \quad (2.44)$$

Quand  $m = 0$ , on a  $x^{\underline{0}} = x^{\overline{0}} = 1$ , car par convention un produit sans facteur est égal à 1 (tout comme une somme sans termes est égale à 0).

La quantité  $x^{\underline{m}}$  est appelée "x puissance  $m$  descendante", tandis que  $x^{\overline{m}}$  est appelée "x puissance  $m$  montante". On désigne aussi ces fonctions sous

les noms de *puissance factorielle descendante* et *puissance factorielle montante* car elles sont étroitement liées à la fonction factorielle  $n! = n(n-1)\dots(1)$ . Effectivement,  $n! = n^n = 1^n$ .

On trouve plusieurs notations pour les puissances factorielles dans la littérature mathématique, notamment le “symbole de Pochhammer”  $(x)_m$  qui désigne  $x^{\overline{m}}$  ou  $x^m$ ; on peut trouver aussi des notations comme  $x^{(m)}$  ou  $x_{(m)}$  pour  $x^m$ . Cependant la convention souligné/surligné l'emporte car elle est facile à écrire, à retenir, et ne contient pas de parenthèses redondantes.

Les puissances factorielles descendantes  $x^m$  se comportent de façon particulièrement agréable lorsqu'on leur applique  $\Delta$ . En effet,

$$\begin{aligned}\Delta(x^m) &= (x+1)^m - x^m \\ &= (x+1)x\dots(x-m+2) - x\dots(x-m+2)(x-m+1) \\ &= m x(x-1)\dots(x-m+2),\end{aligned}$$

ce qui donne au calcul fini une règle pratique, similaire à  $D(x^m) = mx^{m-1}$ :

$$\Delta(x^m) = mx^{m-1}. \quad (2.45)$$

L'opérateur  $D$  du calcul infinitésimal a un inverse, l'opérateur d'intégration  $\int$ . Le Théorème Fondamental du Calcul Infinitésimal lie  $D$  à  $\int$ :

$$g(x) = Df(x) \quad \text{si et seulement si} \quad \int g(x) dx = f(x) + C.$$

où  $\int g(x) dx$ , l'intégrale indéfinie de  $g(x)$ , désigne l'ensemble des fonctions dont la dérivée est  $g(x)$ . De façon analogue,  $\Delta$  a un inverse, l'opérateur de sommation  $\sum$ . On peut ainsi énoncer un autre Théorème Fondamental :

$$g(x) = \Delta f(x) \quad \text{si et seulement si} \quad \sum g(x) \delta x = f(x) + C. \quad (2.46)$$

où  $\sum g(x) \delta x$ , la *somme indéfinie* de  $g(x)$ , désigne l'ensemble des fonctions dont la *différence* est  $g(x)$  (remarquez que le  $\delta$  minuscule est lié au  $\Delta$  majuscule, comme  $d$  l'est à  $D$ ). Le “C”, pour une intégrale indéfinie, représente une constante arbitraire ; lorsqu'il s'agit d'une somme indéfinie, il peut représenter toute fonction  $p(x)$  telle que  $p(x+1) = p(x)$ . Par exemple,  $C$  pourrait être la fonction périodique  $a + b \sin 2\pi x$  ; une telle fonction disparaît quand on calcule les différences, tout comme une constante disparaît lorsqu'on calcule une dérivée. Pour des valeurs entières de  $x$ , la fonction  $C$  est constante.

Nous arrivons maintenant au plat de résistance. En calcul infinitésimal, on rencontre aussi des intégrales *définies* : si  $g(x) = Df(x)$ , alors

$$\int_a^b g(x) dx = f(x) \Big|_a^b = f(b) - f(a).$$

*La terminologie mathématique est parfois complètement dingue : en fait, Pochhammer [293] a utilisé la notation  $(x)_m$  pour désigner le coefficient binomial  $\binom{x}{m}$ , non les puissances factorielles.*

*Quemadmodum ad differentiam denotandam usi sumus signo  $\Delta$ , ita summam indicabimus signo  $\Sigma$ . . . ex quo æquatio  $z = \Delta y$ , si invertatur, dabit quoque  $y = \Sigma z + C$ .*  
—L. Euler [110]

De même, le calcul fini, prenant toujours modèle sur son cousin plus célèbre, a ses *sommes* définies : si  $g(x) = \Delta f(x)$ , alors

$$\sum_a^b g(x) \delta x = f(x) \Big|_a^b = f(b) - f(a). \quad (2.47)$$

Cette formule donne un sens à la notation  $\sum_a^b g(x) \delta x$ , de la même façon que la précédente définit  $\int_a^b g(x) dx$ .

Quel est, intuitivement, le véritable sens de l'expression  $\sum_a^b g(x) \delta x$ ? Nous l'avons définie pour pouvoir retenir aisément les règles du calcul fini ; mais cette notation restera inutile si nous ne comprenons pas sa signification. Essayons de l'appréhender, d'abord en regardant quelques cas particuliers. Supposons que  $g(x) = \Delta f(x) = f(x+1) - f(x)$ . Si  $b = a$ , alors

$$\sum_a^a g(x) \delta x = f(a) - f(a) = 0.$$

Puis, si on pose  $b = a + 1$ , on trouve

$$\sum_a^{a+1} g(x) \delta x = f(a+1) - f(a) = g(a).$$

Plus généralement, si  $b$  augmente de 1, alors on a

$$\begin{aligned} \sum_a^{b+1} g(x) \delta x - \sum_a^b g(x) \delta x &= (f(b+1) - f(a)) - (f(b) - f(a)) \\ &= f(b+1) - f(b) = g(b). \end{aligned}$$

Avec l'aide de l'induction mathématique, ces observations nous permettent de déduire exactement la signification de  $\sum_a^b g(x) \delta x$ , lorsque  $a$  et  $b$  sont des entiers tels que  $b \geq a$  :

$$\sum_a^b g(x) \delta x = \sum_{k=a}^{b-1} g(k) = \sum_{a \leq k < b} g(k), \quad \text{pour } b \geq a \text{ entiers.} \quad (2.48)$$

Et vous appelez ça  
un plat de résis-  
tance ?

En d'autres termes, la somme définie est pareille à une somme ordinaire avec ses limites, sauf qu'on exclut la valeur limite supérieure.

Essayons de présenter ceci un peu différemment. Soit une somme dont nous cherchons une forme close. Supposons que nous pouvons l'exprimer sous la forme  $\sum_{a \leq k < b} g(k) = \sum_a^b g(x) \delta x$ . Selon la théorie du calcul fini, la réponse peut s'écrire  $f(b) - f(a)$ , si on connaît une somme indéfinie ou une fonction  $f$  telle que  $g(x) = f(x+1) - f(x)$ . On comprend ce principe si on écrit entièrement  $\sum_{a \leq k < b}$  en utilisant les points de suspension :

$$\begin{aligned} \sum_{a \leq k < b} (f(k+1) - f(k)) &= (f(a+1) - f(a)) + (f(a+2) - f(a+1)) + \cdots \\ &\quad + (f(b-1) - f(b-2)) + (f(b) - f(b-1)). \end{aligned}$$

Tout ce qui est dans le membre de droite disparaît, sauf  $f(b) - f(a)$ , qui est donc la valeur de la somme (on dit souvent que les sommes du type  $\sum_{a \leq k < b} (f(k-1) - f(k))$  sont *télescopiques*, par analogie avec un télescope pliant, car l'épaisseur des parois d'un télescope replié est uniquement déterminée par le rayon externe du tube externe et le rayon interne du tube interne).

La règle (2.48) est valable seulement quand  $b \geq a$ . Que se passe-t-il si  $b < a$ ? Dans ce cas, (2.47) nous dit qu'on doit avoir

$$\begin{aligned}\sum_a^b g(x) \delta x &= f(b) - f(a) \\ &= -(f(a) - f(b)) = -\sum_b^a g(x) \delta x.\end{aligned}$$

C'est tout à fait analogue à ce qui se passe pour une intégrale définie. Par un argument similaire on prouve que  $\sum_a^b + \sum_b^c = \sum_a^c$ , exactement comme  $\int_a^b + \int_b^c = \int_a^c$ . Plus formellement,

$$\sum_a^b g(x) \delta x + \sum_b^c g(x) \delta x = \sum_a^c g(x) \delta x, \quad (2.49)$$

pour tous  $a, b$  et  $c$  entiers.

En ce moment, il y a certainement des lecteurs qui commencent à se demander à quoi peuvent bien servir ces analogies. Disons pour commencer que la sommation définie nous fournit un moyen simple de calculer des sommes de puissances descendantes : on déduit des règles de base (2.45), (2.47) et (2.48) la règle générale

$$\sum_{0 \leq k < n} k^m = \frac{k^{m+1}}{m+1} \Big|_0^n = \frac{n^{m+1}}{m+1}, \quad \text{pour } m, n \geq 0 \text{ entiers.} \quad (2.50)$$

Cette formule est facile à retenir car elle ressemble beaucoup à  $\int_0^n x^m dx = n^{m+1}/(m+1)$  qui nous est plus familière.

En particulier, lorsque  $m = 1$  on a  $k^1 = k$ . Ainsi, grâce aux principes du calcul fini, on pourra aisément se rappeler que

$$\sum_{0 \leq k < n} k = \frac{n^2}{2} = n(n-1)/2.$$

La méthode des sommes définies nous fait aussi soupçonner que les sommes sur  $0 \leq k < n$  sont souvent plus simples que les sommes sur  $1 \leq k \leq n$  : les premières sont simplement égales à  $f(n) - f(0)$ , tandis qu'on doit voir les autres comme  $f(n+1) - f(1)$ .

On peut aussi exprimer les puissances ordinaires en fonction de puissances descendantes. Par exemple,

$$k^2 = k^2 + k^1,$$

*Tiens ! Jusqu'ici, je croyais que c'était parce qu'une expression très longue se repliait pour en former une beaucoup plus courte.*

*Les autres se le demandent déjà depuis un certain temps.*

et donc

$$\sum_{0 \leq k < n} k^2 = \frac{n^3}{3} + \frac{n^2}{2} = \frac{1}{3}n(n-1)(n-2+\frac{3}{2}) = \frac{1}{3}n(n-\frac{1}{2})(n-1).$$

Avec des amis  
comme ça...

En remplaçant  $n$  par  $n+1$ , on obtient une nouvelle façon de calculer une forme close pour notre vieil ami  $\square_n = \sum_{0 \leq k \leq n} k^2$ .

Plutôt facile, non ? C'est effectivement plus facile que tous les traitements que nous avons infligés à cette pauvre formule dans la section précédente. Essayons maintenant de franchir une étape supplémentaire, en passant des carrés aux *cubes*. Un calcul simple montre que

$$k^3 = k^3 + 3k^2 + k^1.$$

(Nous verrons au chapitre 6 qu'il est toujours possible de convertir les puissances ordinaires en puissances factorielles ; on utilise pour cela les nombres de Stirling). Ainsi

$$\sum_{a \leq k < b} k^3 = \frac{k^4}{4} + k^3 + \frac{k^2}{2} \Big|_a^b.$$

Les puissances descendantes sont donc très pratiques pour les sommes. Cependant, s'il faut convertir nos bonnes vieilles puissances ordinaires en puissances descendantes pour les sommer, puis les reconvertis en puissances ordinaires avant de faire quoi que ce soit d'autre, ce ne sera pas franchement pratique. Heureusement, les puissances factorielles ont de bonnes propriétés additives ; nous pourrons donc souvent travailler directement avec elles. Par exemple, tout comme  $(x+y)^2 = x^2 + 2xy + y^2$ , il se trouve que  $(x+y)^3 = x^3 + 3x^2y^1 + y^3$ . La même analogie est possible entre  $(x+y)^m$  et  $(x+y)^{\underline{m}}$  (la preuve de ce “théorème du binôme factoriel” constitue l'exercice 5.37).

Jusqu'à présent, nous n'avons considéré que les puissances descendantes qui ont des exposants positifs ou nuls. Pour étendre l'analogie avec les puissances ordinaires aux exposants négatifs, il nous faut une définition appropriée de  $x^{\underline{m}}$  pour  $m < 0$ . Si l'on observe la suite

$$\begin{aligned} x^3 &= x(x-1)(x-2), \\ x^2 &= x(x-1), \\ x^1 &= x, \\ x^0 &= 1, \end{aligned}$$

on remarque qu'en divisant par  $x-2$ , puis par  $x-1$ , puis par  $x$ , on passe de  $x^3$  à  $x^2$ , puis à  $x^1$ , puis à  $x^0$ . Il semble raisonnable (sinon impératif) de diviser ensuite par  $x+1$  pour passer de  $x^0$  à  $x^{-1}$ , obtenant  $x^{-1} = 1/(x+1)$ .

En continuant ainsi, on peut écrire les premières puissances factorielles à exposant négatif :

$$\begin{aligned}x^{-1} &= \frac{1}{x+1}, \\x^{-2} &= \frac{1}{(x+1)(x+2)}, \\x^{-3} &= \frac{1}{(x+1)(x+2)(x+3)},\end{aligned}$$

Voici donc la définition des puissances descendantes négatives :

$$x^{-m} = \frac{1}{(x+1)(x+2)\dots(x+m)}, \quad \text{pour } m > 0. \quad (2.51)$$

(On peut aussi définir les puissances descendantes pour des nombres réels ou même complexes ; nous verrons cela au chapitre 5).

Cette définition a le mérite de donner aux puissances descendantes de bonnes propriétés additives. Parmi celles-ci, la plus importante est probablement l'analogue de la règle générale

$$x^{m+n} = x^m x^n$$

pour les puissances ordinaires. Pour les puissances descendantes, cela donne

$$x^{m+n} = x^m (x-m)^n, \quad \text{pour } m \text{ et } n \text{ entiers.} \quad (2.52)$$

Par exemple,  $x^{2+3} = x^2 (x-2)^3$ , et pour un  $n$  négatif on a

$$x^{2-3} = x^2 (x-2)^{-3} = x(x-1) \frac{1}{(x-1)x(x+1)} = \frac{1}{x+1} = x^{-1}.$$

Si nous avions choisi de définir  $x^{-1}$  comme  $1/x$  au lieu de  $1/(x+1)$ , la règle (2.52) n'aurait pas été valable dans certains cas, comme celui où  $m = -1$  et  $n = 1$ . En fait, on aurait pu utiliser (2.52) pour savoir exactement comment définir les puissances descendantes à exposants négatifs ; il suffisait de poser  $m = -n$ . Quand, dans le but d'élargir son champ d'application, on étend une notation existante, il vaut toujours mieux concevoir les définition de sorte que les règles classiques soient toujours valables.

*Les règles des puissances sont puissantes.*

Maintenant, assurons-nous que la propriété cruciale de l'opérateur de différence s'applique encore à nos nouvelles puissances descendantes. Est-il toujours vrai que  $\Delta x^m = mx^{m-1}$  lorsque  $m < 0$  ? Si  $m = 2$ , par exemple, la différence vaut

$$\Delta x^{-2} = \frac{1}{(x+2)(x+3)} - \frac{1}{(x+1)(x+2)}$$

$$\begin{aligned}
 &= \frac{(x+1) - (x+3)}{(x+1)(x+2)(x+3)} \\
 &= -2x^{-3}.
 \end{aligned}$$

Oui, ça marche ! Un argument similaire permet de généraliser à tout  $m < 0$ .

Ainsi la propriété de sommation (2.50) est valable aussi bien pour les puissances descendantes négatives que pour les positives, du moins tant qu'il n'y a pas de division par zéro :

$$\sum_a^b x^m \delta x = \left. \frac{x^{m+1}}{m+1} \right|_a^b, \quad \text{pour } m \neq -1.$$

Qu'arrive-t-il si  $m = -1$  ? Rappelons que, pour l'intégration, on a

$$\int_a^b x^{-1} dx = \ln x \Big|_a^b.$$

Ce serait bien d'avoir un analogue fini de  $\ln x$ . Il s'agit de trouver une fonction  $f(x)$  telle que

$$x^{-1} = \frac{1}{x+1} = \Delta f(x) = f(x+1) - f(x).$$

Il n'est pas très difficile de voir que, si  $x$  est entier,

$$f(x) = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{x}$$

est une telle fonction. Elle est exactement égale au nombre harmonique  $H_x$  de (2.13). Ainsi  $H_x$  est l'analogue discret de la fonction continue  $\ln x$ . (Dans le chapitre 6, nous définirons  $H_x$  pour des  $x$  non entiers, mais pour l'instant les valeurs entières nous suffisent. Nous verrons aussi au chapitre 9 que, lorsque  $x$  est grand, la valeur de  $H_x - \ln x$  est à peu près égale à 0,577 + 1/(2x). Ainsi, non seulement  $H_x$  et  $\ln x$  sont analogues, mais de plus, dans la plupart des cas, leurs valeurs diffèrent de moins de 1).

*Exactement  
0,577 ? Peut-être  
qu'ils veulent dire  
 $1/\sqrt{3}$ . Ou peut-être  
pas.*

Nous pouvons maintenant décrire complètement les sommes des puissances descendantes :

$$\sum_a^b x^m \delta x = \begin{cases} \left. \frac{x^{m+1}}{m+1} \right|_a^b, & \text{si } m \neq -1; \\ H_x \Big|_a^b, & \text{si } m = -1. \end{cases} \quad (2.53)$$

Cette formule explique pourquoi les nombres harmoniques ont tendance à surgir dans la solution de certains problèmes discrets comme l'analyse du tri rapide ; c'est exactement de la même manière que les logarithmes

dits naturels apparaissent naturellement dans les solutions de problèmes continus.

Maintenant que nous avons trouvé un analogue de  $\ln x$ , voyons s'il y en a un pour  $e^x$ . Quelle est la fonction  $f(x)$  pour laquelle  $\Delta f(x) = f(x)$ , ce qui correspond à  $D e^x = e^x$  ? Facile :

$$f(x+1) - f(x) = f(x) \quad \Leftrightarrow \quad f(x+1) = 2f(x);$$

On déduit de cette simple récurrence que  $f(x) = 2^x$  est qualifiée pour être la fonction exponentielle discrète.

Si  $c$  est un nombre quelconque, la différence de  $c^x$  est aussi très simple :

$$\Delta(c^x) = c^{x+1} - c^x = (c-1)c^x.$$

Par conséquent, l'anti-différence de  $c^x$  est  $c^x/(c-1)$  si  $c \neq 1$ . Ce fait, allié aux règles fondamentales (2.47) et (2.48), nous donne un moyen élégant de comprendre la formule générale qui donne la somme d'une progression géométrique :

$$\sum_{a \leq k < b} c^k = \sum_a^b c^x \delta x = \frac{c^x}{c-1} \Big|_a^b = \frac{c^b - c^a}{c-1}, \quad \text{pour } c \neq 1.$$

Chaque fois qu'on cherche une formule close pour une fonction  $f$ , on peut calculer sa différence  $\Delta f = g$ . On obtient alors une fonction  $g$  dont on connaît la somme indéfinie  $\sum g(x) \delta x$ . La table 59 fournit une aide à la sommation : on y trouve la différence et l'anti-différence de quelques fonctions usuelles.

La "table 59" est en page 59. Pigé ?

Malgré tous les parallèles que l'on peut faire entre le mathématiques continues et les mathématiques discrètes, il y a des notions continues qui n'ont pas d'analogie discret. Par exemple, en calcul infinitésimal, il existe une règle pratique pour trouver la dérivée d'une fonction composée. En calcul fini, il n'y a pas d'équivalent à cette règle. Il est généralement difficile d'effectuer un changement de variable discret, sauf dans certains cas où l'on peut, par exemple, remplacer  $x$  par  $c \pm x$ .

En revanche,  $\Delta(f(x)g(x))$  peut s'exprimer de façon assez plaisante, et nous fournit ainsi une règle de *sommation par parties*, l'analogue fini de l'*intégration par parties* du calcul infinitésimal. Rappelons que de la formule

$$D(uv) = u Dv + v Du$$

du calcul infinitésimal on déduit, en intégrant puis réorganisant les termes, la règle d'intégration par partie

$$\int u Dv = uv - \int v Du.$$

**Table 59** Quelle est la différence ?

$f = \sum g$	$\Delta f = g$	$f = \sum g$	$\Delta f = g$
$x^0 = 1$	0	$2^x$	$2^x$
$x^1 = x$	1	$c^x$	$(c - 1)c^x$
$x^2 = x(x - 1)$	$2x$	$c^x/(c - 1)$	$c^x$
$x^m$	$mx^{m-1}$	$cf$	$c\Delta f$
$x^{m+1}/(m + 1)$	$x^m$	$f + g$	$\Delta f + \Delta g$
$H_x$	$x^{-1} = 1/(x + 1)$	$f g$	$f\Delta g + Eg\Delta f$

On peut faire quelque chose de similaire en calcul fini.

Commençons par appliquer l'opérateur de différence au produit de deux fonctions  $u(x)$  et  $v(x)$  :

$$\begin{aligned}\Delta(u(x)v(x)) &= u(x+1)v(x+1) - u(x)v(x) \\ &= u(x+1)v(x+1) - u(x)v(x+1) \\ &\quad + u(x)v(x+1) - u(x)v(x) \\ &= u(x)\Delta v(x) + v(x+1)\Delta u(x).\end{aligned}\tag{2.54}$$

En utilisant l'*opérateur de décalage*  $E$  défini par

$$Ef(x) = f(x+1),$$

on peut mettre cette formule sous une forme plus pratique :

$$\Delta(uv) = u\Delta v + Ev\Delta u.\tag{2.55}$$

(Le  $E$  est un peu gênant, mais il est nécessaire pour que l'équation soit correcte). En prenant la somme indéfinie de chacun des deux termes de cette équation, puis en les réorganisant, on obtient, comme promis, la règle de sommation par parties :

$$\sum u\Delta v = uv - \sum Ev\Delta u.\tag{2.56}$$

Comme en calcul infinitésimal, on peut obtenir ainsi des sommes définies en ajoutant des bornes sur les trois termes.

Cette règle est très pratique quand la somme de gauche est plus difficile à calculer que celle de droite. Prenons un exemple. La fonction  $\int xe^x dx$  est typiquement une fonction qu'on intègre par parties ; son analogue discret est  $\sum x2^x \delta x$ , que nous avons déjà rencontré dans ce chapitre sous la forme  $\sum_{k=0}^n k 2^k$ . Pour sommer ceci par parties, posons  $u(x) = x$  et  $\Delta v(x) = 2^x$ . Ainsi  $\Delta u(x) = 1$ ,  $v(x) = 2^x$  et  $Ev(x) = 2^{x+1}$ . En appliquant (2.56) on trouve

$$\sum x2^x \delta x = x2^x - \sum 2^{x+1} \delta x = x2^x - 2^{x+1} + C.$$

En calcul infinitésimal, on se débarrasse du  $E$  en faisant tendre 1 vers 0.

J'imagine que  $e^x = 2^x$  pour des petites valeurs de 1.

## 60 SOMMES

De plus, en ajoutant des bornes à cette formule, on peut l'utiliser cette formule pour calculer une somme vue précédemment :

$$\begin{aligned}\sum_{k=0}^n k2^k &= \sum_0^{n+1} x2^x \delta x \\ &= x2^x - 2^{x+1} \Big|_0^{n+1} \\ &= ((n+1)2^{n+1} - 2^{n+2}) - (0 \cdot 2^0 - 2^1) = (n-1)2^{n+1} + 2.\end{aligned}$$

Cette méthode est plus facile à utiliser que la méthode de perturbation, parce qu'ici on n'a pas besoin de réfléchir.

Plus haut dans ce chapitre, nous étions heureux de trouver, après beaucoup d'efforts, la formule (2.36) pour  $\sum_{0 \leq k < n} H_k$ . Nous aurions pu la trouver automatiquement si nous avions connu la sommation par parties. Faisons-en la démonstration en nous attaquant à une somme qui a l'air encore plus difficile,  $\sum_{0 \leq k < n} kH_k$ . On trouve aisément la solution en suivant l'analogie avec  $\int x \ln x \, dx$  : prenons  $u(x) = H_x$  et  $\Delta v(x) = x = x^1$  ; alors  $\Delta u(x) = x^{-1}$ ,  $v(x) = x^2/2$ ,  $\mathbb{E}v(x) = (x+1)^2/2$  et

$$\begin{aligned}\sum xH_x \delta x &= \frac{x^2}{2}H_x - \sum \frac{(x+1)^2}{2} x^{-1} \delta x \\ &= \frac{x^2}{2}H_x - \frac{1}{2} \sum x^1 \delta x \\ &= \frac{x^2}{2}H_x - \frac{x^2}{4} + C.\end{aligned}$$

(Pour passer de la première ligne à la deuxième, nous avons combiné deux puissances descendantes  $(x+1)^2 x^{-1}$  en posant  $m = -1$  et  $n = 2$  dans la règle des puissances (2.52)). Nous pouvons maintenant ajouter des bornes et conclure que

$$\sum_{0 \leq k < n} kH_k = \sum_0^n xH_x \delta x = \frac{n^2}{2} \left( H_n - \frac{1}{2} \right). \quad (2.57)$$

## 2.7 · SOMMES INFINIES

Lorsque nous avons défini la notation  $\sum$  au début de ce chapitre, nous avons finassé lorsqu'est apparu le problème des sommes infinies en disant, en substance, "remettons cela à plus tard ; pour l'instant, nous pouvons supposer que toutes les sommes que nous rencontrons ont un nombre fini de termes non nuls". Il est temps maintenant de regarder les choses en face : il existe des sommes infinies. En fait, c'est à la fois une bonne et une mauvaise nouvelle.

*Le but ultime des mathématiques, c'est d'éliminer tout appel à l'intelligence.*

*C'est ça, finasser ?*

La mauvaise nouvelle d'abord : il se trouve que les méthodes que nous avons utilisées pour manipuler les  $\sum$  ne peuvent *pas* toujours être appliquées aux sommes infinies. Et maintenant la bonne nouvelle : il existe une grande classe de sommes infinies, facilement reconnaissables, pour lesquelles toutes les opérations que nous avons utilisées peuvent légitimement être appliquées. Lorsque nous aurons regardé de plus près la signification profonde de la sommation, les raisons de ces deux affirmations seront claires.

Tout le monde sait ce qu'est une somme finie : on ajoute les termes un par un jusqu'à ce qu'il n'y en ait plus. En revanche, une somme infinie doit être définie avec plus de circonspection, faute de quoi on peut aboutir à des paradoxes.

Par exemple, il semblerait naturel de concevoir les choses de telle façon que la somme infinie

$$S = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \dots$$

soit égale à 2, car si on la double on obtient

$$2S = 2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots = 2 + S.$$

Pourtant, ce même raisonnement nous conduirait à affirmer que

$$T = 1 + 2 + 4 + 8 + 16 + 32 + \dots$$

Bien sûr :  $1 + 2 + 4 + 8 + \dots$  est la représentation binaire du nombre  $-1$  en "précision infinie" sur un ordinateur qui manipulerait des données de taille infinie.

est égal à  $-1$ , car en multipliant par 2 on a

$$2T = 2 + 4 + 8 + 16 + 32 + 64 + \dots = T - 1.$$

C'est bizarre : comment se fait-il qu'on obtienne un nombre négatif en additionnant des quantités positives ? Il serait peut-être préférable de considérer que  $T$  est indéfini ; ou alors on devrait dire que  $T = \infty$ , car on ajoute dans  $T$  des termes qui sont plus grands que n'importe quel nombre fixé (remarquez que  $\infty$  est aussi une "solution" de l'équation  $2T = T - 1$  ; ainsi d'ailleurs que de  $2S = 2 + S$ ).

Essayons de trouver une bonne définition de la valeur d'une somme générale  $\sum_{k \in K} a_k$ , où l'ensemble  $K$  peut être infini. Pour commencer, supposons que les termes  $a_k$  sont tous *positifs ou nuls*. Dans ce cas, il n'est pas difficile de trouver une définition convenable : s'il existe une constante  $A$  telle que

$$\sum_{k \in F} a_k \leq A$$

pour tout sous-ensemble fini  $F \subset K$ , alors  $\sum_{k \in K} a_k$  sera le *plus petit A possible* (l'ensemble de ces  $A$  a toujours un plus petit élément ; cela découle

de certaines propriétés bien connues des nombres réels). En revanche, si on ne peut pas trouver une telle borne  $A$ , alors on dira que  $\sum_{k \in K} a_k = \infty$ ; cela signifie que, pour tout nombre réel  $A$ , il existe un ensemble fini de termes  $a_k$  dont la somme excède  $A$ .

En formulant la définition du paragraphe précédent, nous nous sommes appliqués à ce qu'elle ne dépende pas d'un ordre qui pourrait exister sur les éléments de l'ensemble d'indices  $K$ . Ainsi, les résultats que nous allons énoncer pourront s'appliquer non seulement aux sommes sur l'ensemble des entiers, mais aussi aux sommes multiples, faisant appel à plusieurs indices  $k_1, k_2$  etc.

Dans le cas où  $K$  est l'ensemble des entiers positifs ou nuls, il découle de notre définition pour les termes  $a_k$  positifs ou nuls que

$$\sum_{k \geq 0} a_k = \lim_{n \rightarrow \infty} \sum_{k=0}^n a_k.$$

Voici pourquoi : toute suite croissante de nombres réels a une limite (qui peut être  $\infty$ ). Soit  $A$  cette limite, et soit  $F$  un ensemble quelconque d'entiers positifs ou nuls dont tous les éléments sont  $\leq n$ . On a bien sûr  $\sum_{k \in F} a_k \leq \sum_{k=0}^n a_k \leq A$ ; donc, soit  $A = \infty$ , soit  $A$  est un nombre qui majore  $\sum_{k \in F} a_k$  pour tout  $F$ . Soit maintenant  $A'$  un nombre plus petit que  $A$ ; alors il existe un entier  $n$  tel que  $\sum_{k=0}^n a_k > A'$ . Il suffit alors de prendre  $F = \{0, 1, \dots, n\}$  pour voir que  $A'$  n'est pas un majorant de  $\sum_{k \in F} a_k$  pour tout  $F$ .

La définition que nous venons de voir facilite grandement le calcul de certaines sommes infinies. Par exemple, si  $a_k = x^k$ , on a

$$\sum_{k \geq 0} x^k = \lim_{n \rightarrow \infty} \frac{1 - x^{n+1}}{1 - x} = \begin{cases} 1/(1-x), & \text{si } 0 \leq x < 1; \\ \infty, & \text{si } x \geq 1. \end{cases}$$

En particulier, les sommes infinies  $S$  et  $T$  que nous avons récemment rencontrées ont pour valeurs respectives  $2$  et  $\infty$ , exactement comme nous le soupçonnions. Voici un autre exemple intéressant :

$$\begin{aligned} \sum_{k \geq 0} \frac{1}{(k+1)(k+2)} &= \sum_{k \geq 0} k^{-2} \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^n k^{-2} = \lim_{n \rightarrow \infty} \frac{k^{-1}}{-1} \Big|_0^n = 1. \end{aligned}$$

Considérons maintenant le cas d'une somme qui peut avoir des termes négatifs aussi bien que positifs. Par exemple, quelle devrait être la valeur de

$$\sum_{k \geq 0} (-1)^k = 1 - 1 + 1 - 1 + 1 - 1 + \dots ?$$

*L'ensemble  $K$  peut même être non dénombrable. Si la borne  $A$  existe, le nombre de termes non nuls est forcément dénombrable, car  $nA$  termes au plus sont  $\geq 1/n$ .*

*“Aggregatum quantitatum  $a - a + a - a + a - a$  etc. nunc est =  $a$ , nunc = 0, adeoque continuata in infinitum serie ponendus =  $a/2$ , fateor acumen et veritatem animadversionis tuæ.”*

—G. Grandi [163]

En regroupant les termes par paires, on obtient

$$(1 - 1) + (1 - 1) + (1 - 1) + \cdots = 0 + 0 + 0 + \cdots,$$

et ainsi la somme s'annule ; mais si on n'apparie qu'à partir du deuxième terme, on a

$$1 - (1 - 1) - (1 - 1) - (1 - 1) - \cdots = 1 - 0 - 0 - 0 - \cdots;$$

et la somme fait 1.

On peut aussi essayer de poser  $x = -1$  dans la formule  $\sum_{k \geq 0} x^k = 1/(1-x)$ , puisqu'on sait qu'elle est correcte pour  $0 \leq x < 1$  ; mais alors on est forcé de conclure que cette somme infinie est égale à  $\frac{1}{2}$ , bien que ses termes soient tous entiers !

Voici un autre exemple intéressant : la somme doublement infinie  $\sum_k a_k$  où  $a_k = 1/(k+1)$  si  $k \geq 0$  et  $a_k = 1/(k-1)$  si  $k < 0$ . Elle peut s'écrire

$$\cdots + (-\frac{1}{4}) + (-\frac{1}{3}) + (-\frac{1}{2}) + 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots. \quad (2.58)$$

Si on évalue cette somme en partant du terme “central” et en allant vers les extrémités,

$$\cdots + \left( -\frac{1}{4} + \left( -\frac{1}{3} + \left( -\frac{1}{2} + (1) + \frac{1}{2} \right) + \frac{1}{3} \right) + \frac{1}{4} \right) + \cdots,$$

on obtient la valeur 1. On trouve cette même valeur 1 si on décale toutes les parenthèses d'un pas vers la gauche,

$$\cdots + \left( -\frac{1}{5} + \left( -\frac{1}{4} + \left( -\frac{1}{3} + \left( -\frac{1}{2} + 1 \right) + \frac{1}{2} \right) + \frac{1}{3} \right) + \frac{1}{4} \right) + \cdots,$$

car la somme de tous les nombres contenus dans les  $n$  parenthèses internes est

$$-\frac{1}{n+1} - \frac{1}{n} - \cdots - \frac{1}{2} + 1 + \frac{1}{2} + \cdots + \frac{1}{n-1} = 1 - \frac{1}{n} - \frac{1}{n+1}.$$

On peut montrer, avec un argument similaire, qu'on obtient toujours 1 si on décale les parenthèses de n'importe quel nombre fixé de pas ; ceci nous porte à croire que la somme vaut effectivement 1. D'un autre côté, si on regroupe les termes comme suit,

$$\cdots + \left( -\frac{1}{4} + \left( -\frac{1}{3} + \left( -\frac{1}{2} + 1 + \frac{1}{2} \right) + \frac{1}{3} + \frac{1}{4} \right) + \frac{1}{5} + \frac{1}{6} \right) + \cdots,$$

la  $n$ ième paire de parenthèses en partant de l'intérieur contient les nombres

$$-\frac{1}{n+1} - \frac{1}{n} - \cdots - \frac{1}{2} + 1 + \frac{1}{2} + \cdots + \frac{1}{2n-1} + \frac{1}{2n} = 1 + H_{2n} - H_{n+1}.$$

Nous montrerons au chapitre 9 que  $\lim_{n \rightarrow \infty} (H_{2n} - H_{n+1}) = \ln 2$  ; par conséquent cette façon de regrouper les termes suggère que la somme doublement infinie devrait être en réalité égale à  $1 + \ln 2$ .

Il est plutôt déconcertant de voir une somme donner des valeurs différentes selon la manière d'additionner ses termes. Les ouvrages d'analyse avancée proposent toute une variété de définitions permettant de donner des valeurs consistantes à des sommes pathologiques comme celle-là. Toutefois, si nous adoptons ces définitions, nous ne pourrons pas continuer à manipuler les  $\sum$  aussi librement que nous l'avons fait jusqu'ici. Pour atteindre les buts de ce livre, nous n'avons pas besoin des suprêmes raffinements de la "convergence conditionnelle" ; nous allons donc nous tenir à une définition des sommes infinies qui préserve la validité de toutes les opérations que nous avons effectuées dans ce chapitre.

En fait, notre définition des sommes infinies est très simple. Soit  $K$  un ensemble et  $a_k$  un terme à valeur réelle défini pour tout  $k \in K$  (en fait,  $K$  peut être multidimensionnel, donc " $k$ " peut désigner plusieurs indices  $k_1, k_2, \dots$ ). Tout nombre réel  $x$  peut s'écrire comme la différence de sa partie positive et sa partie négative,

$$x = x^+ - x^-, \quad \text{où } x^+ = x \cdot [x > 0] \text{ et } x^- = -x \cdot [x < 0].$$

(Il est clair que  $x^+ = 0$  ou bien  $x^- = 0$ ). Nous avons déjà vu comment on définit les valeurs des sommes  $\sum_{k \in K} a_k^+$  et  $\sum_{k \in K} a_k^-$ , car les  $a_k^+$  et les  $a_k^-$  sont positifs ou nuls. Voici par conséquent notre définition générale :

$$\sum_{k \in K} a_k = \sum_{k \in K} a_k^+ - \sum_{k \in K} a_k^-, \tag{2.59}$$

sauf si les deux sommes de droite sont égales à  $\infty$ . Dans ce dernier cas,  $\sum_{k \in K} a_k$  reste indéfinie.

Soit  $A^+ = \sum_{k \in K} a_k^+$  et  $A^- = \sum_{k \in K} a_k^-$ . Si  $A^+$  et  $A^-$  sont toutes deux finies, on dit que la somme  $\sum_{k \in K}$  converge absolument vers la valeur  $A = A^+ + A^-$ . Si  $A^+ = \infty$  et  $A^-$  est fini, on dit que la somme  $\sum_{k \in K}$  diverge vers  $+\infty$ . De façon similaire, si  $A^- = \infty$  et  $A^+$  est finie, on dit que  $\sum_{k \in K}$  diverge vers  $-\infty$ . Si  $A^+ = A^- = \infty$ , rien ne va plus.

*En d'autres termes, la convergence absolue signifie que la somme des valeurs absolues converge.*

Récapitulons : nous avons commencé par donner une définition valable pour les termes positifs ou nuls ; puis nous l'avons étendue à tous les termes à valeurs réelles. Si les termes  $a_k$  sont des nombres complexes, on peut évidemment étendre encore cette définition : la somme  $\sum_{k \in K} a_k$  est égale à  $\sum_{k \in K} \Re a_k + i \sum_{k \in K} \Im a_k$ , où  $\Re a_k$  et  $\Im a_k$  sont respectivement la partie réelle et la partie imaginaire de  $a_k$ . Cela n'est bien sûr valable que si les deux sommes sont définies ; dans le cas contraire,  $\sum_{k \in K} a_k$  est indéfinie (voir l'exercice 18).

Comme nous l'avons dit plus haut, il y a une mauvaise nouvelle : certaines sommes infinies doivent rester indéfinies, car les manipulations que nous faisons donneraient fatallement lieu à des incohérences (voir l'exercice 34). En revanche, la bonne nouvelle est que toutes les manipulations de ce chapitre sont parfaitement valides chaque fois qu'elles concernent des sommes qui convergent absolument, selon la définition que nous venons de donner. On peut vérifier cette bonne nouvelle en montrant que chacune de nos règles de transformation laisse inchangée la valeur de toute somme absolument convergente. Nous devons prouver cela pour les règles de distributivité, d'associativité et de commutativité ; et aussi pour la règle qui permet de sommer d'abord sur un indice donné. Tout ce que nous avons fait d'autre découlait de ces quatre opérations de base.

La règle de distributivité (2.15) peut être reformulée plus précisément comme suit : si  $\sum_{k \in K} a_k$  converge absolument vers  $A$  et si  $c$  est un nombre complexe, alors  $\sum_{k \in K} ca_k$  converge absolument vers  $cA$ . On peut le démontrer en séparant la somme en parties imaginaire et réelle, puis en parties positive et négative, comme ci-dessus. Après quoi on s'intéresse au cas où  $c > 0$  et chaque terme  $a_k$  est positif ou nul. La preuve de ce cas particulier fonctionne parce que  $\sum_{k \in F} ca_k = c \sum_{k \in F} a_k$  pour tout ensemble fini  $F$  ; cette dernière affirmation se prouve par induction sur la taille de  $F$ .

La règle d'associativité (2.16) peut s'écrire comme suit : si  $\sum_{k \in K} a_k$  et  $\sum_{k \in K} b_k$  convergent absolument vers  $A$  et  $B$  respectivement, alors la somme  $\sum_{k \in K} (a_k + b_k)$  converge absolument vers  $A + B$ . C'est en fait un cas particulier d'un théorème plus général que nous prouverons bientôt.

On n'a pas vraiment besoin de prouver la règle de commutativité (2.17) car nous avons montré lors de la discussion qui suit (2.35) comment la déduire d'un cas particulier d'une règle générale de changement de l'ordre de sommation.

Le principal résultat qu'il nous faut prouver est le principe fondamental des sommes multiples : *les sommes absolument convergentes sur deux ou plusieurs indices peuvent toujours être sommées d'abord sur l'un quelconque de leurs indices*. Formellement, nous allons prouver que si  $J$  et les éléments de  $\{K_j \mid j \in J\}$  sont des ensembles d'indices tels que

$$\sum_{\substack{j \in J \\ k \in K_j}} a_{j,k} \text{ converge absolument vers } A,$$

alors, pour chaque  $j \in J$ , il existe un nombre complexe  $A_j$  tel que

$$\sum_{k \in K_j} a_{j,k} \text{ converge absolument vers } A_j \text{ et}$$

$$\sum_{j \in J} A_j \text{ converge absolument vers } A.$$

*Si c'est la première fois que vous lisez cette page, je vous conseille de le faire en diagonale.*

— Votre sympathique chargé de TD.

Il suffit de montrer que cette assertion est vraie lorsque tous les termes sont positifs ou nuls, car on peut prouver le cas général en séparant tout en parties imaginaire et réelle, puis en parties positive et négative comme nous l'avons déjà vu. Supposons donc que  $a_{j,k} \geq 0$  pour tout couple  $(j, k) \in M$ , où  $M$  est l'ensemble d'indices  $\{(j, k) \mid j \in J, k \in K_j\}$ .

Pour tout ensemble fini  $F \subseteq M$ , nous savons que  $\sum_{(j,k) \in F} a_{j,k}$  est fini, donc que

$$\sum_{(j,k) \in F} a_{j,k} \leq A,$$

et que  $A$  est le plus petit de ces majorants (la borne supérieure). Si  $j$  est un élément quelconque de  $J$ , chaque somme de la forme  $\sum_{k \in F_j} a_{j,k}$ , où  $F_j$  est un sous-ensemble fini de  $K_j$ , est bornée par  $A$ . Donc ces sommes finies ont une borne supérieure  $A_j \geq 0$ , et  $\sum_{k \in K_j} a_{j,k} = A_j$  par définition.

Il nous faut encore prouver que  $A$  est la borne supérieure de  $\sum_{j \in G} A_j$  pour tout ensemble fini  $G \subseteq J$ . Supposons que  $G$  est un sous-ensemble fini de  $J$  et que  $\sum_{j \in G} a_{j,k} = A' > A$ . On peut trouver des sous-ensembles finis  $F_j \subseteq K_j$  tels que  $\sum_{k \in F_j} a_{j,k} > (A/A')A_j$  pour tout  $j \in G$  tel que  $A_j > 0$ . Il existe au moins un tel  $j$ . Mais alors  $\sum_{j \in G, k \in F_j} a_{j,k} > (A/A') \sum_{j \in G} A_j = A$ , ce qui contredit le fait que  $\sum_{(j,k) \in F} a_{j,k} \leq A$  pour tout sous-ensemble fini  $F \subseteq M$ . Donc  $\sum_{j \in G} A_j \leq A$  pour tout sous-ensemble fini  $G \subseteq J$ .

Pour finir, soit  $A'$  un nombre réel plus petit que  $A$ . Notre preuve sera complète si nous pouvons trouver un ensemble fini  $G \subseteq J$  tel que  $\sum_{j \in G} A_j > A'$ . Nous savons qu'il existe un ensemble fini  $F \subseteq M$  tel que  $\sum_{(j,k) \in F} a_{j,k} > A'$ ; soit  $G$  l'ensemble des  $j$  de ce  $F$ , et soit  $F_j = \{k \mid (j, k) \in F\}$ . Alors  $\sum_{j \in G} A_j \geq \sum_{j \in G} \sum_{k \in F_j} a_{j,k} = \sum_{(j,k) \in F} a_{j,k} > A'$ . CQFD.

Bien, nous sommes maintenant légitimés ! Tout ce que nous avons fait avec les sommes infinies est justifié, pourvu qu'il existe une borne finie pour toutes les sommes finies des valeurs absolues des termes. Puisque la somme doublement infinie (2.58) nous a donné deux réponses différentes selon la manière dont nous l'avons évaluée, ses termes positifs  $1 + \frac{1}{2} + \frac{1}{3} + \dots$  doivent forcément diverger vers  $\infty$ ; sinon, nous aurions obtenu la même réponse quelle que soit la façon de grouper les termes.

## Exercices

### Echauffements

1 Que signifie la notation

$$\sum_{k=4}^0 q_k ?$$

2 Simplifiez l'expression  $x \cdot ([x > 0] - [x < 0])$ .

- 3 Montrez que vous avez bien compris la notation  $\sum$  en développant entièrement les sommes

$$\sum_{0 \leq k \leq 5} a_k \quad \text{et} \quad \sum_{0 \leq k^2 \leq 5} a_{k^2}$$

(attention, la seconde somme est un peu délicate).

- 4 Ecrivez la somme triple

$$\sum_{1 \leq i < j < k \leq 4} a_{ijk}$$

comme une sommation en trois parties (avec trois  $\sum$ ),

- a en sommant d'abord sur  $k$ , puis sur  $j$ , enfin sur  $i$  ;  
 b en sommant d'abord sur  $i$ , puis sur  $j$ , enfin sur  $k$ .

Puis écrivez entièrement les sommes triples sans la notation  $\sum$ , en utilisant des parenthèses pour montrer ce qu'on ajoute d'abord.

- 5 Qu'est-ce qui cloche dans le calcul suivant ?

$$\left( \sum_{j=1}^n a_j \right) \left( \sum_{k=1}^n \frac{1}{a_k} \right) = \sum_{j=1}^n \sum_{k=1}^n \frac{a_j}{a_k} = \sum_{k=1}^n \sum_{j=1}^n \frac{a_k}{a_k} = \sum_{k=1}^n n = n^2.$$

- 6 Calculez  $\sum_k [1 \leq j \leq k \leq n]$  en fonction de  $j$  et  $n$ ?

- 7 Soit  $\nabla f(x) = f(x) - f(x-1)$ . Calculez  $\nabla(x^{\underline{m}})$ .

- 8 Si  $m$  est un entier, quelle est la valeur de  $0^{\underline{m}}$  ?

- 9 Trouvez un analogue de la règle (2.52) pour les puissances montantes. Utilisez cette règle pour définir  $x^{-n}$ .

- 10 Nous avons donné la formule suivante pour la différence d'un produit :

$$\Delta(uv) = u\Delta v + v\Delta u.$$

Comment peut-elle être correcte, alors que le membre de gauche est symétrique en  $u$  et  $v$  tandis que le membre de droite ne l'est pas ?

### Exercices de base

- 11 La règle de sommation par parties (2.56) est équivalente à

$$\begin{aligned} \sum_{0 \leq k < n} (a_{k+1} - a_k)b_k &= a_n b_n - a_0 b_0 \\ &\quad - \sum_{0 \leq k < n} a_{k+1}(b_{k+1} - b_k), \quad \text{pour } n \geq 0. \end{aligned}$$

Prouvez cette formule directement, en utilisant les règles de distributivité, d'associativité et de commutativité.

- 12 Montrez que, quelque soit l'entier  $c$ , la fonction  $p(k) = k + (-1)^k c$  est une permutation sur l'ensemble des entiers.
- 13 Utilisez la méthode du répertoire pour trouver une forme close de  $\sum_{k=0}^n (-1)^k k^2$ .
- 14 Calculez  $\sum_{k=1}^n k 2^k$  en la réécrivant  $\sum_{1 \leq j \leq k \leq n} 2^k$ .
- 15 Calculez  $\Re_n = \sum_{k=1}^n k^3$  par la méthode 5 en procédant ainsi : écrivez d'abord  $\Re_n + \square_n = 2 \sum_{1 \leq j \leq k \leq n} j k$ , puis appliquez (2.33).
- 16 Montrez que  $x^m/(x-n)^m = x^n/(x-m)^n$ , pourvu qu'aucun des dénominateurs ne soit nul.
- 17 Montrez qu'on peut utiliser les formules suivantes pour convertir factorielles montantes en factorielles descendantes et inversement ( $x^{-m}$  est défini par la réponse à l'exercice 9) :

$$\begin{aligned} x^{\overline{m}} &= (-1)^m (-x)^{\underline{m}} = (x+m-1)^{\underline{m}} = 1/(x-1)^{\overline{-m}}; \\ x^{\underline{m}} &= (-1)^m (-x)^{\overline{m}} = (x-m+1)^{\overline{m}} = 1/(x+1)^{\underline{-m}}. \end{aligned}$$

- 18 Soient  $\Re z$  et  $\Im z$  les parties réelle et imaginaire d'un nombre complexe  $z$ . Le module de  $z$ , noté  $|z|$ , est égal à  $\sqrt{(\Re z)^2 + (\Im z)^2}$ . On dit qu'une somme  $\sum_{k \in K} a_k$  de termes complexes  $a_k$  est absolument convergente si les sommes réelles  $\sum_{k \in K} \Re a_k$  et  $\sum_{k \in K} \Im a_k$  sont toutes deux absolument convergentes. Prouvez que  $\sum_{k \in K} a_k$  est absolument convergente si et seulement s'il existe une constante  $B$  telle que  $\sum_{k \in F} |a_k| \leq B$  pour tout sous-ensemble fini  $F \subseteq K$ .

### *Devoirs à la maison*

- 19 Utilisez un facteur de sommation pour résoudre la récurrence

$$\begin{aligned} T_0 &= 5; \\ 2T_n &= nT_{n-1} + 3 \cdot n!, \quad \text{pour } n > 0. \end{aligned}$$

- 20 Essayez de calculer  $\sum_{k=0}^n k H_k$  par la méthode de perturbation, mais au lieu de cela trouvez la valeur de  $\sum_{k=0}^n H_k$ .
- 21 Calculez les sommes  $S_n = \sum_{k=0}^n (-1)^{n-k}$ ,  $T_n = \sum_{k=0}^n (-1)^{n-k} k$  et  $U_n = \sum_{k=0}^n (-1)^{n-k} k^2$  avec la méthode de perturbation, en supposant que  $n \geq 0$ .
- 22 Prouvez l'*identité de Lagrange* (sans utiliser l'induction) :

$$\sum_{1 \leq j < k \leq n} (a_j b_k - a_k b_j)^2 = \left( \sum_{k=1}^n a_k^2 \right) \left( \sum_{k=1}^n b_k^2 \right) - \left( \sum_{k=1}^n a_k b_k \right)^2.$$

*Difficile de justifier l'identité de quelqu'un qui est mort depuis 175 ans.*

En fait, il s'agit de trouver une identité pour cette somme double plus générale :

$$\sum_{1 \leq j < k \leq n} (a_j b_k - a_k b_j)(A_j B_k - A_k B_j).$$

- 23 Calculez la somme  $\sum_{k=1}^n (2k+1)/k(k+1)$  de deux manières différentes :
- Développez  $1/k(k+1)$  en éléments simples :  $1/k(k+1) = 1/k - 1/(k+1)$ .
  - Sommez par parties.
- 24 Quelle est la valeur de  $\sum_{0 \leq k < n} H_k/(k+1)(k+2)$  ? *Suggestion :* Généralisez le calcul de (2.57).
- 25 La notation  $\prod_{k \in K} a_k$  désigne le produit de tous les nombres  $a_k$  pour tous les  $k \in K$ . Supposons pour simplifier que le nombre de  $k$  pour lesquels  $a_k \neq 1$  est fini ; on n'a donc pas besoin de définir les produits infinis. Quelles sont les règles, analogues à la distributivité, l'associativité et la commutativité pour les  $\sum$ , que cette notation  $\prod$  satisfait ?
- 26 Exprimez le produit double  $\prod_{1 \leq j < k \leq n} a_j a_k$  en fonction du produit simple  $\prod_{k=1}^n a_k$ , en manipulant la notation  $\prod$ . (Cet exercice nous donne un analogue de la formule du triangle supérieur (2.33)).
- 27 Calculez  $\Delta(c^\infty)$ , puis déduisez du résultat la valeur de  $\sum_{k=1}^n (-2)^k/k$ .
- 28 A quel moment le calcul suivant se fourvoie-t-il ?

$$\begin{aligned} 1 &= \sum_{k \geq 1} \frac{1}{k(k+1)} = \sum_{k \geq 1} \left( \frac{k}{k+1} - \frac{k-1}{k} \right) \\ &= \sum_{k \geq 1} \sum_{j \geq 1} \left( \frac{k}{j}[j=k+1] - \frac{j}{k}[j=k-1] \right) \\ &= \sum_{j \geq 1} \sum_{k \geq 1} \left( \frac{k}{j}[j=k+1] - \frac{j}{k}[j=k-1] \right) \\ &= \sum_{j \geq 1} \sum_{k \geq 1} \left( \frac{k}{j}[k=j-1] - \frac{j}{k}[k=j+1] \right) \\ &= \sum_{j \geq 1} \left( \frac{j-1}{j} - \frac{j}{j+1} \right) = \sum_{j \geq 1} \frac{-1}{j(j+1)} = -1. \end{aligned}$$

### Problèmes d'examen

- 29 Calculez la somme  $\sum_{k=1}^n (-1)^k k / (4k^2 - 1)$ .
- 30 Les joueurs de Cribbage savent depuis longtemps que  $15 = 7 + 8 = 4 + 5 + 6 = 1 + 2 + 3 + 4 + 5$ . Trouvez le nombre de façons d'écrire 1050 comme une somme d'entiers positifs consécutifs. (La représentation triviale

du nombre “1050” par lui-même compte pour une ; il y a donc, non pas trois, mais quatre façons d’écrire 15 comme une somme d’entiers positifs consécutifs. Ajoutons qu’il n’est pas besoin de connaître les règles du Cribbage pour résoudre ce problème).

- 31 La fonction zéta  $\zeta(k)$  de Riemann est définie comme suit :

$$1 + \frac{1}{2^k} + \frac{1}{3^k} + \cdots = \sum_{j \geq 1} \frac{1}{j^k}.$$

Montrez que  $\sum_{k \geq 2} (\zeta(k) - 1) = 1$ . Que vaut  $\sum_{k \geq 1} (\zeta(2k) - 1)$  ?

- 32 Soit  $a \doteq b = \max(0, a - b)$ . Prouvez que

$$\sum_{k \geq 0} \min(k, x \doteq k) = \sum_{k \geq 0} (x \doteq (2k + 1))$$

pour tout réel  $x \geq 0$ , et donnez une formule close pour ces sommes.

#### Questions subsidiaires

- 33 Soit  $\Lambda_{k \in K} a_k$  le minimum des  $a_k$  (ou leur borne inférieure si  $K$  est infini), en supposant que chaque  $a_i$  est soit réel, soit  $\pm\infty$ . Quelles sont les règles, analogues à celles qui régissent  $\sum$  et  $\prod$ , qui sont valides pour  $\bigwedge$  (voir l’exercice 25) ? *Les lois de la jungle*
- 34 Montrez que si une somme  $\sum_{k \in K} a_k$  est indéfinie dans le sens de (2.59), alors pour tout couple de nombres réels  $A^-$  et  $A^+$ , on peut trouver une suite de sous-ensembles finis  $F_1 \subset F_2 \subset F_3 \subset \cdots$  de  $K$  telle que

$$\sum_{k \in F_n} a_k \leq A^-, \text{ si } n \text{ est impair; } \sum_{k \in F_n} a_k \geq A^+, \text{ si } n \text{ est pair.}$$

- 35 Démontrez le théorème de Goldbach

$$1 = \frac{1}{3} + \frac{1}{7} + \frac{1}{8} + \frac{1}{15} + \frac{1}{24} + \frac{1}{26} + \frac{1}{31} + \frac{1}{35} + \cdots = \sum_{k \in P} \frac{1}{k-1},$$

où  $P$  est l’ensemble des “puissantes parfaites”, défini récursivement comme suit :

$$P = \{m^n \mid m \geq 2, n \geq 2, m \notin P\}.$$

*La puissance parfaite corrompt.  
Parfaitement !*

- 36 La “suite auto-descriptive”  $\langle f(1), f(2), f(3), \dots \rangle$  de Solomon Golomb est l’unique suite non décroissante d’entiers positifs qui contient exactement  $f(k)$  occurrences de chaque entier  $k$ . Quelques instants de réflexion suffisent pour trouver le début de la suite :

n	1	2	3	4	5	6	7	8	9	10	11	12
f(n)	1	2	2	3	3	4	4	4	5	5	5	6

Soit  $g(n)$  le plus grand entier  $m$  tel que  $f(m) = n$ . Montrez que

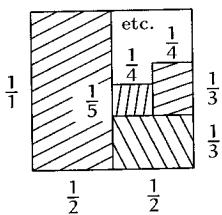
a  $g(n) = \sum_{k=1}^n f(k)$ .

b  $g(g(n)) = \sum_{k=1}^n kf(k)$ .

c  $g(g(g(n))) = \frac{1}{2}ng(n)(g(n)+1) - \frac{1}{2} \sum_{k=1}^{n-1} g(k)(g(k)+1)$ .

**Sujet de recherche**

- 37 Peut-on pavier un carré de côté 1 par tous les rectangles de dimensions  $1/k$  sur  $1/(k+1)$ , pour  $k \geq 1$  (souvenez-vous que la somme de leurs aires est égale à 1) ?



# 3

## Fonctions entières

---

LES NOMBRES ENTIERS sont l'épine dorsale des mathématiques discrètes. Ainsi, il existe beaucoup de situations dans lesquelles on a besoin de convertir des fractions ou des nombres réels en nombres entiers. Le but de ce chapitre est de nous amener à nous familiariser avec ces conversions. Nous apprendrons aussi quelques-unes de leurs remarquables propriétés.

### 3.1 PARTIES ENTIÈRES

Commençons par un compte-rendu sur les fonctions partie entière inférieure et partie entière supérieure, définies pour tout réel  $x$  comme ceci :

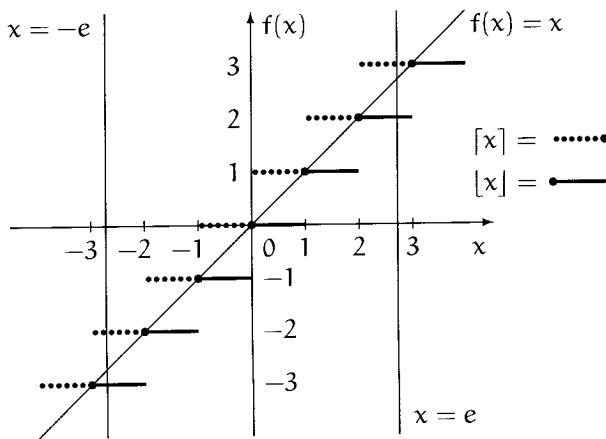
$$\begin{aligned} \lfloor x \rfloor &= \text{le plus grand entier inférieur ou égal à } x; \\ \lceil x \rceil &= \text{le plus petit entier supérieur ou égal à } x. \end{aligned} \tag{3.1}$$

C'est Kenneth E. Iverson qui a, au début des années 60 [191, page 12], introduit ces notations ainsi que les noms anglais “floor” (plancher) pour la partie entière inférieure et “ceiling” (plafond) pour la partie entière supérieure. Il a choisi ces symboles pour qu'ils puissent être imprimés en effaçant, selon le cas, le haut ou le bas de “[” et de “]”. Cette notation est maintenant assez répandue pour que les crochets de partie entière inférieure et supérieure puissent être utilisés dans un article scientifique sans que l'on rappelle leur signification. Il y a encore peu, on écrivait le plus souvent “[x]” pour la partie entière inférieure. Il n'y avait pas de notation spécifique pour la partie entière supérieure. Quelques auteurs avaient proposé “[x[”, sans succès évidemment.

)Aïe.(

Non seulement les notations, mais les fonctions elles-mêmes aussi peuvent varier. Par exemple, certaines calculettes possèdent une fonction INT, dont le résultat est  $\lfloor x \rfloor$  si  $x$  est positif et  $\lceil x \rceil$  si  $x$  est négatif. Peut-être les concepteurs de ces machines voulaient-ils que leur fonction INT satisfît  $\text{INT}(-x) = -\text{INT}(x)$ . Nous allons toutefois en rester aux fonctions que nous avons définies, car elles présentent des propriétés bien plus intéressantes.

Pour se familiariser avec les deux fonctions de partie entière, une bonne approche consiste à observer leurs graphes. Ils sont en forme d'escaliers, situés de part et d'autre de la droite d'équation  $f(x) = x$  :



Par exemple, on voit sur le graphe que, comme  $e = 2,71828\dots$ ,

$$\lfloor e \rfloor = 2, \quad \lceil -e \rceil = -3, \\ \lceil e \rceil = 3, \quad \lfloor -e \rceil = -2.$$

Cette figure va nous permettre de dégager un certain nombre de propriétés des parties entières. D'abord, puisque la courbe de la fonction de partie entière inférieure est en dessous de la diagonale  $f(x) = x$ , on a  $\lfloor x \rfloor \leq x$  ; de même,  $\lceil x \rceil \geq x$  (c'est bien sûr, quasiment évident de par définition). Les deux fonctions sont égales pour les valeurs entières, et elles seulement :

$$\lfloor x \rfloor = x \iff x \text{ est entier} \iff \lceil x \rceil = x.$$

(La notation “ $\iff$ ” signifie “si et seulement si”). De plus, lorsqu'elles diffèrent, leur différence est égale à 1 :

$$\lceil x \rceil - \lfloor x \rfloor = [x \text{ n'est pas un entier}]. \tag{3.2}$$

*Malin.  
Avec la convention  
d'Iverson, cette  
équation est tout  
à fait correcte.*

Si on décale la diagonale d'une unité vers le bas, elle se trouvera en dessous de la fonction de partie entière inférieure ; donc  $x - 1 < \lfloor x \rfloor$  et, de même,  $x + 1 > \lceil x \rceil$ . En combinant ces observations, on trouve

$$x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1. \tag{3.3}$$

Enfin, les deux fonctions sont symétriques l'une de l'autre par rapport à l'origine :

$$\lfloor -x \rfloor = -\lceil x \rceil; \quad \lceil -x \rceil = -\lfloor x \rceil. \tag{3.4}$$

Ainsi chacune d'entre elles s'exprime en fonction de l'autre. On comprend alors mieux pourquoi, à l'origine, seule la partie entière inférieure avait une notation spécifique. Cependant, la partie entière supérieure est maintenant assez répandue pour qu'on s'autorise à lui attribuer aussi une notation propre, tout comme nous avons adopté des notations spécifiques aussi bien pour les puissances montantes que pour les puissances descendantes. Depuis longtemps les mathématiciens disposent du sinus et du cosinus, de la tangente et de la cotangente, de la sécante et de la cosécante, du max et du min ; maintenant, nous avons aussi la partie entière inférieure et la partie entière supérieure.

Nous n'allons pas nous contenter d'observer les propriétés de ces deux fonctions sur un graphe ; nous allons les démontrer. Pour cela, les quatre propriétés suivantes nous seront bien utiles :

$$\begin{aligned} \lfloor x \rfloor = n &\iff n \leq x < n + 1, & (a) \\ \lfloor x \rfloor = n &\iff x - 1 < n \leq x, & (b) \\ \lceil x \rceil = n &\iff n - 1 < x \leq n, & (c) \\ \lceil x \rceil = n &\iff x \leq n < x + 1. & (d) \end{aligned} \tag{3.5}$$

où  $n$  est un entier et  $x$  un réel. Les règles (a) et (c) sont des conséquences immédiates de la définition (3.1) ; les règles (b) et (d) sont les mêmes respectivement que (a) et (c), sauf qu'on y a réorganisé les inégalités pour que  $n$  soit au milieu.

On peut faire entrer ou sortir un terme entier d'une partie entière (inférieure ou supérieure) :

$$\lfloor x + n \rfloor = \lfloor x \rfloor + n, \quad n \text{ entier}, \tag{3.6}$$

car, d'après la règle (3.5(a)), cette assertion est équivalente aux inégalités  $\lfloor x \rfloor + n \leq x + n < \lfloor x \rfloor + n + 1$ . Cependant, d'autres opérations similaires ne sont pas autorisées en général ; par exemple faire sortir un facteur constant :  $\lfloor nx \rfloor \neq n \lfloor x \rfloor$  lorsque  $n = 2$  et  $x = 1/2$ . On peut dire que les crochets de partie entière sont "rigides" en ce qui concerne la comparaison. En général, on s'estime heureux quand on peut s'en débarrasser ou quand on arrive à prouver quoi que ce soit quand ils sont présents.

Dans bien des cas, les crochets de partie entière sont redondants ; on peut alors en ajouter ou en supprimer à volonté. Par exemple, toute inégalité entre un réel et un entier est équivalente à une inégalité entre deux entiers :

$$\begin{aligned} x < n &\iff \lfloor x \rfloor < n, & (a) \\ n < x &\iff n < \lceil x \rceil, & (b) \\ x \leq n &\iff \lceil x \rceil \leq n, & (c) \\ n \leq x &\iff n \leq \lfloor x \rfloor. & (d) \end{aligned} \tag{3.7}$$

Ces règles se démontrent facilement. Par exemple, si  $x < n$ , alors  $\lfloor x \rfloor < n$  car  $\lfloor x \rfloor \leq x$ . Inversement, si  $\lfloor x \rfloor < n$ , alors nécessairement  $x < n$  puisque  $x < \lfloor x \rfloor + 1$  et  $\lfloor x \rfloor + 1 \leq n$ .

On aimerait bien que les quatre règles de (3.7) soient aussi faciles à retenir qu'à prouver. Chaque inégalité sans partie entière correspond à la même inégalité avec une partie entière inférieure ou supérieure ; mais il faut réfléchir à deux fois avant de décider laquelle des deux est appropriée.

La différence entre  $x$  et  $\lfloor x \rfloor$  s'appelle la *partie fractionnaire* de  $x$ . Elle apparaît assez souvent pour qu'on lui attribue une notation spécifique :

$$\{x\} = x - \lfloor x \rfloor. \quad (3.8)$$

On peut se permettre de dire simplement "partie entière" pour la partie entière inférieure  $\lfloor x \rfloor$ , car  $x = \lfloor x \rfloor + \{x\}$ . Si un nombre réel peut s'écrire sous la forme  $x = n + \theta$  où  $n$  est un entier et  $0 \leq \theta < 1$ , alors (3.5(a)) nous permet de conclure que  $n = \lfloor x \rfloor$  et  $\theta = \{x\}$ .

L'égalité (3.6) n'est pas vérifiée si  $n$  est un réel quelconque. Mais on peut déduire de ce qui précède qu'il y a seulement deux valeurs possibles pour l'expression  $\lfloor x+y \rfloor$  : si on écrit  $x = \lfloor x \rfloor + \{x\}$  et  $y = \lfloor y \rfloor + \{y\}$ , alors on a  $\lfloor x+y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor + \lfloor \{x\} + \{y\} \rfloor$  ; et comme  $0 \leq \{x\} + \{y\} < 2$ , on trouve que  $\lfloor x+y \rfloor$  peut être égal soit à  $\lfloor x+y \rfloor$  is  $\lfloor x \rfloor + \lfloor y \rfloor$ , soit à  $\lfloor x \rfloor + \lfloor y \rfloor + 1$ .

*Hum. Il vaut mieux éviter d'écrire  $\{x\}$  pour la partie fractionnaire quand on peut la confondre avec l'ensemble réduit à  $x$ .*

*Le second cas arrive si et seulement s'il y a une retenue au niveau du point décimal quand on additionne les deux parties fractionnaires  $\{x\}$  et  $\{y\}$ .*

## 3.2 APPLICATIONS

Nous venons de voir les outils de base pour manipuler les parties entières inférieure et supérieure. Mettons les maintenant en pratique, en commençant par un problème facile. Combien vaut  $\lceil \lg 35 \rceil$  (nous utilisons "lg" pour le logarithme en base 2, suivant en cela une prescription d'Edward M. Reingold) ? Eh bien, puisque  $2^5 < 35 \leq 2^6$ , on peut appliquer le logarithme pour trouver  $5 < \lg 35 \leq 6$  ; donc, d'après la relation (3.5(c)), on a  $\lceil \lg 35 \rceil = 6$ .

Notez que le nombre 35 s'écrit en base deux avec six bits :  $35 = (100011)_2$ . Peut-on dire que  $\lceil \lg n \rceil$  est toujours la longueur de l'écriture binaire de  $n$  ? Pas tout à fait. Il faut aussi six bits pour écrire  $32 = (100000)_2$  ; donc  $\lceil \lg n \rceil$  n'est pas la bonne réponse (il est vrai qu'elle n'échoue que lorsque  $n$  est une puissance de 2, mais ça fait quand même un nombre infini d'échecs). On peut trouver une réponse correcte en réalisant qu'il faut  $m$  bits pour écrire tout nombre  $n$  tel que  $2^{m-1} \leq n < 2^m$  ; alors (3.5(a)) nous indique que  $m-1 = \lfloor \lg n \rfloor$ , et donc  $m = \lceil \lg n \rceil + 1$ . En d'autres termes, il faut  $\lceil \lg n \rceil + 1$  bits pour écrire un nombre  $n$  en binaire, pour tout  $n > 0$ . L'expression  $\lceil \lg(n+1) \rceil$  est une autre solution, que l'on peut trouver de façon similaire. Cette dernière formule étant valable aussi pour  $n = 0$ , on peut dire qu'il faut zéro bit pour écrire  $n = 0$  en binaire.

Examinons maintenant des expressions avec plusieurs parties entières inférieure ou supérieure. Combien vaut  $\lceil \lfloor x \rfloor \rceil$  ? Facile : comme  $\lfloor x \rfloor$  est entier,  $\lceil \lfloor x \rfloor \rceil$  est tout simplement égal à  $\lfloor x \rfloor$ . Il en est ainsi de toute expression contenant un  $\lfloor x \rfloor$  intérieur, encadré par n'importe quel nombre de parties entières inférieures ou supérieures.

Voyons un problème plus ardu : démontrer ou réfuter l'assertion

$$\lceil \sqrt{\lfloor x \rfloor} \rceil = \lceil \sqrt{x} \rceil, \quad x \geq 0 \text{ réel.} \quad (3.9)$$

L'égalité est évidente si  $x$  est entier, car  $x = \lfloor x \rfloor$ . Il y a aussi égalité pour les cas particuliers  $\pi = 3,14159\dots$ ,  $e = 2,71828\dots$  et  $\phi = (1 + \sqrt{5})/2 = 1,61803\dots$ , car on obtient  $1 = 1$ . Notre incapacité à trouver un contre-exemple suggère que l'égalité est vraie en général. Essayons donc de la prouver.

Soit dit en passant, quand on se trouve face à un “démontrer ou réfuter”, il est généralement préférable d'essayer d'abord de réfuter avec un contre-exemple. C'est vrai pour deux raisons : d'une part il est a priori plus facile de réfuter (un contre-exemple suffit) ; d'autre part, le fait de chercher des poux dans une formule fait travailler notre imagination. Même si l'assertion donnée est vraie, la recherche d'un contre-exemple mène souvent à une preuve, dès lors que l'on voit pourquoi un contre-exemple est impossible. En outre, le scepticisme est une saine attitude.

Essayons de prouver que  $\lceil \sqrt{\lfloor x \rfloor} \rceil = \lceil \sqrt{x} \rceil$  en utilisant le calcul infinitésimal. Commençons par décomposer  $x$  en ses parties entière et fractionnaire  $\lfloor x \rfloor + \{x\} = n + \theta$ , puis développons la racine carrée en utilisant la formule du binôme :  $(n + \theta)^{1/2} = n^{1/2} + n^{-1/2}\theta/2 - n^{-3/2}\theta^2/8 + \dots$ . C'est une méthode pas très propre.

Il est bien plus facile d'utiliser les outils que nous avons développés. Voici une stratégie possible : trouver un moyen de sortir la partie entière inférieure la plus extérieure et la racine carrée de  $\lceil \sqrt{\lfloor x \rfloor} \rceil$ , puis supprimer la partie entière intérieure, enfin remettre en place la couche extérieure pour obtenir  $\lceil \sqrt{x} \rceil$ . Bon, allons-y. Prenons  $m = \lfloor \sqrt{\lfloor x \rfloor} \rceil$  et invoquons (3.5(a)) ; cela nous donne  $m \leq \sqrt{\lfloor x \rfloor} < m + 1$ . Nous avons supprimé la partie entière extérieure sans perdre une seule information. Comme les trois expressions sont positives ou nulles, on obtient  $m^2 \leq \lfloor x \rfloor < (m + 1)^2$  en élevant le tout au carré. Nous sommes débarrassés de la racine carrée. Maintenant, supprimons la partie entière inférieure en utilisant (3.7(d)) pour l'inégalité de gauche et (3.7(a)) pour celle de droite :  $m^2 \leq x < (m + 1)^2$ . C'est maintenant une simple formalité que de revenir sur nos pas, en prenant la racine carrée pour obtenir  $m \leq \sqrt{x} < m + 1$ , puis en invoquant (3.5(a)) pour avoir  $m = \lceil \sqrt{x} \rceil$ . Ainsi  $\lceil \sqrt{\lfloor x \rfloor} \rceil = m = \lceil \sqrt{x} \rceil$ . L'assertion est donc vraie. On peut prouver de façon similaire que

$$\lceil \sqrt{\lceil x \rceil} \rceil = \lceil \sqrt{x} \rceil, \quad x \geq 0 \text{ réel.}$$

*Bien sûr,  $\pi$ ,  $e$  et  $\phi$  sont évidemment les premiers réels qu'il faut tester. N'est-ce pas ?*

*Le scepticisme est sain seulement dans une certaine mesure. Etre sceptique sur les preuves ou les programmes (particulièrement les vôtres) vous permet probablement de garder un esprit sain et d'effectuer un travail assez sûr. Cependant, si vous êtes sceptique à ce point, vous risquez fort de travailler sans arrêt, sans jamais sortir vous aérer ou vous relaxer. Trop de scepticisme peut mener à la rigidité, un état tel que l'on devient si préoccupé d'être correct et rigoureux que l'on ne peut jamais rien achever.*

— Un sceptique

La preuve que nous venons de faire ne dépend pas fortement des propriétés des racines carrées. En regardant de plus près, on voit que l'idée peut être généralisée pour prouver plus encore : soit  $f(x)$  une fonction continue, strictement croissante et satisfaisant la propriété

$$f(x) \text{ est entier} \implies x \text{ est entier.}$$

(Le symbole “ $\implies$ ” signifie “implique”). Alors on a

*(Cela a été remarqué par R. J. McEliece, quand il était en licence).*

$[f(x)] = [f([x])]$  et  $[f(x)] = \lceil f(\lceil x \rceil) \rceil$  (3.10) dès lors que  $f(x)$ ,  $f([x])$  et  $f(\lceil x \rceil)$  sont définies. Nous allons démontrer cette propriété générale pour les parties entières supérieures, parce que jusqu'à présent nous ne nous sommes occupés que des parties entières inférieures. La preuve pour les parties entières inférieures est quasiment identique. Si  $x = \lceil x \rceil$ , on n'a rien à prouver. Sinon  $x < \lceil x \rceil$  et  $f(x) < f(\lceil x \rceil)$  car  $f$  est strictement croissante. Donc  $[f(x)] \leq [f(\lceil x \rceil)]$  car  $\lceil \cdot \rceil$  est croissante. Si  $[f(x)] < [f(\lceil x \rceil)]$ , alors, du fait que  $f$  est continue, il existe un nombre  $y$  tel que  $x \leq y < \lceil x \rceil$  et  $f(y) = [f(x)]$ . Ce nombre  $y$  est forcément entier en raison de la propriété particulière de  $f$ . Il y a là une contradiction, car il ne peut pas exister d'entier compris entre  $x$  et  $\lceil x \rceil$  strictement. Donc on a forcément  $[f(x)] = [f(\lceil x \rceil)]$ .

Voici un cas particulier important de ce théorème qui vaut la peine d'être écrit explicitement : si  $m$  est un entier et  $n$  un entier strictement positif, alors

$$\left\lfloor \frac{x+m}{n} \right\rfloor = \left\lfloor \frac{\lceil x \rceil + m}{n} \right\rfloor \quad \text{et} \quad \left\lceil \frac{x+m}{n} \right\rceil = \left\lceil \frac{\lceil x \rceil + m}{n} \right\rceil. \quad (3.11)$$

Par exemple, si  $m = 0$ , on a  $\lfloor \lfloor \lfloor x/10 \rfloor / 10 \rfloor / 10 \rfloor = \lfloor x/1000 \rfloor$ . Diviser trois fois par 10 et éliminer le reste à chaque division revient exactement à diviser une fois par 1000 puis supprimer le reste après coup.

Essayons maintenant de démontrer ou réfuter une autre assertion :

$$\lceil \sqrt{\lceil x \rceil} \rceil \stackrel{?}{=} \lceil \sqrt{x} \rceil, \quad x \geq 0 \text{ réel.}$$

Cela marche pour  $x = \pi$  ou  $x = e$ , mais pas pour  $x = \phi$  ; nous savons donc que l'assertion n'est pas vraie en général.

Avant d'aller plus loin, ouvrons une parenthèse pour passer en revue les différents niveaux de problèmes qui peuvent apparaître dans les livres de mathématiques :

**Niveau 1.** Etant donnés un objet  $x$  et une propriété explicite  $P(x)$ , montrer que  $P(x)$  est vraie. Par exemple, “montrez que  $\lfloor \pi \rfloor = 3$ ”. Le problème consiste donc à trouver une preuve d'un fait donné.

**Niveau 2.** Etant donnés un ensemble  $X$  et une propriété  $P(x)$ , montrer que  $P(x)$  est vraie pour tout  $x \in X$ . Par exemple, “montrez que  $\lfloor x \rfloor \leq x$  pour tout réel  $x$ ”. Il s’agit encore de trouver une preuve, mais cette fois-ci elle doit être générale. C’est de l’algèbre, et non simplement de l’arithmétique.

**Niveau 3.** Etant donnés un ensemble  $X$  et une propriété  $P(x)$ , prouver ou réfuter le fait que  $P(x)$  est vraie pour tout  $x \in X$ . Par exemple “prouvez ou réfutez l’assertion suivante :  $\lceil \sqrt{\lfloor x \rfloor} \rceil = \lceil \sqrt{x} \rceil$  pour tout réel  $x \geq 0$ ”. Il y a là un niveau d’incertitude supplémentaire : le résultat n’est pas suggéré par la question. Cette situation est plus proche de celles auxquelles les mathématiciens sont confrontés en permanence. Les assertions que l’on peut lire dans les livres sont vraies en général ; en revanche, les choses nouvelles doivent être étudiées d’un œil circonspect. Si l’assertion est fausse, notre travail consiste à trouver un contre-exemple. Si elle est vraie, il faut en trouver une preuve, tout comme au niveau 2.

**Niveau 4.** Etant donnés un ensemble  $X$  et une propriété  $P(x)$ , trouver une condition nécessaire et suffisante  $Q(x)$  pour que  $P(x)$  soit vraie. Par exemple, “donnez une condition nécessaire et suffisante pour que  $\lfloor x \rfloor \geq \lceil x \rceil$ ”. Le problème consiste à trouver  $Q$  telle que  $P(x) \iff Q(x)$ . Bien entendu, il y a toujours une réponse triviale : on n’a qu’à prendre  $P(x) = Q(x)$ . Cependant, ce qui est implicitement demandé, c’est une condition aussi simple que possible. Il faut faire preuve de créativité pour trouver une condition simple qui conviendra. (par exemple, dans le cas présent, “ $\lfloor x \rfloor \geq \lceil x \rceil \iff x$  est entier”). L’élément supplémentaire qu’est l’imagination nécessaire pour trouver  $Q(x)$  rend ce genre de problèmes plus difficile ; c’est aussi plus typique de ce que doivent faire les mathématiciens dans le “monde réel”. Pour finir, il faut évidemment donner une preuve du fait que  $P(x)$  est vraie si et seulement si  $Q(x)$  est vraie.

**Niveau 5.** Etant donné un ensemble  $X$ , trouver une propriété intéressante  $P(x)$  satisfaite par ses éléments. Ici on entre dans le domaine effrayant de la recherche pure, où les étudiants pourraient penser qu’il règne un chaos total. Les vraies mathématiques sont là. Les auteurs de manuels osent rarement poser des problèmes de niveau 5.

Fermons la parenthèse. Toutefois, nous allons accorder une promotion la dernière question que nous avons étudiée : elle va passer du niveau 3 au niveau 4. Cherchons une condition nécessaire et suffisante pour que  $\lceil \sqrt{\lfloor x \rfloor} \rceil = \lceil \sqrt{x} \rceil$ . Nous avons vu qu’il y a égalité quand  $x = 3,142$  mais pas quand  $x = 1,618$  ; d’autres expériences montrent qu’il n’y a pas égalité non plus quand  $x$  est compris entre 9 et 10. Ah ! On voit que l’égalité est fausse quand  $m^2 < x < m^2 + 1$ , car cela donne  $m$  dans le membre gauche et  $m + 1$  dans le membre droit. Dans tous les autres cas où  $\sqrt{x}$  est défini, c’est-à-dire lorsque  $x = 0$  ou  $m^2 + 1 \leq x \leq (m + 1)^2$ , il y a égalité. Voici donc une condition nécessaire et suffisante pour l’égalité : soit  $x$  est entier,

Dans mes autres manuels, “prouvez ou réfutez” semble avoir le même sens que “prouvez” dans 99,44% des cas environ ; mais pas dans ce livre.

soit  $\sqrt{\lfloor x \rfloor}$  ne l'est pas.

Avant d'étudier le problème suivant, nous allons voir une nouvelle notation bien pratique, suggérée par C. A. R. Hoare et Lyle Ramshaw, pour décrire les intervalles réels :  $[\alpha .. \beta]$  désigne l'ensemble des nombres réels  $x$  tels que  $\alpha \leq x \leq \beta$ . On appelle cet ensemble un *intervalle fermé* car il contient les deux extrémités  $\alpha$  et  $\beta$ . L'intervalle ne contenant aucune extrémité, que l'on note  $(\alpha .. \beta)$ , est l'ensemble des  $x$  tels que  $\alpha < x < \beta$  ; c'est un *intervalle ouvert*. On définit de même les intervalles *semi-ouverts*  $[\alpha .. \beta)$  et  $(\alpha .. \beta]$ , qui contiennent seulement une extrémité.

Ou semi-fermés pour les pessimistes.

Les lecteurs français sont peut-être plus familiarisés avec les notations  $[\alpha, \beta]$ ,  $]\alpha, \beta[$ ,  $[\alpha, \beta[,$   $]\alpha, \beta]$ , pour les intervalles fermé, ouvert, semi-ouvert à droite et semi-ouvert à gauche respectivement (N.d.T.).

Combien y a-t-il d'entiers dans de tels intervalles ? La réponse est plus facile pour les intervalles semi-ouverts que pour les autres. Commençons donc par les premiers. En fait, il est presque toujours plus agréable de manipuler les intervalles semi-ouverts que les intervalles ouverts ou fermés. Par exemple, ils sont additifs, au sens qu'on peut combiner les deux intervalles semi-ouverts  $[\alpha .. \beta)$  et  $[\beta .. \gamma)$  pour obtenir le semi-ouvert  $[\alpha .. \gamma)$ . Ceci ne marcherait pas avec les intervalles ouverts, car le point  $\beta$  serait exclu du résultat ; et pour les intervalles fermés, des problèmes pourraient survenir du fait que  $\beta$  serait compté deux fois.

Revenons à notre problème. La réponse est évidente si  $\alpha$  et  $\beta$  sont des entiers :  $[\alpha .. \beta)$  contient les  $\beta - \alpha$  entiers  $\alpha, \alpha + 1, \dots, \beta - 1$ , en supposant que  $\alpha \leq \beta$ . De la même manière,  $(\alpha .. \beta]$  contient aussi  $\beta - \alpha$  entiers. Mais notre problème est plus difficile car  $\alpha$  et  $\beta$  sont des réels quelconques. On peut néanmoins se ramener au problème plus simple, car, d'après (3.7),

$$\begin{aligned} \alpha \leq n < \beta &\iff \lceil \alpha \rceil \leq n < \lceil \beta \rceil, \\ \alpha < n \leq \beta &\iff \lfloor \alpha \rfloor < n \leq \lfloor \beta \rfloor, \end{aligned}$$

lorsque  $n$  est un entier. Les intervalles de droite ont des extrémités entières et contiennent le même nombre d'entiers que ceux de gauche, qui ont des extrémités réelles. Ainsi l'intervalle  $[\alpha .. \beta)$  contient exactement  $\lceil \beta \rceil - \lceil \alpha \rceil$  entiers, tandis que  $(\alpha .. \beta]$  en contient  $\lfloor \beta \rfloor - \lfloor \alpha \rfloor$ . Voici un cas où, au lieu d'essayer de nous débarrasser des crochets de partie entière, nous en introduisons à dessein.

A propos, il y a un truc pour retenir dans quels cas il faut utiliser la partie entière inférieure ou la partie entière supérieure : on rencontre plus souvent des intervalles semi-ouverts à droite (comme  $0 \leq \theta < 1$ ) que des intervalles semi-ouverts à gauche (comme  $0 < \theta \leq 1$ ) ; de même, les parties entières inférieures sont plus répandues que les parties entières supérieures. Donc, en vertu de la Loi de Murphy, la règle à suivre est le contraire de ce à quoi on s'attend. Il faut donc utiliser la partie entière supérieure pour  $[\alpha .. \beta)$  et la partie entière inférieure pour  $(\alpha .. \beta]$ .

Une réflexion similaire montre que l'intervalle fermé  $[\alpha .. \beta]$  contient exactement  $\lfloor \beta \rfloor - \lceil \alpha \rceil + 1$  entiers, tandis que l'intervalle ouvert  $(\alpha .. \beta)$  en

Exactement comme on retient l'année du départ de Christophe Colomb vers l'Amérique en chantant "In fourteen hundred and ninety-three/ Columbus sailed the deep blue sea."

contient  $\lceil \beta \rceil - \lfloor \alpha \rfloor - 1$ . Cependant, dans ce dernier cas, il faut ajouter la restriction  $\alpha \neq \beta$  pour que la formule ne puisse nous embarrasser en prétendant que l'intervalle vide  $(\alpha \dots \alpha)$  contient  $-1$  entiers. En résumé, nous avons établi les faits suivants :

intervalle	entiers contenus	restrictions	
$[\alpha \dots \beta]$	$\lfloor \beta \rfloor - \lceil \alpha \rceil + 1$	$\alpha \leq \beta$ ,	
$[\alpha \dots \beta)$	$\lceil \beta \rceil - \lceil \alpha \rceil$	$\alpha \leq \beta$ ,	(3.12)
$(\alpha \dots \beta]$	$\lfloor \beta \rfloor - \lfloor \alpha \rfloor$	$\alpha \leq \beta$ ,	
$(\alpha \dots \beta)$	$\lceil \beta \rceil - \lfloor \alpha \rfloor - 1$	$\alpha < \beta$ .	

Voici maintenant un problème que nous ne pouvons refuser d'étudier. Le Club des Mathématiques Concrètes possède un casino (où seuls sont admis les acquéreurs de ce livre) dans lequel tourne une roulette contenant mille cases numérotées de 1 à 1000. Si le numéro  $n$  qui est tiré lors d'un jeu est divisible par la partie entière inférieure de sa racine cubique, c'est-à-dire si

$$\lfloor \sqrt[3]{n} \rfloor \mid n,$$

alors c'est un numéro gagnant et l'établissement donne 5 dollars à chaque joueur ; sinon, le numéro est perdant et chaque joueur paie 1 dollar. (La notation  $a \backslash b$ , qui se dit "a divise b", signifie que b est un multiple de a ; cette relation est précisément étudiée au chapitre 4). Peut-on espérer gagner à ce jeu ?

On peut calculer le gain moyen, c'est-à-dire la somme qu'on gagne (ou qu'on perd) en moyenne lors d'une partie. Pour cela, on commence par compter le nombre  $W$  de numéros gagnants et le nombre  $L = 1000 - W$  de numéros perdants. Si chaque numéro apparaît une fois au cours de 1000 parties, alors on gagne  $5W$  dollars et on en perd  $L$ , donc le gain moyen sera

$$\frac{5W - L}{1000} = \frac{5W - (1000 - W)}{1000} = \frac{6W - 1000}{1000}.$$

S'il y a au moins 167 numéros gagnants, le joueur est avantagé ; sinon, l'avantage est du côté du casino.

Comment compter le nombre de numéros gagnants entre 1 et 1000 ? On peut facilement en donner une idée. Les numéros entre 1 et  $2^3 - 1 = 7$  sont tous gagnants car  $\lfloor \sqrt[3]{n} \rfloor = 1$  pour chacun d'eux. Parmi les nombres de  $2^3 = 8$  à  $3^3 - 1 = 26$ , seuls les numéros pairs sont gagnants. Parmi les numéros entre  $3^3 = 27$  et  $4^3 - 1 = 63$ , les gagnants sont ceux qui sont divisibles par 3, et ainsi de suite.

On peut utiliser les techniques de sommation du chapitre 2 pour analyser systématiquement le problème, en profitant des crochets d'Iverson pour

(D'après un sondage effectué à ce moment dans la classe, 28 étudiants pensent qu'il vaut mieux ne pas jouer, tandis que 13 veulent parler ; ceux qui n'ont rien compris s'abstiennent).

(Ils sont donc virés du Club).

les formules logiques :

$$\begin{aligned}
 W &= \sum_{n=1}^{1000} [n \text{ est gagnant}] \\
 &= \sum_{1 \leq n \leq 1000} [\lfloor \sqrt[3]{n} \rfloor \setminus n] = \sum_{k,n} [k = \lfloor \sqrt[3]{n} \rfloor] [k \setminus n] [1 \leq n \leq 1000] \\
 &= \sum_{k,m,n} [k^3 \leq n < (k+1)^3] [n = km] [1 \leq n \leq 1000] \\
 &= 1 + \sum_{k,m} [k^3 \leq km < (k+1)^3] [1 \leq k < 10] \\
 &= 1 + \sum_{k,m} [m \in [k^2 .. (k+1)^3/k]] [1 \leq k < 10] \\
 &= 1 + \sum_{1 \leq k < 10} (\lceil k^2 + 3k + 3 + 1/k \rceil - \lceil k^2 \rceil) \\
 &= 1 + \sum_{1 \leq k < 10} (3k + 4) = 1 + \frac{7+31}{2} \cdot 9 = 172.
 \end{aligned}$$

Ce calcul mérite d'être étudié soigneusement. Remarquez que la ligne 6 fait appel à notre formule (3.12) qui donne le nombre d'entiers contenus dans un intervalle semi-ouvert. La seule manœuvre "difficile" est la décision de traiter à part le cas  $n = 1000$  que nous avons prise entre les lignes 3 et 4. (L'inégalité  $k^3 \leq n < (k+1)^3$  ne se combine pas facilement avec  $1 \leq n \leq 1000$  lorsque  $k = 10$ ). De façon générale, les conditions aux bornes sont souvent les passages les plus délicats des manipulations de sommes.

Ça c'est bien vrai.

Où disiez-vous qu'il était, ce casino ?

La dernière ligne du calcul nous indique que  $W = 172$ . De ce fait on déduit que le gain moyen par partie est de  $(6 \cdot 172 - 1000)/1000$  dollars, ce qui nous fait 3,2 cents. On peut donc espérer être plus riche de 3,2 dollars après avoir fait 100 parties à 1 dollar (sauf bien sûr si le casino fait en sorte que certains numéros soient plus égaux que d'autres).

Le problème que nous venons de résoudre est une version "déguisée" d'une question plus abstraite : "parmi les entiers  $n$  tels que  $1 \leq n \leq 1000$ , combien satisfont la relation  $\lfloor \sqrt[3]{n} \rfloor \setminus n$ " ? Mathématiquement parlant, les deux questions sont les mêmes, mais il est parfois bon de déguiser ainsi un problème. Cela permet d'utiliser un vocabulaire plus riche (comme "numéro gagnant" ou "numéro perdant") qui peut nous aider à mieux comprendre ce qui se passe.

Voyons les choses de façon plus générale. Supposons que le 1000 devienne 1000000, ou même un nombre arbitraire  $N$ , très grand. (Nous supposerons que le casino a les moyens de se payer une roue plus grande). Combien y a-t-il alors de numéros gagnants ?

Les mêmes arguments sont valables, mais il nous faut manipuler avec

## 82 FONCTIONS ENTIÈRES

plus de précautions encore la plus grande valeur de  $k$ , que nous appellerons  $K$  :

$$K = \lfloor \sqrt[3]{N} \rfloor.$$

(Jusqu'ici,  $K$  était égal à 10). Le nombre total de numéros gagnants pour un  $N$  quelconque est

$$\begin{aligned} W &= \sum_{1 \leq k < K} (3k + 4) + \sum_m [K^3 \leq km \leq N] \\ &= \frac{1}{2}(7 + 3K + 1)(K - 1) + \sum_m [m \in [K^2 .. N/K]] \\ &= \frac{3}{2}K^2 + \frac{5}{2}K - 4 + \sum_m [m \in [K^2 .. N/K]]. \end{aligned}$$

Nous savons que la somme qui reste est  $[N/K] - [K^2] + 1 = [N/K] - K^2 + 1$ . Par conséquent la formule

$$W = [N/K] + \frac{1}{2}K^2 + \frac{5}{2}K - 3, \quad K = \lfloor \sqrt[3]{N} \rfloor \quad (3.13)$$

nous donne la réponse pour une roue de taille  $N$ .

Les deux premiers termes de cette formule valent approximativement  $N^{2/3} + \frac{1}{2}N^{2/3} = \frac{3}{2}N^{2/3}$ ; les deux autres termes sont bien plus petits en comparaison lorsque  $N$  est grand. Nous apprendrons au chapitre 9 à calculer des expressions du genre de

$$W = \frac{3}{2}N^{2/3} + O(N^{1/3}),$$

où  $O(N^{1/3})$  désigne une quantité qui n'est pas plus grande qu'un nombre constant de fois  $N^{1/3}$ . Quelle que soit cette constante multiplicative, nous savons qu'elle est indépendante de  $N$ ; donc, pour  $N$  grand, la contribution du terme  $O(N^{1/3})$  à  $W$  sera petite par rapport à  $\frac{3}{2}N^{2/3}$ . Le tableau qui suit illustre ce fait, en montrant comment  $\frac{3}{2}N^{2/3}$  se rapproche de  $W$  lorsque  $N$  croît :

$N$	$\frac{3}{2}N^{2/3}$	$W$	% erreur
1 000	150,0	172	12,791
10 000	696,2	746	6,670
100 000	3231,7	3343	3,331
1 000 000	15000,0	15247	1,620
10 000 000	69623,8	70158	0,761
100 000 000	323165,2	324322	0,357
1 000 000 000	1500000,0	1502497	0,166

C'est donc une très bonne approximation.

On utilise des formules approchées parce qu'elles sont plus simples que les formules contenant des parties entières. Cependant, il est souvent important de connaître l'exacte vérité, particulièrement pour les petites valeurs de  $N$ , qui apparaissent souvent en pratique. Par exemple, en utilisant la formule approchée, le propriétaire du casino aurait pu penser à tort qu'il y a seulement  $\frac{3}{2}N^{2/3} = 150$  numéros gagnants lorsque  $N = 1000$  (auquel cas la maison gagnerait en moyenne 10 cents par partie).

La dernière application de cette section concerne ce qu'on appelle les spectres. On appelle *spectre* d'un nombre réel  $\alpha$  le multi-ensemble d'entiers suivant :

$$\text{Spec}(\alpha) = \{\lfloor \alpha \rfloor, \lfloor 2\alpha \rfloor, \lfloor 3\alpha \rfloor, \dots\}.$$

(Un multi-ensemble est comme un ensemble, à ceci près qu'il peut contenir des mêmes éléments plusieurs fois). Voici par exemple le début du spectre de  $1/2$  :  $\{0, 1, 1, 2, 2, 3, 3, \dots\}$ .

On peut facilement montrer que si deux nombres  $\alpha$  et  $\beta$  sont différents, alors leurs spectres sont différents. Pour cela, on peut supposer sans perte de généralité que  $\alpha < \beta$ . Il existe donc un entier positif  $m$  tel que  $m(\beta - \alpha) \geq 1$ . (En fait, n'importe quel  $m \geq \lceil 1/(\beta - \alpha) \rceil$  peut convenir ; mais pour le moment nous n'avons pas besoin d'étaler notre science des parties entières). Donc  $m\beta - m\alpha \geq 1$ , et  $\lfloor m\beta \rfloor > \lfloor m\alpha \rfloor$ . On en conclut que  $\text{Spec}(\beta)$  contient moins de  $m$  éléments  $\leq \lfloor m\alpha \rfloor$ , tandis que  $\text{Spec}(\alpha)$  en contient au moins  $m$ .

Les spectres présentent un grand nombre de propriétés remarquables. Considérons par exemple les deux multi-ensembles

$$\text{Spec}(\sqrt{2}) = \{1, 2, 4, 5, 7, 8, 9, 11, 12, 14, 15, 16, 18, 19, 21, 22, 24, \dots\},$$

$$\text{Spec}(2 + \sqrt{2}) = \{3, 6, 10, 13, 17, 20, 23, 27, 30, 34, 37, 40, 44, 47, 51, \dots\}.$$

Une calculette suffit pour calculer  $\text{Spec}(\sqrt{2})$ , et d'après (3.6) le  $n$ ième élément de  $\text{Spec}(2 + \sqrt{2})$  s'obtient en ajoutant  $2n$  au  $n$ ième élément de  $\text{Spec}(\sqrt{2})$ . En regardant de plus près, on s'aperçoit que ces deux spectres sont même liés de façon bien plus surprenante : on dirait que chaque nombre qui manque dans l'un se trouve dans l'autre, et qu'aucun nombre n'est dans les deux à la fois ! Et c'est effectivement vrai : l'ensemble des entiers strictement positifs est l'union disjointe de  $\text{Spec}(\sqrt{2})$  et  $\text{Spec}(2 + \sqrt{2})$ . On dit que ces deux spectres forment une *partition* de l'ensemble des nombres entiers strictement positifs.

Pour prouver cette assertion, nous allons compter le nombre d'éléments de  $\text{Spec}(\sqrt{2})$  qui sont  $\leq n$  et le nombre d'éléments de  $\text{Spec}(2 + \sqrt{2})$  qui sont  $\leq n$ . Si la somme des deux est égale à  $n$  pour tout  $n$ , alors ces deux spectres forment en effet une partition des entiers.

... sans mettre de généralité...

"If  $x$  be an incommensurable number less than unity, one of the series of quantities  $m/x$ ,  $m/(1-x)$ , where  $m$  is a whole number, can be found which shall lie between any given consecutive integers, and but one such quantity can be found."

—Rayleigh [304]

C'est vrai, car, quand on ajoute 1 à  $n$ , on ajoute un nouvel élément à un seul des deux spectres.

Soit  $\alpha$  un réel strictement positif. Le nombre d'éléments de  $\text{Spec}(\alpha)$  qui sont  $\leq n$  est

$$\begin{aligned}
 N(\alpha, n) &= \sum_{k>0} [\lfloor k\alpha \rfloor \leq n] \\
 &= \sum_{k>0} [\lfloor k\alpha \rfloor < n+1] \\
 &= \sum_{k>0} [k\alpha < n+1] \\
 &= \sum_k [0 < k < (n+1)/\alpha] \\
 &= \lceil (n+1)/\alpha \rceil - 1. \tag{3.14}
 \end{aligned}$$

Il y a deux points particulièrement intéressants dans ce calcul. D'abord, on y utilise la règle

$$m \leq n \iff m < n+1, \quad m \text{ et } n \text{ entiers} \tag{3.15}$$

pour transformer " $\leq$ " en " $<$ ", de sorte qu'on puisse supprimer, en vertu de (3.7), les crochets de partie entière inférieure. D'autre part — et c'est plus subtil — on somme pour  $k > 0$  au lieu de  $k \geq 1$ , car  $(n+1)/\alpha$  pourrait être plus petit que 1 pour certaines valeurs de  $n$  et  $\alpha$ . Si nous avions appliqué (3.12) pour trouver le nombre d'entiers contenus dans  $[1 \dots (n+1)/\alpha]$  au lieu du nombre d'entiers contenus dans  $(0 \dots (n+1)/\alpha)$ , nous aurions aussi trouvé la bonne réponse ; mais notre calcul aurait été incorrect car les conditions d'application n'auraient pas été respectées.

Bon. Nous avons donc une formule pour  $N(\alpha, n)$ . Maintenant, voyons si  $\text{Spec}(\sqrt{2})$  et  $\text{Spec}(2 + \sqrt{2})$  forment une partition des entiers strictement positifs. Pour cela, nous allons regarder, en utilisant (3.14), si  $N(\sqrt{2}, n) + N(2 + \sqrt{2}, n) = n$  pour tout  $n > 0$  :

$$\begin{aligned}
 \left\lceil \frac{n+1}{\sqrt{2}} \right\rceil - 1 + \left\lceil \frac{n+1}{2+\sqrt{2}} \right\rceil - 1 &= n \\
 \iff \left\lfloor \frac{n+1}{\sqrt{2}} \right\rfloor + \left\lfloor \frac{n+1}{2+\sqrt{2}} \right\rfloor &= n, \quad \text{d'après (3.2)}; \\
 \iff \frac{n+1}{\sqrt{2}} - \left\{ \frac{n+1}{\sqrt{2}} \right\} + \frac{n+1}{2+\sqrt{2}} - \left\{ \frac{n+1}{2+\sqrt{2}} \right\} &= n, \quad \text{d'après (3.8)}.
 \end{aligned}$$

La formule se simplifie grâce à la jolie petite identité

$$\frac{1}{\sqrt{2}} + \frac{1}{2+\sqrt{2}} = 1,$$

et on n'a plus qu'à vérifier si

$$\left\{ \frac{n+1}{\sqrt{2}} \right\} + \left\{ \frac{n+1}{2+\sqrt{2}} \right\} = 1,$$

pour tout  $n > 0$ . C'est gagné, car le membre de gauche représente l'addition des parties fractionnaires de deux nombres non entiers dont la somme fait 1. C'est donc bien une partition.

### 3.3 RÉCURRENCES ET PARTIES ENTIÈRES

Avec les parties entières, on ajoute une nouvelle dimension à l'étude des relations de récurrence. Commençons par étudier la récurrence

$$\begin{aligned} K_0 &= 1; \\ K_{n+1} &= 1 + \min(2K_{\lfloor n/2 \rfloor}, 3K_{\lfloor n/3 \rfloor}), \quad \text{pour } n \geq 0. \end{aligned} \tag{3.16}$$

Ainsi, par exemple,  $K_1 = 1 + \min(2K_0, 3K_0) = 3$ . Les premiers termes de la suite sont 1, 3, 3, 4, 7, 7, 7, 9, 9, 10, 13, ... L'un des auteurs de cet ouvrage a modestement décidé d'appeler ces nombres les nombres de Knuth.

Dans l'exercice 25, il est demandé de prouver ou réfuter le fait que  $K_n \geq n$  pour tout  $n \geq 0$ . Comme les premiers  $K_i$  satisfont cette inégalité, il y a de bonnes chances pour que ce soit vrai en général. Essayons de le prouver par induction : on tire la base  $n = 0$  directement de la définition. Pour l'étape d'induction, supposons que l'inégalité est vraie pour tous les nombres compris entre 0 et un entier fixé  $n$  positif ou nul, et essayons de montrer que  $K_{n+1} \geq n+1$ . D'après la récurrence, nous savons que  $K_{n+1} = 1 + \min(2K_{\lfloor n/2 \rfloor}, 3K_{\lfloor n/3 \rfloor})$ . L'hypothèse d'induction nous indique que  $2K_{\lfloor n/2 \rfloor} \geq 2\lfloor n/2 \rfloor$  et  $3K_{\lfloor n/3 \rfloor} \geq 3\lfloor n/3 \rfloor$ . Cependant,  $2\lfloor n/2 \rfloor$  peut être égal à  $n-1$  et  $3\lfloor n/3 \rfloor$  peut être égal à  $n-2$ . Tout ce que nous pouvons donc conclure à partir de notre hypothèse d'induction, c'est que  $K_{n+1} \geq 1+(n-2)$ . On n'atteint pas de cette façon l'inégalité  $K_{n+1} \geq n+1$ .

Nous avons maintenant des raisons d'être moins sûrs que l'inégalité  $K_n \geq n$  soit vraie. Essayons donc de la réfuter. Si nous trouvons un  $n$  tel que soit  $2K_{\lfloor n/2 \rfloor} < n$ , soit  $3K_{\lfloor n/3 \rfloor} < n$ , nous aurons  $K_{n+1} < n+1$ . Est-ce possible ? Pour ne pas déflorer l'exercice 25, nous ne répondrons pas ici.

En informatique, on rencontre souvent des relations de récurrence qui font appel à des parties entières inférieures et/ou supérieures. C'est dû au fait que beaucoup d'algorithmes utilisent la précieuse technique "diviser pour régner" : il s'agit de résoudre un problème de taille  $n$  en le divisant en plusieurs problèmes similaires dont les tailles sont des fractions entières de  $n$ . Par exemple, pour trier  $n$  enregistrements ( $n > 1$ ), on peut les partager en deux parts à peu près égales, de tailles  $\lceil n/2 \rceil$  et  $\lfloor n/2 \rfloor$ . (Notons en

passant que

$$n = \lceil n/2 \rceil + \lfloor n/2 \rfloor; \quad (3.17)$$

cette formule pourra nous être utile). Une fois que chaque part a été triée séparément (en appliquant récursivement la même méthode), au plus  $n - 1$  comparaisons suffisent pour fusionner les enregistrements dans l'ordre. Ainsi le nombre total de comparaisons effectuées est au plus  $f(n)$ , avec

$$\begin{aligned} f(1) &= 0; \\ f(n) &= f(\lceil n/2 \rceil) + f(\lfloor n/2 \rfloor) + n - 1, \quad \text{pour } n > 1. \end{aligned} \quad (3.18)$$

On trouvera une solution de cette récurrence dans l'exercice 34.

Le problème de Josèphe que nous avons vu au chapitre 1 contient une récurrence similaire, que l'on peut écrire ainsi :

$$\begin{aligned} J(1) &= 1; \\ J(n) &= 2J(\lfloor n/2 \rfloor) - (-1)^n, \quad \text{pour } n > 1. \end{aligned}$$

Avec les outils dont nous disposons maintenant, nous pouvons considérer la version plus authentique du problème, dans laquelle, au lieu d'éliminer une personne sur deux, on en élimine une sur trois. Si on applique la méthode du chapitre 1 à ce problème plus difficile, on arrive à la récurrence

$$J_3(n) = \left\lceil \frac{3}{2} J_3(\lfloor \frac{2}{3}n \rfloor) + a_n \right\rceil \bmod n + 1,$$

où "mod" est une fonction que nous décrirons d'ici peu, et  $a_n = -2, +1$  ou  $-\frac{1}{2}$  selon que  $n \bmod 3 = 0, 1$  ou  $2$ . Cette récurrence est trop affreuse pour que nous continuions ainsi ; tentons une autre approche. Chaque fois qu'on compte une personne, on peut lui attribuer un nouveau numéro. Ainsi, 1 et 2 deviennent  $n + 1$  et  $n + 2$ , puis 3 est exécuté ; 4 et 5 deviennent  $n + 3$  et  $n + 4$ , puis 6 est exécuté ; ... ;  $3k + 1$  et  $3k + 2$  deviennent  $n + 2k + 1$  et  $n + 2k + 2$ , puis  $3k + 3$  est exécuté ; ... enfin  $3n$  est exécuté (ou laissé en vie). Par exemple, voici la suite des numéros successifs pour  $n = 10$  :

1	2	3	4	5	6	7	8	9	10
11	12		13	14		15	16		17
18			19	20			21		22
			23	24				25	
			26					27	
			28						
			29						
			30						

La  $k$ ième personne éliminée finit avec le numéro  $3k$ . On peut donc prédire qui sera le survivant si on sait trouver le numéro original de la personne qui aura finalement le numéro  $3n$ .

Si  $N > n$ , la personne numéro  $n$  a forcément eu un autre numéro précédemment. On peut le trouver de la façon suivante : on sait que  $N = n + 2k + 1$  ou  $N = n + 2k + 2$ , donc  $k = \lfloor (N - n - 1)/2 \rfloor$ . Le numéro précédent de la personne considérée était  $3k + 1$  ou  $3k + 2$  respectivement. Il était donc égal à  $3k + (N - n - 2k) = k + N - n$ . On peut alors calculer le numéro  $J_3(n)$  du survivant comme suit :

```

N := 3n;
tant que N > n faire N := ⌊(N - n - 1) / 2⌋ + N - n;
J3(n) := N.
```

Ce n'est pas une formule close ; ce n'est pas même une récurrence ; mais ça nous dit au moins comment trouver assez rapidement la réponse si  $n$  est grand.

*"Not too slow,  
not too fast."*

—L. Armstrong

Par bonheur, on peut simplifier cet algorithme en utilisant la variable  $D = 3n + 1 - N$  au lieu de  $N$ . (Ce changement de variable signifie que l'on numérote les participants en décroissant de  $3n$  vers 1, au lieu de les numérotter de 1 à  $3n$  ; on fait en quelque sorte un compte à rebours). Alors l'affectation compliquée de  $N$  se simplifie en

$$\begin{aligned}
D &:= 3n + 1 - \left( \left\lfloor \frac{(3n + 1 - D) - n - 1}{2} \right\rfloor + (3n + 1 - D) - n \right) \\
&= n + D - \left\lfloor \frac{2n - D}{2} \right\rfloor = D - \left\lfloor \frac{-D}{2} \right\rfloor = D + \left\lceil \frac{D}{2} \right\rceil = \left\lceil \frac{3}{2}D \right\rceil,
\end{aligned}$$

et on peut réécrire l'algorithme ainsi :

```

D := 1;
tant que D ≤ 2n faire D := ⌈ 3/2 D ⌉;
J3(n) := 3n + 1 - D.
```

Bon ! C'est bien mieux qu'avant, car  $n$  participe au calcul de façon très simple. Avec ce même raisonnement, on peut montrer que, si on élimine chaque  $q$ ième personne, alors le survivant  $J_q(n)$  se calcule comme suit :

```

D := 1;
tant que D ≤ (q - 1)n faire D := ⌈ q/(q-1) D ⌉; (3.19)
Jq(n) := qn + 1 - D.
```

Ainsi, dans le cas déjà connu où  $q = 2$ ,  $D$  croît jusqu'à  $2^{m+1}$  lorsque  $l = 2^m + l$ ; donc  $J_2(n) = 2(2^m + l) + 1 - 2^{m+1} = 2l + 1$ . Parfait.

Avec la recette de (3.19), on calcule une suite d'entiers qui peut être définie par la récurrence suivante :

$$\begin{aligned} D_0^{(q)} &= 1; \\ D_n^{(q)} &= \left\lceil \frac{q}{q-1} D_{n-1}^{(q)} \right\rceil \quad \text{pour } n > 0. \end{aligned} \tag{3.20}$$

Apparemment, ces nombres ne sont pas liés à des fonctions connues, sauf lorsque  $q = 2$ ; ils n'ont donc probablement pas de jolie forme close. Cependant, si on considère comme "connue" la suite  $D_n^{(q)}$ , alors on peut facilement décrire la solution du problème de Josèphe généralisé : le survivant  $J_q(n)$  est  $qn + 1 - D_k^{(q)}$ , où  $k$  est le plus petit entier tel que  $D_k^{(q)} > (q-1)n$ .

"Connue" comme, par exemple, la suite des nombres harmoniques. A.M. Odlyzko et H.S. Wilf ont montré [283] que  $D_n^{(3)} = \left\lfloor \left(\frac{3}{2}\right)^n C \right\rfloor$ , où  $C \approx 1,6222705$ .

### 3.4 "MOD" : L'OPÉRATION BINAIRE

Si  $m$  et  $n$  sont des entiers strictement positifs, le quotient de la division de  $n$  par  $m$  est  $\lfloor n/m \rfloor$ . Le reste de cette division est noté " $n$  mod  $m$ ". La formule de base

$$n = m \underbrace{\lfloor n/m \rfloor}_{\text{quotient}} + \underbrace{n \bmod m}_{\text{reste}}$$

indique que  $n \bmod m$  est égal à  $n - m \lfloor n/m \rfloor$ . Ceci peut être généralisé aux entiers négatifs, et même à tous les nombres réels :

$$x \bmod y = x - y \lfloor x/y \rfloor, \quad \text{pour } y \neq 0. \tag{3.21}$$

Ainsi, "mod" est une opération binaire, tout comme l'addition et la soustraction. Les mathématiciens utilisent mod ainsi depuis longtemps, en manipulant des nombres mod 10 ou mod  $2\pi$  par exemple ; mais cette notion n'est formellement définie que depuis une vingtaine d'années. Vieille notion, jeune notation.

Pourquoi appellent-ils mod "l'opération binaire" ? Réponse dans le passionnant prochain chapitre.

La signification intuitive de  $x \bmod y$  est facile à saisir lorsque  $x$  et  $y$  sont des entiers positifs. Imaginons un cercle de circonférence  $y$  dont les points représentent les nombres réels de l'intervalle  $[0 \dots y)$ . Si on suit la circonférence sur une distance  $x$  en partant de 0, le point où on s'arrête est  $x \bmod y$  (et le nombre de fois qu'on est passé au point 0 est égal à  $\lfloor x/y \rfloor$ ).

Si l'un des nombres  $x$  ou  $y$  est négatif, il faut étudier de plus près la définition pour voir exactement ce qu'elle signifie. Voici quelques exemples :

$$5 \bmod 3 = 5 - 3 \lfloor 5/3 \rfloor = 2;$$

Attention, certains langages informatiques utilisent une définition différente

$$\begin{aligned} 5 \bmod -3 &= 5 - (-3)\lfloor 5/(-3) \rfloor = -1; \\ -5 \bmod 3 &= -5 - 3\lfloor -5/3 \rfloor = 1; \\ -5 \bmod -3 &= -5 - (-3)\lfloor -5/(-3) \rfloor = -2. \end{aligned}$$

*Que diriez-vous de l’appeler le modumeur ?*

Le nombre situé après “mod” est appelé le *module* ; personne n’a encore donné un nom au nombre avant “mod”. En pratique, le module est généralement positif, mais la définition garde tout son sens quand il est négatif. Dans les deux cas, la valeur de  $x \bmod y$  est entre 0 et le module :

$$\begin{aligned} 0 \leq x \bmod y < y, &\quad \text{pour } y > 0; \\ 0 \geq x \bmod y > y, &\quad \text{pour } y < 0. \end{aligned}$$

Que se passe-t-il si  $y = 0$  ? Pour éviter de diviser par 0, ce cas a été évité dans la définition (3.21). Toutefois, pour être complet, on peut convenir que

$$x \bmod 0 = x. \tag{3.22}$$

Ainsi on préserve le fait que la différence entre  $x \bmod y$  et  $x$  est un multiple de  $y$ . (Il pourrait sembler plus naturel de faire en sorte que la fonction soit continue en 0 en posant  $x \bmod 0 = \lim_{y \rightarrow 0} x \bmod y = 0$  ; mais nous verrons au chapitre 4 que ce serait moins intéressant. La continuité n’est pas un élément très important dans l’opération mod).

Nous avons déjà vu, dans une version déguisée, un cas particulier de mod : lorsque nous avons écrit  $x$  en fonction de ses parties entière et fractionnaire,  $x = \lfloor x \rfloor + \{x\}$ . La partie fractionnaire peut aussi s’écrire  $x \bmod 1$  car on a

$$x = \lfloor x \rfloor + x \bmod 1.$$

Notez qu’on n’a pas besoin de mettre des parenthèses dans cette formule : on considère que mod est prioritaire face à l’addition ou à la soustraction.

Pour définir mod, nous avons fait appel à la partie entière inférieure. La partie entière supérieure doit légitimement se sentir frustrée. Peut-être pourrait-on l’utiliser pour définir un analogue de mod, comme ceci :

$$x \text{ marmot } y = y \lceil x/y \rceil - x.$$

Dans le cercle défini plus haut, cela représente la distance qui reste à parcourir pour atteindre le point 0 après avoir parcouru une distance  $x$ . Bien entendu, il faudrait trouver un terme plus adéquat que “marmot”. Si nous trouvons des applications pratiques de cette opération, il est probable qu’un terme approprié surgira naturellement.

*A un moment dans les années 70, il fallait être “mod” pour être à la mode. Peut-être qu’on pourrait rebaptiser “punk” cette fonction marmot ?*

*Non ! j’aime “marmot”.*

La propriété algébrique la plus importante de mod est sa distributivité :  
on a

$$c(x \bmod y) = (cx) \bmod (cy) \quad (3.23)$$

pour tous réels  $c$ ,  $x$  et  $y$ . (Ceux qui considèrent que la multiplication est prioritaire par rapport à mod peuvent ôter les parenthèses du membre de droite). Cette loi est facile à prouver à partir de la définition (3.21), car

$$c(x \bmod y) = c(x - y\lfloor x/y \rfloor) = cx - cy\lfloor cx/cy \rfloor = cx \bmod cy$$

si  $cy \neq 0$ , et la formule est trivialement vraie si le module est nul. Si on prend  $c = -1$ , la distributivité est illustrée deux fois dans nos quatre exemples avec  $\pm 5$  et  $\pm 3$ . L'identité (3.23) est rassurante, en ce sens qu'elle nous donne des raisons de penser que "mod" n'a pas été indûment définie.

Dans le reste de cette section, nous allons voir une application pratique dans laquelle "mod" est fort utile, bien que ne jouant pas un rôle de premier plan. Ce problème arrive couramment dans des situations variées : on veut répartir  $n$  objets en  $m$  groupes aussi égaux que possible.

Ah oui, le reste.  
Ha ! ha !

Supposons par exemple que nous voulions mettre en page  $n$  courtes lignes de texte dans  $m$  colonnes. Pour des raisons d'esthétique, nous voulons que les colonnes soient positionnées dans l'ordre des longueurs décroissantes. De plus, les longueurs des colonnes doivent être à peu près égales : les longueurs de deux colonnes distinctes ne doivent pas différer de plus d'une ligne de texte. Par exemple, s'il faut diviser 37 lignes de texte en cinq colonnes, on préférera la disposition de droite :

8	8	8	8	5	8	8	7	7	7
ligne 1	ligne 9	ligne 17	ligne 25	ligne 33	ligne 1	ligne 9	ligne 17	ligne 24	ligne 31
ligne 2	ligne 10	ligne 18	ligne 26	ligne 34	ligne 2	ligne 10	ligne 18	ligne 25	ligne 32
ligne 3	ligne 11	ligne 19	ligne 27	ligne 35	ligne 3	ligne 11	ligne 19	ligne 26	ligne 33
ligne 4	ligne 12	ligne 20	ligne 28	ligne 36	ligne 4	ligne 12	ligne 20	ligne 27	ligne 34
ligne 5	ligne 13	ligne 21	ligne 29	ligne 37	ligne 5	ligne 13	ligne 21	ligne 28	ligne 35
ligne 6	ligne 14	ligne 22	ligne 30		ligne 6	ligne 14	ligne 22	ligne 29	ligne 36
ligne 7	ligne 15	ligne 23	ligne 31		ligne 7	ligne 15	ligne 23	ligne 30	ligne 37
ligne 8	ligne 16	ligne 24	ligne 32		ligne 8	ligne 16			

En outre, si on considère que les lignes sont numérotées, on veut que deux lignes successives d'une même colonne aient des numéros successifs, car c'est dans ce sens qu'on lit en général. Il faudra donc d'abord décider combien de lignes iront dans la première colonne, puis considérer la deuxième colonne, puis la troisième et ainsi de suite. Si on distribuait les lignes rangée par rangée, on aurait un bon nombre de lignes dans chaque colonne, mais l'ordre des lignes serait incorrect (on obtiendrait quelque chose ressemblant à la disposition de droite, mais la colonne 1 contiendrait les lignes 1, 6, 11, ..., 36 au lieu de 1, 2, 3, ... 8).

On ne peut donc utiliser une stratégie de rangement "rangée par rangée". Cependant cette stratégie peut quand même nous indiquer combien

il faut mettre de lignes dans chaque colonne. Si  $n$  n'est pas un multiple de  $m$ , il est clair que cette procédure aurait pour résultat de mettre  $\lceil n/m \rceil$  lignes dans les colonnes les plus longues et  $\lfloor n/m \rfloor$  lignes dans les courtes (et il y aurait exactement  $n$  marmots  $m$  colonnes courtes).

Pour généraliser notre propos, parlons dorénavant d’“objets” et de “groupes” plutôt que de “lignes” et de “colonnes”. Nous venons de décider que le premier groupe contiendrait  $\lceil n/m \rceil$  objets. Pour distribuer séquentiellement  $n$  objets dans  $m$  urnes (avec  $m > 0$ ), on doit donc pouvoir appliquer le plan suivant : mettre  $\lceil n/m \rceil$  objets dans un groupe, puis appliquer récursivement la même procédure pour ranger les  $n' = n - \lceil n/m \rceil$  objets restant dans les  $m' = m - 1$  groupes restant.

Par exemple, si  $n = 314$  et  $m = 6$ , la distribution évolue de la manière suivante :

objets restant	groupes restant	$\lceil \text{objets/groupe} \rceil$
314	6	53
261	5	53
208	4	52
156	3	52
104	2	52
52	1	52

Ça marche. On obtient bien des groupes de taille à peu près égale, même si le diviseur varie au cours du temps.

Au fait, pourquoi est-ce que ça marche ? Posons  $n = qm + r$ , où  $q = \lfloor n/m \rfloor$  et  $r = n \bmod m$ . Si  $r = 0$ , le procédé est simple : on met  $\lceil n/m \rceil = q$  objets dans le premier groupe et on remplace  $n$  par  $n' = n - q$ . Il reste alors  $n' = qm'$  objets à ranger dans  $m' = m - 1$  groupes. D'autre part, si  $r > 0$ , on met  $\lceil n/m \rceil = q + 1$  objets dans le premier groupe et on remplace  $n$  par  $n' = n - q - 1$ , ce qui laisse  $n' = qm' + r - 1$  objets à répartir dans les autres groupes. Le nouveau reste est  $r' = r - 1$ , mais  $q$  est inchangé. Il s'ensuit qu'il y aura  $r$  groupes de  $q + 1$  objets suivis de  $m - r$  groupes de  $q$  objets.

Combien y a-t-il d’objets dans le  $k$ ième groupe ? Il nous faudrait une formule donnant  $\lceil n/m \rceil$  si  $k \leq n \bmod m$  et  $\lfloor n/m \rfloor$  sinon. Il n'est pas difficile de voir que

$$\left\lceil \frac{n - k + 1}{m} \right\rceil$$

convient, car c'est égal à  $q + \lceil (r - k + 1)/m \rceil$  si on écrit, comme dans le paragraphe précédent,  $n = qm + r$ , avec  $q = \lfloor n/m \rfloor$ . On sait que  $\lceil (r - k + 1)/m \rceil = [k \leq r]$  si  $1 \leq k \leq m$  et  $0 \leq r < m$ . Nous pouvons donc donner une identité qui décrit la partition de  $n$  en  $m$  parts “les plus égales

possibles" :

$$n = \left\lfloor \frac{n}{m} \right\rfloor + \left\lfloor \frac{n-1}{m} \right\rfloor + \cdots + \left\lfloor \frac{n-m+1}{m} \right\rfloor. \quad (3.24)$$

Cette identité est valable pour tout entier strictement positif  $m$  et pour tout entier  $n$  (positif, négatif ou nul). Nous avons déjà rencontré le cas  $m = 2$  en (3.17) sous une forme un peu différente :  $n = \lceil n/2 \rceil + \lfloor n/2 \rfloor$ .

Si nous avions voulu que les parts soient en ordre de tailles décroissantes nous aurions procédé de manière similaire, la seule différence étant que nous aurions mis  $\lfloor n/m \rfloor$  objets dans le premier groupe. Nous aurions alors trouvé l'identité correspondante :

$$n = \left\lfloor \frac{n}{m} \right\rfloor + \left\lfloor \frac{n+1}{m} \right\rfloor + \cdots + \left\lfloor \frac{n+m-1}{m} \right\rfloor. \quad (3.25)$$

On peut passer de (3.25) à (3.24) et réciproquement en utilisant, au choix, (3.4) ou l'identité de l'exercice 12.

Maintenant, en remplaçant le  $n$  de (3.25) par  $\lfloor mx \rfloor$  et en appliquant la règle (3.11) pour supprimer les parties entières à l'intérieur d'autres parties entières, on obtient une identité valable pour tout réel  $x$  :

$$\lfloor mx \rfloor = \lfloor x \rfloor + \left\lfloor x + \frac{1}{m} \right\rfloor + \cdots + \left\lfloor x + \frac{m-1}{m} \right\rfloor. \quad (3.26)$$

C'est assez étonnant. En effet, nous savons bien que la fonction partie entière inférieure donne une approximation entière d'une valeur réelle ; mais ce qui se passe ici, c'est que l'unique approximation de gauche donne le même résultat que la somme des approximations de droite. Si on considère que  $\lfloor x \rfloor$  est en moyenne à peu près égal à  $x - \frac{1}{2}$ , le membre de gauche vaut à peu près  $mx - \frac{1}{2}$ , tandis que celui de droite donne, en gros,  $(x - \frac{1}{2}) + (x - \frac{1}{2} + \frac{1}{m}) + \cdots + (x - \frac{1}{2} + \frac{m-1}{m}) = mx - \frac{1}{2}$  ; et il se trouve que la somme de toutes ces approximations grossières est exacte !

### 3.5 SOMMES DE PARTIES ENTIÈRES

L'équation (3.26) montre qu'il existe au moins un type de sommes contenant des  $\lfloor \cdot \rfloor$  pour lesquelles on peut trouver une formule close. Est-ce possible pour d'autres sommes ? La réponse est oui. Il y a un truc qui marche généralement dans ce genre de cas : se débarrasser des parties entières en introduisant une nouvelle variable.

Par exemple, voyons s'il est possible de trouver une formule close pour la somme

$$\sum_{0 \leq k < n} \lfloor \sqrt{k} \rfloor.$$

Choisissons d'introduire la variable  $m = \lfloor \sqrt{k} \rfloor$ . Cela peut être fait “mécaniquement”, en procédant comme nous l'avons fait dans le problème de la roulette :

$$\begin{aligned}\sum_{0 \leq k < n} \lfloor \sqrt{k} \rfloor &= \sum_{k, m \geq 0} m[k < n] [m = \lfloor \sqrt{k} \rfloor] \\&= \sum_{k, m \geq 0} m[k < n] [m \leq \sqrt{k} < m + 1] \\&= \sum_{k, m \geq 0} m[k < n] [m^2 \leq k < (m + 1)^2] \\&= \sum_{k, m \geq 0} m[m^2 \leq k < (m + 1)^2 \leq n] \\&\quad + \sum_{k, m \geq 0} m[m^2 \leq k < n < (m + 1)^2].\end{aligned}$$

Encore une fois, les conditions aux bornes sont un peu délicates à gérer. Supposons pour commencer que  $n = a^2$  est un carré parfait. Alors la seconde somme est nulle et on peut calculer la première en utilisant notre méthode habituelle :

$$\begin{aligned}\sum_{k, m \geq 0} m[m^2 \leq k < (m + 1)^2 \leq a^2] &= \sum_{m \geq 0} m((m + 1)^2 - m^2)[m + 1 \leq a] \\&= \sum_{m \geq 0} m(2m + 1)[m < a] \\&= \sum_{m \geq 0} (2m^2 + 3m^1)[m < a] \\&= \sum_0^a (2m^2 + 3m^1) \delta m \\&= \frac{2}{3}a(a - 1)(a - 2) + \frac{3}{2}a(a - 1) = \frac{1}{6}(4a + 1)a(a - 1).\end{aligned}$$

*Les puissances descendantes font tomber la somme.*

Dans le cas général, on peut poser  $a = \lfloor \sqrt{n} \rfloor$ . Il suffit alors d'ajouter les termes pour  $a^2 \leq k < n$ , qui sont tous égaux à  $a$  ; leur somme vaut donc  $(n - a^2)a$ . On obtient ainsi la formule close désirée :

$$\sum_{0 \leq k < n} \lfloor \sqrt{k} \rfloor = na - \frac{1}{3}a^3 - \frac{1}{2}a^2 - \frac{1}{6}a, \quad a = \lfloor \sqrt{n} \rfloor. \quad (3.27)$$

Une autre approche possible consiste à remplacer une expression de la forme  $\lfloor x \rfloor$  par  $\sum_j [1 \leq j \leq x]$  ; c'est autorisé si  $x \geq 0$ . Voici comment cette

## 94 FONCTIONS ENTIÈRES

méthode fonctionne pour la somme des [racines carrées], en supposant pour simplifier que  $n = a^2$  :

$$\begin{aligned} \sum_{0 \leq k < n} [\sqrt{k}] &= \sum_{j,k} [1 \leq j \leq \sqrt{k}] [0 \leq k < a^2] \\ &= \sum_{1 \leq j < a} \sum_k [j^2 \leq k < a^2] \\ &= \sum_{1 \leq j < a} (a^2 - j^2) = a^3 - \frac{1}{3}a(a + \frac{1}{2})(a + 1). \end{aligned}$$

Voici un autre exemple dans lequel un changement de variable permet de transformer une somme. Un théorème remarquable a été découvert vers 1909 indépendamment par trois mathématiciens : Bohl [34], Sierpiński [326] et Weyl [368] : si  $\alpha$  est irrationnel, alors, lorsque  $n \rightarrow \infty$ , les parties fractionnaires  $\{n\alpha\}$  sont distribuées de façon très uniforme entre 0 et 1. Plus formellement, on peut dire que pour tout  $\alpha$  irrationnel et pour toute fonction  $f$  presque partout continue,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{0 \leq k < n} f(\{k\alpha\}) = \int_0^1 f(x) dx \quad (3.28)$$

Par exemple, on peut trouver la valeur *moyenne* de  $\{n\alpha\}$  en posant  $f(x) = x$  ; on obtient  $\frac{1}{2}$  (c'est exactement ce à quoi on pouvait s'attendre, mais il est bon de savoir que c'est véritablement prouvé).

Le théorème de Bohl, Sierpiński et Weyl se démontre en encadrant  $f(x)$  par des "fonctions en escalier" qui sont des combinaisons linéaires des fonctions simples

$$f_v(x) = [0 \leq x < v]$$

avec  $0 \leq v \leq 1$ . Notre but n'est pas de prouver ce théorème ; c'est un boulot qu'on trouve dans les livres de calcul infinitésimal. Nous allons cependant essayer de découvrir intuitivement pourquoi il marche. Pour cela, nous allons voir s'il se comporte bien lorsque  $f(x) = f_v(x)$ . En d'autres termes, observons à quelle distance la somme

$$\sum_{0 \leq k < n} [\{k\alpha\} < v]$$

se rapproche de la valeur "idéale"  $nv$  quand  $n$  est grand et  $\alpha$  irrationnel.

Pour cela, on définit la *discrépance*  $D(\alpha, n)$  comme le maximum, pour tout  $0 \leq v \leq 1$ , de la valeur absolue de la somme

$$s(\alpha, n, v) = \sum_{0 \leq k < n} ([\{k\alpha\} < v] - v). \quad (3.29)$$

*Avertissement : ce qui suit est de haut niveau. En première lecture, il vaut mieux sauter les deux pages qui suivent ; elles ne sont pas d'une importance capitale.*

— Votre sympathique chargé de TD

Sautez  
à partir d'ici

Notre but est de prouver que  $D(\alpha, n)$  “n'est pas trop grand” comparé à  $n$ , en montrant que  $|s(\alpha, n, v)|$  est toujours raisonnablement petit lorsque  $\alpha$  est irrationnel.

Ecrivons tout d'abord  $s(\alpha, n, v)$  plus simplement et introduisons une nouvelle variable d'indice  $j$  :

$$\begin{aligned} \sum_{0 \leq k < n} ([k\alpha] < v) - v &= \sum_{0 \leq k < n} ([k\alpha] - [k\alpha - v] - v) \\ &= -nv + \sum_{0 \leq k < n} \sum_j [k\alpha - v < j \leq k\alpha] \\ &= -nv + \sum_{0 \leq j < \lceil n\alpha \rceil} \sum_{k < n} [j\alpha^{-1} \leq k < (j+v)\alpha^{-1}]. \end{aligned}$$

Avec un peu de chance, on va pouvoir calculer la somme interne ; mais il nous faut introduire de nouvelles variables si on ne veut pas faire trop de gâchis. On peut supposer sans perte de généralité que  $0 < \alpha < 1$ . Posons alors

$$\begin{aligned} a &= \lfloor \alpha^{-1} \rfloor, & \alpha^{-1} &= a + \alpha'; \\ b &= \lceil v\alpha^{-1} \rceil, & v\alpha^{-1} &= b - v'. \end{aligned}$$

*Bien vu : nommer pour régner. Le point crucial est le changement de  $k$  en  $j$ . — Votre sympathique chargé de TD*

Ainsi,  $\alpha' = \{\alpha^{-1}\}$  est la partie fractionnaire de  $\alpha^{-1}$ , et  $v'$  est la partie fractionnaire de marmot de  $v\alpha^{-1}$ .

Encore une fois, nos seuls motifs d'inquiétude sont les conditions aux bornes. Oublions pour le moment la restriction “ $k < n$ ” et évaluons la somme sur  $k$  :

$$\begin{aligned} \sum_k [k \in [j\alpha^{-1} \dots (j+v)\alpha^{-1}]] &= [(j+v)(a+\alpha')] - \lceil j(a+\alpha') \rceil \\ &= b + \lceil j\alpha' - v' \rceil - \lceil j\alpha' \rceil. \end{aligned}$$

Parfait. C'est très facile en fait. Cela nous donne

$$s(\alpha, n, v) = -nv + \lceil n\alpha \rceil b + \sum_{0 \leq j < \lceil n\alpha \rceil} (\lceil j\alpha' - v' \rceil - \lceil j\alpha' \rceil) - S, \quad (3.30)$$

où  $S$  est un terme correcteur pour les cas où  $k \geq n$ , cas dont nous n'avons pas pu nous débarrasser. La quantité  $j\alpha'$  ne pourra jamais être entière (sauf si  $j = 0$ ), car  $\alpha$  (et donc  $\alpha'$ ) est irrationnel ; et  $j\alpha' - v'$  ne pourra être un entier que pour au plus une valeur de  $j$ . On peut donc transformer les parties entières supérieures en parties entières inférieures :

$$s(\alpha, n, v) = -nv + \lceil n\alpha \rceil b - \sum_{0 \leq j < \lceil n\alpha \rceil} (\lfloor j\alpha' \rfloor - \lfloor j\alpha' - v' \rfloor) - S + \{0 \text{ ou } 1\}.$$

## 96 FONCTIONS ENTIÈRES

Voilà qui est intéressant. Au lieu d'une forme close, nous obtenons une somme qui ressemble tout à fait à  $s(\alpha, n, v)$ , mais avec des paramètres différents :  $\alpha'$  au lieu de  $\alpha$ ,  $\lceil n\alpha \rceil$  au lieu de  $n$  et  $v'$  au lieu de  $v$ . Nous allons ainsi obtenir une récurrence pour  $s(\alpha, n, v)$  qui (espérons le) va nous nous conduire à une récurrence pour la discrépance  $D(\alpha, n)$ . Prenons donc

$$s(\alpha', \lceil n\alpha \rceil, v') = \sum_{0 \leq j < \lceil n\alpha \rceil} (\lfloor j\alpha' \rfloor - \lfloor j\alpha' - v' \rfloor - v')$$

(La formule  $\{0 \text{ ou } 1\}$  dénote quelque chose qui est soit 0, soit 1 ; cela ne nous engage en rien, car en fait ces détails n'ont pas vraiment d'importance).

et portons le dans

$$s(\alpha, n, v) = -nv + \lceil n\alpha \rceil b - \lceil n\alpha \rceil v' - s(\alpha', \lceil n\alpha \rceil, v') - S + \{0 \text{ ou } 1\}.$$

Si on se souvient que  $b - v' = v\alpha^{-1}$ , on voit que tout cela va magnifiquement se simplifier en remplaçant  $\lceil n\alpha \rceil(b - v')$  par  $n\alpha(b - v') = nv$  :

$$s(\alpha, n, v) = -s(\alpha', \lceil n\alpha \rceil, v') - S + \epsilon + \{0 \text{ or } 1\}.$$

Le terme  $\epsilon$  désigne une erreur positive inférieure ou égale à  $v\alpha^{-1}$ . L'exercice 18 montre que, de manière similaire,  $S$  est compris entre 0 et  $\lceil v\alpha^{-1} \rceil$ . De plus, on peut sortir de la somme le terme correspondant à  $j = \lceil n\alpha \rceil - 1 = \lfloor n\alpha \rfloor$  car il est égal soit à  $v'$ , soit à  $v' - 1$ . Par conséquent, en prenant le maximum des valeurs absolues pour tout  $v$ , on trouve

$$D(\alpha, n) < D(\alpha', \lfloor n\alpha \rfloor) + \alpha^{-1} + 2. \quad (3.31)$$

Les méthodes que nous apprendrons dans les chapitres suivants nous permettront de déduire de cette récurrence que  $D(\alpha, n)$  est toujours beaucoup plus petit que  $n$  lorsque  $n$  est suffisamment grand. Ainsi le théorème (3.28) est vrai ; cependant la convergence n'est pas toujours très rapide (voir les exercices 9.45 et 9.61).

Ouf ! C'était un sacré exercice de manipulation de sommes et de parties entières ! Les lecteurs qui n'ont pas l'habitude de "prouver que l'erreur est négligeable" ont peut-être du mal à croire qu'il vaille le coup de s'attaquer à des sommes aussi alambiquées. En fait, si on y regarde de plus près, on voit qu'il y a un fil conducteur dans tout les calcul que nous avons effectués. L'idée principale est la suivante : une certaine somme  $s(\alpha, n, v)$  de  $n$  termes peut se réduire à une somme similaire de  $\lfloor n\alpha \rfloor$  termes au plus. Tout le reste disparaît, sauf un petit résidu qui provient des termes proches des bornes. A présent, respirons un grand coup et voyons une nouvelle somme. Elle n'est pas triviale mais possède un avantage non négligeable (par rapport à la précédente) : sa forme close sera facile à vérifier. Notre but est de trouver une expression pour

$$\sum_{0 \leq k < m} \left\lfloor \frac{nk + x}{m} \right\rfloor, \quad m > 0 \text{ entier}, \quad n \text{ entier}.$$

↓ Atterrir  
ici

*Autant vous prévenir tout de suite : voici la première application d'un précepte selon lequel un chapitre doit finir par la résolution d'un problème long et difficile, demandant une motivation un peu plus importante que la simple curiosité.*

—Des étudiants

*Touché. Mais alors, les gars, il faut obligatoirement vous parler d'applications pratiques pour que vous vous intéressiez à quelque chose ? Cette somme apparaît, par exemple, lorsque l'on étudie ou teste des générateurs de nombres aléatoires. Cependant, les mathématiciens s'y sont intéressés bien avant l'apparition des ordinateurs car ils trouvaient naturel de se demander s'il existait un moyen de sommer des progressions arithmétiques "arrondies".*

—Votre enseignant

Cela nous permettra de généraliser la somme de (3.26). Trouver une forme close de cette somme est un exercice plus coriace que tout ce que nous avons fait jusqu'à présent (sauf peut-être le problème de la discrépance que nous venons de voir). Cela ne nous empêchera pas d'en venir à bout, et ce sera très instructif.

Comme nous le faisons d'habitude, particulièrement pour les problèmes coriaces, commençons par regarder quelques petits exemples. Si  $n = 1$ , on a l'égalité (3.26), dans laquelle  $x$  est remplacé par  $x/m$  :

$$\left\lfloor \frac{x}{m} \right\rfloor + \left\lfloor \frac{1+x}{m} \right\rfloor + \cdots + \left\lfloor \frac{m-1+x}{m} \right\rfloor = \lfloor x \rfloor.$$

Tout comme dans le chapitre 1, il peut-être utile de regarder le cas  $n = 0$  pour avoir un peu plus d'information :

$$\left\lfloor \frac{x}{m} \right\rfloor + \left\lfloor \frac{x}{m} \right\rfloor + \cdots + \left\lfloor \frac{x}{m} \right\rfloor = m \left\lfloor \frac{x}{m} \right\rfloor.$$

Deux paramètres,  $m$  et  $n$ , interviennent dans notre problème. Observons quelques petits exemples pour des valeurs de  $m$ . Quand  $m = 1$ , il n'y a qu'un terme dans la somme et elle vaut  $\lfloor x \rfloor$ . Quand  $m = 2$ , sa valeur est  $\lfloor x/2 \rfloor + \lfloor (x+n)/2 \rfloor$ . Pour éviter l'interaction entre  $x$  et  $n$ , on peut chasser  $n$  de la partie entière inférieure, mais pour ce faire il nous faut considérer séparément les cas où  $n$  est pair ou impair. Si  $n$  est pair, alors  $n/2$  est entier et il peut sortir de la partie entière :

$$\left\lfloor \frac{x}{2} \right\rfloor + \left( \left\lfloor \frac{x}{2} \right\rfloor + \frac{n}{2} \right) = 2 \left\lfloor \frac{x}{2} \right\rfloor + \frac{n}{2}.$$

Si  $n$  est impair, alors  $(n-1)/2$  est entier et on a donc

$$\left\lfloor \frac{x}{2} \right\rfloor + \left( \left\lfloor \frac{x+1}{2} \right\rfloor + \frac{n-1}{2} \right) = \lfloor x \rfloor + \frac{n-1}{2}.$$

La dernière étape se déduit de (3.26) en prenant  $m = 2$ .

Les formules pour  $n$  pair et  $n$  impair ressemblent un peu à celles correspondant à  $n = 0$  et  $n = 1$ , mais aucun motif n'apparaît clairement encore. Examinons donc quelques petits exemples supplémentaires. Pour  $m = 3$ , la somme est

$$\left\lfloor \frac{x}{3} \right\rfloor + \left\lfloor \frac{x+n}{3} \right\rfloor + \left\lfloor \frac{x+2n}{3} \right\rfloor,$$

et il nous faut considérer trois cas : soit  $n$  est un multiple de 3, soit il est plus grand d'une unité, soit il est plus grand de deux unités qu'un multiple de 3. En d'autres termes,  $n \bmod 3 = 0, 1$ , ou 2. Si  $n \bmod 3 = 0$ , alors  $n/3$

et  $2n/3$  sont des entiers, donc la somme devient

$$\left\lfloor \frac{x}{3} \right\rfloor + \left( \left\lfloor \frac{x}{3} \right\rfloor + \frac{n}{3} \right) + \left( \left\lfloor \frac{x}{3} \right\rfloor + \frac{2n}{3} \right) = 3 \left\lfloor \frac{x}{3} \right\rfloor + n.$$

Si  $n \bmod 3 = 1$ , alors  $(n - 1)/3$  et  $(2n - 2)/3$  sont entiers et on a

$$\left\lfloor \frac{x}{3} \right\rfloor + \left( \left\lfloor \frac{x+1}{3} \right\rfloor + \frac{n-1}{3} \right) + \left( \left\lfloor \frac{x+2}{3} \right\rfloor + \frac{2n-2}{3} \right) = \lfloor x \rfloor + n - 1.$$

De nouveau, on déduit la dernière étape de (3.26), en posant cette fois  $m = 3$ . Enfin, si  $n \bmod 3 = 2$ , alors

$$\left\lfloor \frac{x}{3} \right\rfloor + \left( \left\lfloor \frac{x+2}{3} \right\rfloor + \frac{n-2}{3} \right) + \left( \left\lfloor \frac{x+1}{3} \right\rfloor + \frac{2n-1}{3} \right) = \lfloor x \rfloor + n - 1.$$

Notre cerveau gauche a terminé son travail pour  $m = 3$ ; mais le cerveau droit ne voit toujours pas de motif. Continuons donc avec  $m = 4$ :

$$\left\lfloor \frac{x}{4} \right\rfloor + \left\lfloor \frac{x+n}{4} \right\rfloor + \left\lfloor \frac{x+2n}{4} \right\rfloor + \left\lfloor \frac{x+3n}{4} \right\rfloor.$$

Au moins, nous savons maintenant qu'il faut considérer les cas qui dépendent de  $n \bmod m$ . Si  $n \bmod 4 = 0$ , alors

$$\left\lfloor \frac{x}{4} \right\rfloor + \left( \left\lfloor \frac{x}{4} \right\rfloor + \frac{n}{4} \right) + \left( \left\lfloor \frac{x}{4} \right\rfloor + \frac{2n}{4} \right) + \left( \left\lfloor \frac{x}{4} \right\rfloor + \frac{3n}{4} \right) = 4 \left\lfloor \frac{x}{4} \right\rfloor + \frac{3n}{2}.$$

Si  $n \bmod 4 = 1$ ,

$$\begin{aligned} \left\lfloor \frac{x}{4} \right\rfloor + \left( \left\lfloor \frac{x+1}{4} \right\rfloor + \frac{n-1}{4} \right) + \left( \left\lfloor \frac{x+2}{4} \right\rfloor + \frac{2n-2}{4} \right) + \left( \left\lfloor \frac{x+3}{4} \right\rfloor + \frac{3n-3}{4} \right) \\ = \lfloor x \rfloor + \frac{3n}{2} - \frac{3}{2}. \end{aligned}$$

*Inventive genius requires pleasurable mental activity as a condition for its vigorous exercise. ‘Necessity is the mother of invention’ is a silly proverb. ‘Necessity is the mother of futile dodges’ is much nearer to the truth. The basis of the growth of modern invention is science, and science is almost wholly the outgrowth of pleasurable intellectual curiosity.”*

— A. N. Whitehead [371]

Il se trouve que le cas  $n \bmod 4 = 3$  donne exactement la même réponse. Pour finir, le cas  $n \bmod 4 = 2$ , qui est légèrement différent, nous donne un indice important pour le comportement de la somme en général :

$$\begin{aligned} \left\lfloor \frac{x}{4} \right\rfloor + \left( \left\lfloor \frac{x+2}{4} \right\rfloor + \frac{n-2}{4} \right) + \left( \left\lfloor \frac{x}{4} \right\rfloor + \frac{2n}{4} \right) + \left( \left\lfloor \frac{x+2}{4} \right\rfloor + \frac{3n-2}{4} \right) \\ = 2 \left( \left\lfloor \frac{x}{4} \right\rfloor + \left\lfloor \frac{x+2}{4} \right\rfloor \right) + \frac{3n}{2} - 1 = 2 \left\lfloor \frac{x}{2} \right\rfloor + \frac{3n}{2} - 1. \end{aligned}$$

La dernière égalité de ce calcul consiste à simplifier quelque chose de la forme  $\lfloor y/2 \rfloor + \lfloor (y+1)/2 \rfloor$ . C'est de nouveau un cas particulier de (3.26).

Voici les valeurs que prend notre somme lorsque  $m$  varie de 1 à 3 :

$m$	$n \bmod m = 0$	$n \bmod m = 1$	$n \bmod m = 2$	$n \bmod m = 3$
1	$\lfloor x \rfloor$			
2	$2 \left\lfloor \frac{x}{2} \right\rfloor + \frac{n}{2}$	$\lfloor x \rfloor + \frac{n}{2} - \frac{1}{2}$		
3	$3 \left\lfloor \frac{x}{3} \right\rfloor + n$	$\lfloor x \rfloor + n - 1$	$\lfloor x \rfloor + n - 1$	
4	$4 \left\lfloor \frac{x}{4} \right\rfloor + \frac{3n}{2}$	$\lfloor x \rfloor + \frac{3n}{2} - \frac{3}{2}$	$2 \left\lfloor \frac{x}{2} \right\rfloor + \frac{3n}{2} - 1$	$\lfloor x \rfloor + \frac{3n}{2} - \frac{3}{2}$

On dirait qu'on obtient quelque chose de la forme

$$a \left\lfloor \frac{x}{a} \right\rfloor + bn + c,$$

où  $a$ ,  $b$  et  $c$  dépendent d'une certaine manière de  $m$  et  $n$ . Même un myope peut voir que  $b$  est probablement égal à  $(m - 1)/2$ . Il est plus difficile de deviner une expression pour  $a$ ; pourtant, le cas  $n \bmod 4 = 2$  semble nous suggérer que  $a$  pourrait être  $\text{pgcd}(m, n)$ , le plus grand commun diviseur de  $m$  et  $n$ . Ce ne serait pas très étonnant, car  $\text{pgcd}(m, n)$  est la valeur par laquelle il faut diviser  $m$  et  $n$  pour réduire la fraction  $n/m$ , et cette fraction apparaît dans notre somme. (Nous étudierons de près le pgcd au chapitre 4). La valeur de  $c$  est encore mystérieuse, mais peut-être nos preuves pour  $a$  et  $b$  pourront-elles nous aider à la trouver.

Dans nos calculs pour des petites valeurs de  $m$ , nous avons réécrit chacun des termes de la somme comme ceci :

$$\left\lfloor \frac{x + kn}{m} \right\rfloor = \left\lfloor \frac{x + kn \bmod m}{m} \right\rfloor + \frac{kn}{m} - \frac{kn \bmod m}{m},$$

car  $(kn - kn \bmod m)/m$  est un entier et peut donc sortir de la partie entière. De la même façon, notre somme d'origine peut être développée comme dans le tableau suivant :

$$\begin{array}{cccc}
 \left\lfloor \frac{x}{m} \right\rfloor & + & \frac{0}{m} & - \quad \frac{0 \bmod m}{m} \\
 + \quad \left\lfloor \frac{x + n \bmod m}{m} \right\rfloor & + & \frac{n}{m} & - \quad \frac{n \bmod m}{m} \\
 + \quad \left\lfloor \frac{x + 2n \bmod m}{m} \right\rfloor & + & \frac{2n}{m} & - \quad \frac{2n \bmod m}{m} \\
 & \vdots & \vdots & \vdots \\
 + \quad \left\lfloor \frac{x + (m-1)n \bmod m}{m} \right\rfloor & + & \frac{(m-1)n}{m} & - \quad \frac{(m-1)n \bmod m}{m}.
 \end{array}$$

Les trois colonnes de ce tableau correspondent respectivement aux valeurs  $a\lfloor x/a \rfloor$ ,  $bn$  et  $c$  de nos précédents calculs.

On peut voir en particulier comment  $b$  apparaît. La seconde colonne représente une progression arithmétique dont la somme est égale, nous le savons, à la moyenne du premier et du dernier terme multipliée par le nombre de termes :

$$\frac{1}{2} \left( 0 + \frac{(m-1)n}{m} \right) \cdot m = \frac{(m-1)n}{2}.$$

Nous venons ainsi de vérifier que  $b = (m+1)/2$ , comme nous l'avions présumé.

Les première et troisième colonnes paraissent plus coriaces. Pour déterminer les valeurs de  $a$  et  $c$ , regardons d'abord de plus près la suite de nombres

$$0 \bmod m, \quad n \bmod m, \quad 2n \bmod m, \quad \dots, \quad (m-1)n \bmod m.$$

Supposons par exemple que  $m = 12$  et  $n = 5$ . Si on considère la suite comme les heures d'une horloge, nous avons 0 heure (= 12 heures), puis 5 heures, 10 heures, 3 heures (= 15 heures), 8 heures etc. Chaque heure se trouve exactement une fois dans la suite.

Prenons maintenant  $m = 12$  et  $n = 8$ . La suite contient 0 heure, 8 heures, 4 heures (= 16 heures), puis on retrouve 0, 8, 4 et ainsi de suite. Comme 8 et 12 sont des multiples de 4 et le début de la suite est 0 (qui est aussi un multiple de 4), on ne sortira jamais de ce cycle.

Remarquons que  $\text{pgcd}(12, 5) = 1$  et  $\text{pgcd}(12, 8) = 4$ . Nous démontrerons dans le prochain chapitre que, si  $d = \text{pgcd}(m, n)$ , alors on obtient les nombres  $0, d, 2d, \dots, m-d$  dans un certain ordre, suivis par  $d-1$  copies de la même suite. Par exemple, si  $m = 12$  et  $n = 8$ , la suite  $0, 8, 4$  apparaît quatre fois.

Maintenant, la première colonne de notre somme prend tout son sens. Elle contient  $d$  copies des termes  $\lfloor x/m \rfloor, \lfloor (x+d)/m \rfloor, \dots, \lfloor (x+m-d)/m \rfloor$ , dans un certain ordre, donc sa somme est

$$\begin{aligned} & d \left( \left\lfloor \frac{x}{m} \right\rfloor + \left\lfloor \frac{x+d}{m} \right\rfloor + \dots + \left\lfloor \frac{x+m-d}{m} \right\rfloor \right) \\ &= d \left( \left\lfloor \frac{x/d}{m/d} \right\rfloor + \left\lfloor \frac{x/d + 1}{m/d} \right\rfloor + \dots + \left\lfloor \frac{x/d + m/d - 1}{m/d} \right\rfloor \right) \\ &= d \left\lfloor \frac{x}{d} \right\rfloor. \end{aligned}$$

*Un lemme maintenant, un dilemme plus tard.*

La dernière égalité est, une fois de plus, une application de (3.26). Nous

venons de montrer que, comme nous l'avions deviné,

$$a = d = \text{pgcd}(m, n).$$

De plus, comme nous le pensions, nous pouvons maintenant trouver  $c$ , car on voit bien mieux maintenant ce qui se passe dans la troisième colonne. Elle contient  $d$  copies de la progression arithmétique  $0/m, d/m, 2d/m, \dots, (m-d)/m$ ; sa somme vaut donc

$$d \left( \frac{1}{2} \left( 0 + \frac{m-d}{m} \right) \cdot \frac{m}{d} \right) = \frac{m-d}{2}.$$

En fait, la troisième colonne n'est pas additionnées mais soustraite, par conséquent

$$c = \frac{d-m}{2}.$$

Ainsi le mystère est dévoilé et notre quête touche à sa fin. La forme close que nous cherchions est

$$\sum_{0 \leq k < m} \left\lfloor \frac{nk+x}{m} \right\rfloor = d \left\lfloor \frac{x}{d} \right\rfloor + \frac{m-1}{2}n + \frac{d-m}{2},$$

où  $d = \text{pgcd}(m, n)$ . Juste pour vérifier, calculons la pour les cas déjà connus  $n=0$  et  $n=1$ . Quand  $n=0$ , on a  $d = \text{pgcd}(m, 0) = m$ ; les deux derniers termes de la formule sont nuls et on obtient bien  $m \lfloor x/m \rfloor$ . Lorsque  $n=1$ ,  $d = \text{pgcd}(m, 1) = 1$ ; les deux derniers termes se neutralisent mutuellement et la somme vaut  $\lfloor x \rfloor$  comme prévu.

En triturant encore un peu la forme close, on peut la rendre symétrique en  $m$  et  $n$ :

$$\begin{aligned} \sum_{0 \leq k < m} \left\lfloor \frac{nk+x}{m} \right\rfloor &= d \left\lfloor \frac{x}{d} \right\rfloor + \frac{m-1}{2}n + \frac{d-m}{2} \\ &= d \left\lfloor \frac{x}{d} \right\rfloor + \frac{(m-1)(n-1)}{2} + \frac{m-1}{2} + \frac{d-m}{2} \\ &= d \left\lfloor \frac{x}{d} \right\rfloor + \frac{(m-1)(n-1)}{2} + \frac{d-1}{2}. \end{aligned} \quad (3.32)$$

*Ma surprise est entière.*

C'est surprenant, car, algébriquement parlant, il n'y a pas de raison de se douter qu'une telle somme puisse être symétrique. Nous venons de démontrer la "loi de réciprocité" suivante :

$$\sum_{0 \leq k < m} \left\lfloor \frac{nk+x}{m} \right\rfloor = \sum_{0 \leq k < n} \left\lfloor \frac{mk+x}{n} \right\rfloor, \quad m, n > 0 \text{ entiers.}$$

Par exemple, si  $m = 41$  et  $n = 127$ , la somme de gauche contient 41 termes tandis que celle de droite en a 127 ; et pourtant, elles sont égales pour tout réel  $x$ .

## Exercices

### Echauffements

- 1 Au cours de l'analyse du problème de Josèphe au chapitre 1, nous avons décidé d'écrire un entier  $n$  quelconque sous la forme  $n = 2^m + l$ , où  $0 \leq l < 2^m$ . Calculez explicitement  $l$  et  $m$  en fonction de  $n$ , en utilisant les crochets de partie entière inférieure et/ou supérieure.
- 2 Trouvez une formule qui donne l'entier le plus proche d'un nombre réel donné  $x$ . Si  $x$  se trouve exactement au milieu de deux entiers, donnez une expression qui arrondit (a) par excès ; (b) par défaut.
- 3 Calculez  $\lfloor \lfloor m\alpha \rfloor n/\alpha \rfloor$ , où  $m$  et  $n$  sont des entiers strictement positifs et  $\alpha$  un nombre irrationnel supérieur à  $n$ .
- 4 On définit dans ce chapitre les problèmes de niveau 1 à 5. Qu'est-ce qu'un problème de niveau 0 ? (Celui-ci n'est *pas* un problème de niveau 0).
- 5 Donnez une condition nécessaire et suffisante pour que  $\lfloor nx \rfloor = n \lfloor x \rfloor$  lorsque  $n$  est un entier strictement positif. (On a le droit d'utiliser  $\{x\}$  dans la condition).
- 6 Peut-on dire quelque chose d'intéressant sur  $\lfloor f(x) \rfloor$  quand  $f(x)$  est une fonction continue et strictement *décroissante* dont la valeur ne peut être entière que si  $x$  est un entier ?
- 7 Résolvez la récurrence

$$\begin{aligned} X_n &= n, && \text{pour } 0 \leq n < m; \\ X_n &= X_{n-m} + 1, && \text{pour } n \geq m. \end{aligned}$$

- 8 Démontrez le *principe des boîtes de Dirichlet* : si on range  $n$  objets dans  $m$  boîtes, il y a au moins une boîte qui contient  $\geq \lceil n/m \rceil$  objets, et au moins une boîte qui en contient  $\leq \lfloor n/m \rfloor$ .
- 9 En 1800 avant Jésus-Christ, les mathématiciens égyptiens représentaient les nombres rationnels entre 0 et 1 comme des sommes de fractions unitaires  $1/x_1 + \dots + 1/x_k$ , où les  $x_i$  étaient des entiers strictement positifs distincts. Par exemple, ils écrivaient  $\frac{1}{3} + \frac{1}{15}$  au lieu de  $\frac{2}{5}$ . Montrez qu'on peut toujours le faire de façon systématique : si  $0 < m/n < 1$ , alors

$$\frac{m}{n} = \frac{1}{q} + \left\{ \text{représentation de } \frac{m}{n} - \frac{1}{q} \right\}, \quad q = \left\lceil \frac{n}{m} \right\rceil.$$

*Comment sait-on si on est à l'université ? C'est quand votre livre ne vous dit pas comment on doit prononcer "Dirichlet".*

(C'est l'*algorithme de Fibonacci*, dû à Leonardo Fibonacci (1202)).

### Exercices de base

- 10** Montrez que l'expression

$$\left\lceil \frac{2x+1}{2} \right\rceil - \left\lceil \frac{2x+1}{4} \right\rceil + \left\lceil \frac{2x+1}{4} \right\rceil$$

est toujours égale soit à  $\lfloor x \rfloor$ , soit à  $\lceil x \rceil$ . Dans quelles circonstances chacun de ces cas se produit-il ?

- 11** Ecrivez en détail la preuve, éludée dans le chapitre, du fait que l'intervalle ouvert  $(\alpha \dots \beta)$  contient exactement  $\lceil \beta \rceil - \lfloor \alpha \rfloor - 1$  entiers lorsque  $\alpha < \beta$ . Pourquoi la preuve ne peut-elle être correcte que si on exclut le cas  $\alpha = \beta$  ?

- 12** Montrez que

$$\left\lceil \frac{n}{m} \right\rceil = \left\lceil \frac{n+m-1}{m} \right\rceil$$

pour tout entier  $n$  et tout entier strictement positif  $m$ . (Cette identité nous fournit une nouvelle méthode, différente de la règle (3.4), pour convertir des parties entières supérieures en parties entières inférieures et vice-versa).

- 13** Soient  $\alpha$  et  $\beta$  deux réels strictement positifs. Montrez que  $\text{Spec}(\alpha)$  et  $\text{Spec}(\beta)$  forment une partition de l'ensemble des entiers strictement positifs si et seulement si  $\alpha$  et  $\beta$  sont irrationnels et  $1/\alpha + 1/\beta = 1$ .
- 14** Prouvez ou réfutez l'égalité suivante :

$$(x \bmod ny) \bmod y = x \bmod y, \quad n \text{ entier.}$$

- 15** Existe-t-il une identité analogue à (3.26) faisant intervenir des parties entières inférieures au lieu de parties entières supérieures ?
- 16** Montrez que  $n \bmod 2 = (1 - (-1)^n)/2$ . Trouvez et prouvez une expression similaire pour  $n \bmod 3$ , de la forme  $a + b\omega^n + c\omega^{2n}$ , où  $\omega$  est le nombre complexe  $(-1 + i\sqrt{3})/2$ . *Suggestion* : remarquez que  $\omega^3 = 1$  et  $1 + \omega + \omega^2 = 0$ .
- 17** Calculez la somme  $\sum_{0 \leq k < m} \lfloor x + k/m \rfloor$  dans le cas où  $x \geq 0$ , en substituant  $\sum_j [1 \leq j \leq x + k/m]$  à  $\lfloor x + k/m \rfloor$  et en sommant d'abord sur  $k$ . Votre réponse s'accorde-t-elle avec (3.26) ?
- 18** Montrez que le terme d'erreur  $S$  de (3.30) est au plus égal à  $\lceil \alpha^{-1}v \rceil$ . *Suggestion* : montrez que les petites valeurs de  $j$  n'y contribuent pas.

***Devoirs à la maison***

- 19 Donnez une condition nécessaire et suffisante que doit satisfaire le nombre  $b > 1$  pour que

$$\lfloor \log_b x \rfloor = \lfloor \log_b \lfloor x \rfloor \rfloor$$

pour tout réel  $x \geq 1$ .

- 20 Calculez la somme de tous les multiples de  $x$  situés dans l'intervalle fermé  $[\alpha \dots \beta]$ , pour  $x > 0$ .

- 21 Combien existe-t-il de nombres  $2^m$ , pour  $0 \leq m \leq M$ , dont le premier chiffre (le plus à gauche) en base 10 est 1 ?

- 22 Calculez  $S_n = \sum_{k \geq 1} \lfloor n/2^k + \frac{1}{2} \rfloor$  et  $T_n = \sum_{k \geq 1} 2^k \lfloor n/2^k + \frac{1}{2} \rfloor^2$ .

- 23 Montrez que le  $n$ ième élément de la suite

$$1, 2, 2, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, \dots$$

est égal à  $\lfloor \sqrt{2n} + \frac{1}{2} \rfloor$ . (La suite contient exactement  $m$  occurrences de chaque entier  $m$ ).

- 24 On établit dans l'exercice 13 une intéressante relation entre les deux multi-ensembles  $\text{Spec}(\alpha)$  et  $\text{Spec}(\alpha/(\alpha - 1))$  lorsque  $\alpha$  est un nombre irrationnel  $> 1$ , du fait que  $1/\alpha + (\alpha - 1)/\alpha = 1$ . Trouvez (et prouvez) une intéressante relation entre les deux multi-ensembles  $\text{Spec}(\alpha)$  et  $\text{Spec}(\alpha/(\alpha + 1))$  lorsque  $\alpha$  est un réel strictement positif quelconque.

- 25 Prouvez ou réfutez le fait que les nombres de Knuth, définis en (3.16), satisfont  $K_n \geq n$  pour tout entier positif ou nul  $n$ .

- 26 Montrez que les nombres auxiliaires de Josèphe (3.20) satisfont

$$\left( \frac{q}{q-1} \right)^n \leq D_n^{(q)} \leq q \left( \frac{q}{q-1} \right)^n, \quad \text{pour } n \geq 0.$$

- 27 Montrez que parmi les nombres  $D_n^{(3)}$  définis par (3.20), il y a une infinité de nombres pairs et une infinité de nombres impairs.

- 28 Résolvez la récurrence

$$\begin{aligned} a_0 &= 1; \\ a_n &= a_{n-1} + \lfloor \sqrt{a_{n-1}} \rfloor, \quad \text{pour } n > 0. \end{aligned}$$

- 29 Montrez qu'on peut ajouter à (3.31) l'inégalité suivante :

$$D(\alpha, n) \geq D(\alpha', \lfloor \alpha n \rfloor) - \alpha^{-1} - 2.$$

- 30** Montrez que si  $m$  est un entier strictement supérieur à 2, alors la récurrence

$$\begin{aligned} X_0 &= m, \\ X_n &= X_{n-1}^2 - 2, \quad \text{pour } n > 0, \end{aligned}$$

a pour solution  $X_n = \lceil \alpha^{2^n} \rceil$ , où  $\alpha + \alpha^{-1} = m$  et  $\alpha > 1$ . Par exemple, si  $m = 3$ , la solution est

$$X_n = \lceil \phi^{2^{n+1}} \rceil, \quad \phi = \frac{1 + \sqrt{5}}{2}, \quad \alpha = \phi^2.$$

- 31** Prouvez ou réfutez :  $|x| + |y| + |x+y| \leq \lfloor 2x \rfloor + \lfloor 2y \rfloor$ .

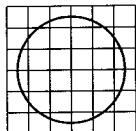
- 32** Soit  $\|x\| = \min(x - \lfloor x \rfloor, \lceil x \rceil - x)$  la distance entre  $x$  et l'entier le plus proche de  $x$ . Quelle est la valeur de

$$\sum_k 2^k \|x/2^k\|^2 ?$$

(Notez que cette somme peut être doublement infinie. Par exemple, quand  $x = 1/3$ , les termes sont non nuls lorsque  $k \rightarrow -\infty$ , ainsi que lorsque  $k \rightarrow +\infty$ ).

### Problèmes d'examen

- 33** Un cercle de diamètre  $2n-1$  est tracé au centre d'un échiquier  $2n \times 2n$ . En voici une illustration pour  $n = 3$  :



- a Combien de cases contiennent un arc du cercle ?  
 b Trouvez une fonction  $f(n, k)$  telle que le nombre de cellules entièrement contenues dans le cercle soit exactement égal à  $\sum_{k=1}^{n-1} f(n, k)$ .

- 34** Soit  $f(n) = \sum_{k=1}^n \lceil \lg k \rceil$ .  
 a Trouvez une formule close pour  $f(n)$  lorsque  $n \geq 1$ .  
 b Montrez que  $f(n) = n - 1 + f(\lceil n/2 \rceil) + f(\lfloor n/2 \rfloor)$  pour tout  $n \geq 1$ .

Simplifiez la, mais ne changez pas sa valeur.

- 35** Simplifiez la formule  $\lfloor (n+1)^2 n! e \rfloor \bmod n$ .

- 36** En supposant que  $n$  est un entier positif ou nul, trouvez une forme close de la somme

$$\sum_{1 \leq k < 2^{2^n}} \frac{1}{2^{\lceil \lg k \rceil} 4^{\lceil \lg \lg k \rceil}}.$$

## 37 Démontrez l'identité

$$\sum_{0 \leq k < m} \left( \left\lfloor \frac{m+k}{n} \right\rfloor - \left\lfloor \frac{k}{n} \right\rfloor \right) = \left\lfloor \frac{m^2}{n} \right\rfloor - \left\lfloor \frac{\min(m \bmod n, (-m) \bmod n)^2}{n} \right\rfloor$$

pour tous entiers strictement positifs  $m$  et  $n$ .

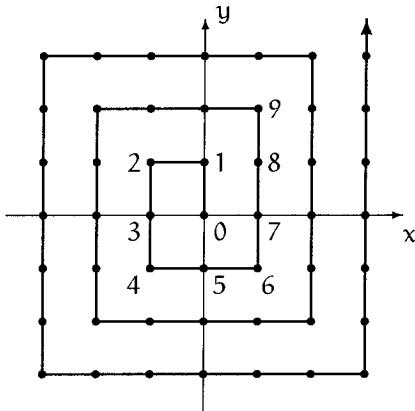
38 Soient  $x_1, \dots, x_n$  des nombres réels tels que l'identité

$$\sum_{k=1}^n \lfloor mx_k \rfloor = \left\lfloor m \sum_{1 \leq k \leq n} x_k \right\rfloor$$

soit vraie pour tout entier positif  $m$ . Démontrez quelque chose d'intéressant concernant  $x_1, \dots, x_n$ .

- 39 Montrez que la somme double  $\sum_{0 \leq k \leq \log_b x} \sum_{0 < j < b} \lceil (x + jb^k)/b^{k+1} \rceil$  est égale à  $(b-1)(\lfloor \log_b x \rfloor + 1) + \lceil x \rceil - 1$ , pour tout réel  $x \geq 1$  et tout entier  $b > 1$ .
- 40 La fonction en spirale  $\sigma(n)$ , décrite par la figure ci-dessous, associe à tout entier positif ou nul  $n$  un couple d'entiers  $(x(n), y(n))$ . Par exemple, l'image de 9 est le couple  $(1, 2)$ .

*Dans l'hémisphère Sud, la spirale est dans l'autre sens.*



- a Montrez que, si  $m = \lfloor \sqrt{n} \rfloor$ , alors

$$x(n) = (-1)^m \left( (n - m(m+1)) \cdot [2\sqrt{n} \text{ est pair}] + \lceil \frac{1}{2}m \rceil \right),$$

et donnez une formule similaire pour  $y(n)$ . *Suggestion :* partagez la spirale en segments  $W_k$ ,  $S_k$ ,  $E_k$  et  $N_k$ , selon que  $\lfloor 2\sqrt{n} \rfloor = 4k-2$ ,  $4k-1$ ,  $4k$ , ou  $4k+1$ .

- b Montrez qu'on peut, inversement, déterminer  $n$  en fonction de  $\sigma(n)$  par une formule de la forme

$$n = (2k)^2 \pm (2k + x(n) + y(n)), \quad k = \max(|x(n)|, |y(n)|).$$

Donnez une règle permettant de déterminer quand le signe doit être + et quand il doit être -.

### Questions subsidiaires

- 41 Soient  $f$  et  $g$  deux fonctions strictement croissantes telles que les ensembles  $\{f(1), f(2), \dots\}$  et  $\{g(1), g(2), \dots\}$  forment une partition de l'ensemble des entiers strictement positifs. Supposons de plus que  $f$  et  $g$  sont liées par la relation  $g(n) = f(f(n)) + 1$  pour tout  $n > 0$ . Prouvez que  $f(n) = \lfloor n\phi \rfloor$  et  $g(n) = \lfloor n\phi^2 \rfloor$ , avec  $\phi = (1 + \sqrt{5})/2$ .
- 42 Existe-t-il des nombres réels  $\alpha$ ,  $\beta$  et  $\gamma$  tels que  $\text{Spec}(\alpha)$ ,  $\text{Spec}(\beta)$  et  $\text{Spec}(\gamma)$  forment une partition de l'ensemble des entiers strictement positifs ?
- 43 En dépliant la récurrence (3.16), trouvez une interprétation intéressante des nombres de Knuth.
- 44 Montrez qu'il existe des entiers  $a_n^{(q)}$  et  $d_n^{(q)}$  tels que

$$a_n^{(q)} = \frac{D_{n-1}^{(q)} + d_n^{(q)}}{q-1} = \frac{D_n^{(q)} + d_n^{(q)}}{q}, \quad \text{pour } n > 0,$$

où  $D_n^{(q)}$  est la solution de (3.20). Utilisez cette propriété pour trouver une autre forme de la solution du problème de Josèphe généralisé :

$$J_q(n) = 1 + d_k^{(q)} + q(n - a_k^{(q)}), \quad \text{pour } a_k^{(q)} \leq n < a_{k+1}^{(q)}.$$

- 45 Inspirez-vous de l'exercice 30 pour trouver une formule close pour la solution de

$$\begin{aligned} Y_0 &= m, \\ Y_n &= 2Y_{n-1}^2 - 1, \quad \text{pour } n > 0, \end{aligned}$$

où  $m$  est un entier positif.

- 46 Montrez que si  $n = \lfloor (\sqrt{2}^l + \sqrt{2}^{l-1})m \rfloor$ , où  $m$  et  $l$  sont des entiers positifs ou nuls, alors  $\lfloor \sqrt{2n(n+1)} \rfloor = \lfloor (\sqrt{2}^{l+1} + \sqrt{2}^l)m \rfloor$ . Utilisez cette remarquable propriété pour trouver une solution en forme close de la récurrence

$$\begin{aligned} L_0 &= a, & a > 0 \text{ entier}; \\ L_n &= \lfloor \sqrt{2L_{n-1}(L_{n-1} + 1)} \rfloor, \quad \text{pour } n > 0. \end{aligned}$$

*Suggestion :*  $\lfloor \sqrt{2n(n+1)} \rfloor = \lfloor \sqrt{2}(n + \frac{1}{2}) \rfloor$ .

**47** Une fonction  $f(x)$  est dite *réplicative* si elle satisfait

$$f(mx) = f(x) + f\left(x + \frac{1}{m}\right) + \cdots + f\left(x + \frac{m-1}{m}\right)$$

pour tout entier strictement positif  $m$ . Trouvez des conditions nécessaires et suffisantes que doit respecter le réel  $c$  pour que les fonctions suivantes soient réplicatives :

- a  $f(x) = x + c$ .
- b  $f(x) = [x + c \text{ est un entier}]$ .
- c  $f(x) = \max([x], c)$ .
- d  $f(x) = x + c[x] - \frac{1}{2}[x \text{ n'est pas un entier}]$ .

**48** Prouvez l'identité

$$x^3 = 3x[x[x]] + 3\{x\}\{x[x]\} + \{x\}^3 - 3[x]\{x[x]\} + [x]^3$$

et montrez comment obtenir des formules similaires pour  $x^n$  lorsque  $n > 3$ .

**49** Trouvez une condition nécessaire et suffisante sur les nombres réels  $0 \leq \alpha < 1$  et  $\beta \geq 0$  pour qu'il soit possible, étant donné le multi-ensemble infini

$$\{[n\alpha] + [n\beta] \mid n > 0\}.$$

de déterminer les valeurs de  $\alpha$  et  $\beta$ .

#### Sujets de recherche

**50** Trouver une condition nécessaire et suffisante sur les nombres réels  $\alpha$  et  $\beta$  pour que  $\alpha$  et  $\beta$  puissent être déterminés par le multi-ensemble infini suivant :

$$\{[n\alpha]\beta \mid n > 0\}.$$

**51** Soit  $x$  un nombre réel  $\geq \phi = \frac{1}{2}(1 + \sqrt{5})$ . Si  $x$  est entier, la solution de la récurrence

$$\begin{aligned} Z_0(x) &= x, \\ Z_n(x) &= Z_{n-1}(x)^2 - 1, \quad \text{pour } n > 0, \end{aligned}$$

peut s'écrire  $Z_n(x) = \lceil f(x)^{2^n} \rceil$ , où

$$f(x) = \lim_{n \rightarrow \infty} Z_n(x)^{1/2^n},$$

car, dans ce cas,  $Z_n(x) - 1 < f(x)^{2^n} < Z_n(x)$ . Quelles autres propriétés intéressantes cette fonction  $f(x)$  présente-t-elle ?

- 52 Etant donnés deux réels positifs ou nuls  $\alpha$  et  $\beta$ , soit le multi-ensemble

$$\text{Spec}(\alpha; \beta) = \{\lfloor \alpha + \beta \rfloor, \lfloor 2\alpha + \beta \rfloor, \lfloor 3\alpha + \beta \rfloor, \dots\},$$

*Je suspecte que ça va être difficile.*

généralisation de  $\text{Spec}(\alpha) = \text{Spec}(\alpha; 0)$ . Prouvez ou réfutez l'assertion suivante : si les  $m \geq 3$  multi-ensembles  $\text{Spec}(\alpha_1; \beta_1)$ ,  $\text{Spec}(\alpha_2; \beta_2)$ , ...,  $\text{Spec}(\alpha_m; \beta_m)$  partitionnent l'ensemble des entiers strictement positifs, et si les paramètres  $\alpha_1 < \alpha_2 < \dots < \alpha_m$  sont rationnels, alors

$$\alpha_k = \frac{2^m - 1}{2^{k-1}}, \quad \text{pour } 1 \leq k \leq m.$$

- 53 L'algorithme de Fibonacci (exercice 9) est un “algorithme glouton”, car, à chaque étape, il choisit le plus petit  $q$  possible. Il existe un autre algorithme, plus compliqué, qui permet d'écrire toute fraction  $m/n$  telle que  $n$  est impair comme une somme de fractions distinctes  $1/q_1 + \dots + 1/q_k$  ayant toutes un dénominateur *impair*. L'algorithme glouton permettant d'arriver à cette décomposition se termine-t-il toujours ?

# 4

## Théorie des Nombres

LES NOMBRES ENTIERS occupent une place centrale dans les mathématiques discrètes. C'est la raison pour laquelle nous allons explorer la *théorie des nombres*, une branche importante des mathématiques qui concerne les propriétés des entiers.

Nous avons déjà mis un pied dans la théorie des nombres au cours du chapitre précédent, lorsque nous avons introduit les opérations binaires "mod" et "pgcd". Maintenant, plongeons-nous hardiment dedans pour véritablement nous imprégner du sujet.

*Autrement dit,  
préparez-vous à  
couler.*

### 4.1 DIVISIBILITÉ

On dit que  $m$  divise  $n$  (ou que  $n$  est divisible par  $m$ ) si  $m > 0$  et la fraction  $n/m$  est un entier. Cette propriété est sous-jacente à toute la théorie des nombres ; il convient donc de lui donner une notation spécifique. Nous écrirons donc

$$m \mid n \iff m > 0 \text{ et il existe un entier } k \text{ tel que } n = mk. \quad (4.1)$$

En fait, dans la littérature mathématique, la notation " $m \mid n$ " est plus communément utilisée que " $m \backslash n$ ". Néanmoins, nous jugeons que le trait vertical est bien trop souvent utilisé — pour les valeurs absolues, les délimiteurs d'ensembles, les probabilités conditionnelles etc — et que le caractère \ ne l'est pas assez. De plus, la notation " $m \backslash n$ " rappelle que  $m$  est le dénominateur d'une fraction implicite. Nous allons donc hardiment laisser notre signe de divisibilité pencher vers la gauche. Si  $m$  ne divise pas  $n$ , on écrit " $m \nmid n$ ".

Il existe une autre relation, que l'on exprime par " $n$  est un multiple de  $m$ ", qui signifie presque la même chose ; la différence est que  $m$  n'a pas besoin d'être strictement positif. Dans ce cas, on veut simplement dire qu'il existe un entier  $k$  tel que  $n = mk$ . Ainsi, par exemple, il existe un (et un seul) multiple de 0 (0 lui-même), bien qu'aucun nombre ne soit divisible

“... aucun entier n'est, à proprement parler, divisible par -1.”

—Graham, Knuth et Patashnik [161]

En Grande-Bretagne, on le note “hcf” (highest common factor).

Et aux Etats-Unis, c'est “gcd” (greatest common divisor) (N.d.T.).

Le terme américain est “least common multiple” et se note “lcm” (N.d.T.).

A ne pas confondre avec le plus grand commun multiple.

par 0. Tout entier est un multiple de -1, mais aucun entier n'est, à proprement parler, divisible par -1. Ces définitions peuvent aussi s'appliquer aux nombres réels ; par exemple,  $2\pi$  est divisible par  $\pi$ . Cependant nous les utiliserons seulement dans le cas où  $m$  et  $n$  sont entiers. N'oublions pas que nous faisons de la théorie des nombres.

Le plus grand commun diviseur de deux entiers  $m$  et  $n$  est le plus grand entier qui les divise tous deux :

$$\text{pgcd}(m, n) = \max\{k \mid k \mid m \text{ et } k \mid n\}. \quad (4.2)$$

Par exemple,  $\text{pgcd}(12, 18) = 6$ . C'est une notion bien connue, car on apprend à l'école qu'il faut diviser  $n$  et  $m$  par ce nombre pour réduire la fraction  $n/m$  :  $12/18 = (12/6)/(18/6) = 2/3$ . Notons que si  $n > 0$ , alors  $\text{pgcd}(0, n) = n$  car tout entier strictement positif divise 0, et  $n$  est le plus grand diviseur de lui-même. La valeur de  $\text{pgcd}(0, 0)$  est indéfinie.

Voici une autre notion familière : le plus petit commun multiple,

$$\text{ppcm}(m, n) = \min\{k \mid k > 0, \quad m \mid k \text{ et } n \mid k\}; \quad (4.3)$$

il n'est pas défini si  $m \leq 0$  ou  $n \leq 0$ . Ceux qui apprennent l'arithmétique le connaissent sous un autre nom : c'est le plus petit dénominateur commun, que l'on doit calculer pour additionner deux fractions de dénominateurs respectifs  $m$  et  $n$ . Par exemple,  $\text{ppcm}(12, 18) = 36$ , et tous les collégiens savent que  $\frac{7}{12} + \frac{1}{18} = \frac{21}{36} + \frac{2}{36} = \frac{23}{36}$ . Le ppcm est en quelque sorte analogue au pgcd ; toutefois ce dernier présente des propriétés bien plus intéressantes, c'est pourquoi nous en parlerons plus longuement.

Une des propriétés les plus agréables du pgcd est le fait qu'il est facile à calculer. On utilise pour cela une méthode, connue depuis 2300 ans, que l'on appelle l'*algorithme d'Euclide*. Étant donnés deux nombres  $m$  et  $n$  tels que  $0 \leq m < n$ , l'algorithme calcule  $\text{pgcd}(m, n)$  au moyen de la récurrence

$$\begin{aligned} \text{pgcd}(0, n) &= n; \\ \text{pgcd}(m, n) &= \text{pgcd}(n \bmod m, m), \quad \text{pour } m > 0. \end{aligned} \quad (4.4)$$

Ainsi, par exemple,  $\text{pgcd}(12, 18) = \text{pgcd}(6, 12) = \text{pgcd}(0, 6) = 6$ . Cette récurrence est correcte car tout commun diviseur de  $m$  et  $n$  doit aussi être un commun diviseur de  $m$  et  $n \bmod m$ , ce dernier nombre étant égal à  $n - \lfloor n/m \rfloor m$ . Il n'existe apparemment pas de récurrence aussi simple pour le ppcm (voir l'exercice 2).

L'algorithme d'Euclide nous donne plus encore : on peut le généraliser pour calculer deux entiers  $m'$  et  $n'$  tels que

$$m'm + n'n = \text{pgcd}(m, n). \quad (4.5)$$

Voici comment. Si  $m = 0$ , on prend simplement  $m' = 0$  et  $n' = 1$ . Sinon, on définit  $r = n \bmod m$  et on applique récursivement la méthode en remplaçant respectivement  $m$  et  $n$  par  $r$  et  $m$ , pour calculer  $\bar{r}$  et  $\bar{m}$  tels que

*N'oubliez pas que  $m'$  et  $n'$  peuvent être négatifs.*

$$\bar{r}r + \bar{m}m = \text{pgcd}(r, m).$$

Comme  $r = n - \lfloor n/m \rfloor m$  et  $\text{pgcd}(r, m) = \text{pgcd}(m, n)$ , on déduit de cette équation que

$$\bar{r}(n - \lfloor n/m \rfloor m) + \bar{m}m = \text{pgcd}(m, n).$$

Le membre gauche peut être réécrit de façon à montrer la dépendance en  $m$  et  $n$ :

$$(\bar{m} - \lfloor n/m \rfloor \bar{r})m + \bar{r}n = \text{pgcd}(m, n);$$

ce qui permet de conclure que  $m' = \bar{m} - \lfloor n/m \rfloor \bar{r}$  et  $n' = \bar{r}$  sont les deux entiers recherchés en (4.5). Par exemple, dans notre cas préféré  $m = 12$  et  $n = 18$ , on obtient par cette méthode  $6 = 0 \cdot 0 + 1 \cdot 6 = 1 \cdot 6 + 0 \cdot 12 = (-1) \cdot 12 + 1 \cdot 18$ .

Voyons maintenant en quoi (4.5) est si intéressant. La raison principale en est que, dans un certain sens, les nombres  $m'$  et  $n'$  *fournissent la preuve* que l'algorithme d'Euclide donne la bonne réponse dans tous les cas. Supposons que notre ordinateur nous dise, après un certain nombre de calculs, que  $\text{pgcd}(m, n) = d$  et que  $m'm + n'n = d$ . Qu'est-ce qui nous dit que le résultat est vrai, que l'ordinateur n'est pas passé à côté du vrai  $\text{pgcd}$ ? En fait, c'est tout simplement impossible, car tout diviseur commun de  $m$  et  $n$  doit aussi diviser  $m'm + n'n$ , donc il doit diviser  $d$ , donc il doit être  $\leq d$ . De plus, on peut facilement vérifier que  $d$  divise  $m$  et  $n$ . (De tels algorithmes, qui fournissent eux-mêmes leur preuve de correction, sont dits *auto-certifiants*).

Nous ferons beaucoup appel à (4.5) dans la suite de ce chapitre. Le mini-théorème suivant en est une des conséquences les plus importantes :

$$k \mid m \text{ et } k \mid n \iff k \mid \text{pgcd}(m, n). \quad (4.6)$$

(Preuve : si  $k$  divise à la fois  $m$  et  $n$ , alors il divise  $m'm + n'n$ , donc il divise  $\text{pgcd}(m, n)$ . Réciproquement, si  $k$  divise  $\text{pgcd}(m, n)$ , alors il divise un diviseur de  $m$  et un diviseur de  $n$  donc il divise  $m$  et  $n$ ). Nous savions déjà que tout commun diviseur de  $m$  et  $n$  doit être *inférieur ou égal à* leur  $\text{pgcd}$ . Nous savons maintenant que tout commun diviseur de  $m$  et  $n$  est en fait *un diviseur de* leur  $\text{pgcd}$ .

On a parfois besoin de sommer sur tous les diviseurs d'un entier  $n$ .

Dans ce cas, la règle suivante s'avère souvent utile :

$$\sum_{m \mid n} a_m = \sum_{m \mid n} a_{n/m}, \quad n > 0 \text{ entier.} \quad (4.7)$$

Elle se déduit du fait que, si  $m$  parcourt l'ensemble des diviseurs de  $n$ ,  $n/m$  le fait aussi. Par exemple, si  $n = 12$ , cela donne  $a_1 + a_2 + a_3 + a_4 + a_6 + a_{12} = a_{12} + a_6 + a_4 + a_3 + a_2 + a_1$ .

Il existe aussi une identité un peu plus générale,

$$\sum_{m \mid n} a_m = \sum_k \sum_{m>0} a_m[n=mk], \quad (4.8)$$

conséquence immédiate de la définition (4.1). Si  $n$  est positif, le membre droit de (4.8) est  $\sum_{k \mid n} a_{n/k}$ ; donc (4.8) entraîne (4.7). En outre, l'équation (4.8) est valable aussi lorsque  $n$  est négatif. (Dans ce cas, les termes de droite non nuls sont ceux pour lesquels  $-k$  est un diviseur de  $n$ ).

D'autre part, une somme double sur des diviseurs peut se réécrire selon la règle

$$\sum_{m \mid n} \sum_{k \mid m} a_{k,m} = \sum_{k \mid n} \sum_{l \mid (n/k)} a_{k,kl}. \quad (4.9)$$

Par exemple, pour  $n = 12$ , on a

$$\begin{aligned} a_{1,1} &+ (a_{1,2} + a_{2,2}) + (a_{1,3} + a_{3,3}) \\ &\quad + (a_{1,4} + a_{2,4} + a_{4,4}) + (a_{1,6} + a_{2,6} + a_{3,6} + a_{6,6}) \\ &\quad + (a_{1,12} + a_{2,12} + a_{3,12} + a_{4,12} + a_{6,12} + a_{12,12}) \\ &= (a_{1,1} + a_{1,2} + a_{1,3} + a_{1,4} + a_{1,6} + a_{1,12}) \\ &\quad + (a_{2,2} + a_{2,4} + a_{2,6} + a_{2,12}) + (a_{3,3} + a_{3,6} + a_{3,12}) \\ &\quad + (a_{4,4} + a_{4,12}) + (a_{6,6} + a_{6,12}) + a_{12,12}. \end{aligned}$$

L'équation (4.9) peut se démontrer en utilisant la notation d'Iverson. Voici le membre gauche :

$$\sum_{j,l} \sum_{k,m>0} a_{k,m}[n=jm][m=kl] = \sum_j \sum_{k,l>0} a_{k,kl}[n=jkl];$$

et le membre droit :

$$\sum_{j,m} \sum_{k,l>0} a_{k,kl}[n=jk][n/k=ml] = \sum_m \sum_{k,l>0} a_{k,kl}[n=mlk].$$

Ils sont identiques à un renommage des indices près. Nous voyons à travers cet exemple que les techniques présentées au chapitre 2 sont bien pratiques pour la théorie des nombres.

## 4.2 NOMBRES PREMIERS

Un entier positif  $p$  est dit *premier* s'il a exactement deux diviseurs, 1 et  $p$  lui-même. *Dans toute la suite du le chapitre, la lettre p désignera toujours un nombre premier, même si ce n'est pas explicitement indiqué.* Par convention, le nombre 1 n'est pas premier. Voici le tout début de la suite des nombres premiers :

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, \dots$$

Certains nombres ont l'air premiers mais ne le sont pas, comme 91 ( $= 7 \cdot 13$ ) et 161 ( $= 7 \cdot 23$ ). Ces nombres, et tous ceux qui ont trois diviseurs ou plus, sont dits *composites*. Tout nombre strictement supérieur à 1 est soit premier, soit composite.

Les nombres premiers sont extrêmement importants, car ils constituent les fondations de tout l'ensemble des entiers positifs. Tout entier positif  $n$  peut s'écrire comme un produit de nombres premiers,

$$n = p_1 \dots p_m = \prod_{k=1}^m p_k, \quad p_1 \leq \dots \leq p_m. \quad (4.10)$$

Par exemple,  $12 = 2 \cdot 2 \cdot 3$ ;  $11011 = 7 \cdot 11 \cdot 11 \cdot 13$ ;  $11111 = 41 \cdot 271$ . (On note les produits par  $\prod$  exactement de la même façon qu'on note les sommes par  $\sum$ , comme on peut le voir dans l'exercice 2.25. Si  $m = 0$ , on considère que le produit est vide et qu'il vaut 1 par définition; c'est le cas lorsque  $n = 1$  dans (4.10)). Cette factorisation est toujours possible car, si  $n > 1$  n'est pas premier, alors il a un diviseur  $n_1$  tel que  $1 < n_1 < n$ ; on peut donc écrire  $n = n_1 \cdot n_2$ , et on sait (par induction) que  $n_1$  et  $n_2$  peuvent s'écrire comme des produits de nombres premiers.

De plus, la décomposition de (4.10) est *unique*: il y a une seule façon d'écrire  $n$  comme un produit de nombres premiers en ordre croissant. Cette propriété constitue le Théorème Fondamental de l'Arithmétique. Elle paraît tellement évidente que l'on aurait tendance à se demander s'il est vraiment nécessaire de la prouver. Comment pourrait-il y avoir deux ensembles de nombres premiers différents qui donnent le même produit? C'est impossible en effet, mais cela ne s'explique pas simplement "par définition des nombres premiers". Par exemple, si on considère l'ensemble des nombres réels de la forme  $m + n\sqrt{10}$  où  $m$  et  $n$  sont des entiers, le produit de deux quelconques de ces nombres appartient à ce même ensemble. Convenons qu'un de ces nombres est "premier" s'il ne peut pas se factoriser de façon non triviale. Le nombre 6 a deux représentations,  $2 \cdot 3 = (4 + \sqrt{10})(4 - \sqrt{10})$ ; et l'exercice 36 montre que  $2, 3, 4 + \sqrt{10}$  et  $4 - \sqrt{10}$  sont tous "premiers" dans ce système.

Il nous faut donc prouver rigoureusement que la décomposition (4.10) est unique. Si  $n = 1$ , il y a bien sûr une seule possibilité car le produit est vide. Prenons alors  $n > 1$ , et supposons que tous les nombres strictement

Même la lettre  $p$  de "explicitement"?

inférieurs à  $n$  ont une unique décomposition. Supposons maintenant qu'il existe deux décompositions de  $n$ ,

$$n = p_1 \dots p_m = q_1 \dots q_k, \quad p_1 \leq \dots \leq p_m \quad \text{et} \quad q_1 \leq \dots \leq q_k,$$

où les  $p_i$  et les  $q_i$  sont tous premiers. Nous allons montrer que  $p_1 = q_1$ . Si ce n'est pas le cas, on peut poser  $p_1 < q_1$ , de sorte que  $p_1$  est plus petit que chacun des  $q_i$ . Comme  $p_1$  et  $q_1$  sont premiers, leur pgcd est égal à 1 ; donc l'algorithme d'Euclide renvoie deux entiers  $a$  et  $b$  tels que  $ap_1 + bq_1 = 1$ . Donc

$$ap_1 q_2 \dots q_k + bq_1 q_2 \dots q_k = q_2 \dots q_k.$$

Puisque  $q_1 q_2 \dots q_k = n$ ,  $p_1$  divise les deux termes du membre gauche ; donc  $p_1$  divise le membre de droite,  $q_2 \dots q_k$ . Par conséquent,  $q_2 \dots q_k / p_1$  est un entier, et  $q_2 \dots q_k$  a une décomposition en facteurs premiers dans laquelle  $p_1$  apparaît. Cependant, comme  $q_2 \dots q_k < n$ , il a une seule décomposition (par induction). Il y a contradiction, donc finalement  $p_1$  doit être égal à  $q_1$ . Ainsi on peut diviser les deux décompositions de  $n$  par  $p_1$ , pour obtenir  $p_2 \dots p_m = q_2 \dots q_k < n$ . Par hypothèse d'induction, tous ces facteurs sont égaux. Notre preuve d'unicité est donc terminée.

*C'est la factorisation qui est unique, pas le théorème.*

Il est parfois plus pratique d'exprimer le Théorème Fondamental d'une autre manière : *Tout entier strictement positif peut s'écrire de façon unique sous la forme*

$$n = \prod_p p^{n_p}, \quad \text{où tout } n_p \geq 0. \quad (4.11)$$

Le membre droit est un produit sur un nombre infini de nombres premiers ; mais, pour tout  $n$  donné, presque tous les exposants sont nuls, de sorte que les facteurs correspondants sont égaux à 1. C'est donc en réalité un produit fini, tout comme certaines sommes "infinies" qui s'avèrent en réalité finies car la plupart de leurs termes sont nuls.

La formule (4.11) donne une représentation unique de  $n$ . On peut donc considérer la suite  $\langle n_2, n_3, n_5, \dots \rangle$  comme un *système de représentation* des entiers strictement positifs. Par exemple, la représentation en puissances premières de 12 est  $\langle 2, 1, 0, 0, \dots \rangle$  et celle de 18 est  $\langle 1, 2, 0, 0, \dots \rangle$ . Pour multiplier deux nombres, il suffit d'additionner leurs représentations. En d'autres termes,

$$k = mn \iff k_p = m_p + n_p \quad \text{pour tout } p. \quad (4.12)$$

On en déduit que

$$m \mid n \iff m_p \leq n_p \quad \text{pour tout } p, \quad (4.13)$$

et il s'ensuit immédiatement que

$$k = \text{pgcd}(m, n) \iff k_p = \min(m_p, n_p) \text{ pour tout } p; \quad (4.14)$$

$$k = \text{ppcm}(m, n) \iff k_p = \max(m_p, n_p) \text{ pour tout } p. \quad (4.15)$$

Par exemple, comme  $12 = 2^2 \cdot 3^1$  et  $18 = 2^1 \cdot 3^2$ , on trouve leur pgcd et leur ppcm en prenant les min et max de leurs exposants communs :

$$\text{pgcd}(12, 18) = 2^{\min(2, 1)} \cdot 3^{\min(1, 2)} = 2^1 \cdot 3^1 = 6;$$

$$\text{ppcm}(12, 18) = 2^{\max(2, 1)} \cdot 3^{\max(1, 2)} = 2^2 \cdot 3^2 = 36.$$

Si un nombre premier  $p$  divise un produit  $mn$ , alors, en vertu du théorème d'unique décomposition,  $p$  divise  $m$ , ou  $n$ , ou peut-être même les deux. En revanche, les nombres composites ne présentent pas cette propriété : par exemple, le nombre 4 divise  $60 = 6 \cdot 10$ , mais il ne divise ni 6 ni 10. La raison en est simple : dans la décomposition  $60 = 6 \cdot 10 = (2 \cdot 3)(2 \cdot 5)$ , les deux facteurs premiers de 4 = 2 · 2 sont séparés. Par contre, un nombre premier n'est pas décomposable, donc il divise forcément l'un des facteurs originaux.

### 4.3 PREMIERS EXEMPLES PREMIERS

Combien existe-t-il de nombres premiers ? Beaucoup. Une infinité, en fait. Euclide a prouvé cela il y a fort longtemps dans son théorème 9 : 20, en procédant de la façon suivante. Supposons qu'il n'existe qu'un nombre fini  $k$  de nombres premiers :  $2, 3, 5, \dots, P_k$ . Considérons alors, dit Euclide, le nombre  $M = 2 \cdot 3 \cdot 5 \dots \cdot P_k + 1$ . Aucun des  $k$  nombres premiers n'est un diviseur de  $M$ , car chacun d'eux divise  $M - 1$ . Donc il existe au moins un autre nombre premier qui divise  $M$ . Peut être  $M$  lui-même est-il premier. Ceci contredit l'hypothèse selon laquelle  $2, 3, \dots, P_k$  sont les seuls nombres premiers ; on en déduit donc qu'il en existe une infinité.

C'est l'occasion ou jamais de présenter les *nombres d'Euclide*, définis par la récurrence

$$e_n = e_1 e_2 \dots e_{n-1} + 1, \quad \text{pour } n \geq 1. \quad (4.16)$$

La suite commence par

$$e_1 = 1 + 1 = 2;$$

$$e_2 = 2 + 1 = 3;$$

$$e_3 = 2 \cdot 3 + 1 = 7;$$

$$e_4 = 2 \cdot 3 \cdot 7 + 1 = 43;$$

*"Οἱ πρῶτοι ἀριθμοὶ πλείους εἰσὶ παντὸς τοῦ προτεθέντος πλήθους πρώτων ἀριθμῶν."*

— Euclide [98]

[Traduction :  
"Il existe plus de nombres premiers qu'il n'en existe dans tout ensemble donné de nombres premiers."]

qui sont tous des nombres premiers . Cependant, le suivant,  $e_5$ , est  $1807 = 13 \cdot 139$ . Son successeur  $e_6 = 3263443$  est premier, tandis que

$$\begin{aligned} e_7 &= 547 \cdot 607 \cdot 1033 \cdot 31051; \\ e_8 &= 29881 \cdot 67003 \cdot 9119521 \cdot 6212157481. \end{aligned}$$

On sait que  $e_9, \dots, e_{17}$  sont composites, et les autres  $e_n$  le sont probablement aussi. Toutefois, les nombres d'Euclide sont tous *premiers entre eux*, ce qui signifie que

$$\text{pgcd}(e_m, e_n) = 1, \quad \text{lorsque } m \neq n.$$

L'algorithme d'Euclide (encore lui !) nous le démontre en trois étapes, car  $e_n \bmod e_m = 1$  lorsque  $n > m$  :

$$\text{pgcd}(e_m, e_n) = \text{pgcd}(1, e_m) = \text{pgcd}(0, 1) = 1.$$

Ceci dit, si on définit  $q_j$  comme le plus petit facteur de  $e_j$  pour tout  $j \geq 1$ , alors les nombres premiers  $q_1, q_2, q_3, \dots$  sont tous différents. On obtient ainsi une suite infinie de nombres premiers.

Arrêtons-nous un instant pour considérer les nombres d'Euclide sous le même angle qu'au chapitre 1. Existe-t-il une forme close pour  $e_n$  ? La récurrence (4.16) se simplifie si on supprime les points de suspension : si  $n > 1$ , alors

$$e_n = e_1 \dots e_{n-2} e_{n-1} + 1 = (e_{n-1} - 1)e_{n-1} + 1 = e_{n-1}^2 - e_{n-1} + 1.$$

Ainsi,  $e_n$  s'écrit avec environ deux fois plus de chiffres que  $e_{n-1}$ . On montre dans l'exercice 37 qu'il existe une constante  $E \approx 1.264$  telle que

$$e_n = \lfloor E^{2^n} + \frac{1}{2} \rfloor. \tag{4.17}$$

D'autre part, l'exercice 60 nous fournit une formule similaire, qui ne donne que des nombres premiers :

$$p_n = \lfloor P^{3^n} \rfloor, \tag{4.18}$$

pour une certaine constante  $P$ . Cependant, on ne peut pas vraiment considérer les équations telles que (4.17) ou (4.18) comme des formes closes, car les constantes  $E$  et  $P$  sont, de façon un peu sournoise, calculées à partir des nombres  $e_n$  et  $p_n$ . On ne connaît pas de relation "indépendante" qui permettrait de les lier à d'autres constantes mathématiques connues.

En fait, on ne connaît *aucune* formule pratique qui donne une suite infinie de nombres distincts tous premiers. Les informaticiens de Chevron Geosciences ont cependant trouvé un bon filon en 1984. Un programme,

développé par David Slowinski pour tester un nouveau super-ordinateur Cray X-MP, leur a permis de découvrir le plus grand nombre premier connu à l'époque,

$$2^{216091} - 1.$$

N'importe quel ordinateur personnel peut calculer ce nombre en quelques millisecondes. En effet, les ordinateurs modernes travaillent en mode binaire, et ce nombre s'écrit simplement  $(11\dots1)_2$ . Chacun de ses 216 091 bits est égal à “1”. Il est cependant bien plus difficile de prouver que ce nombre est premier. En fait, le moindre calcul sur ce nombre prend un temps énorme à cause de sa très grande taille. Par exemple, il faut, même si on utilise un algorithme sophistiqué, plusieurs minutes à un ordinateur personnel pour seulement convertir  $2^{216091} - 1$  en base 10. Si on imprime ses 65 050 décimales, il faut payer 78 cents de timbres pour l'envoyer par la poste américaine en courrier rapide.

Notons en passant que  $2^{216091} - 1$  est le nombre de déplacements nécessaires pour résoudre le problème de la Tour de Hanoi avec 216 091 disques. Les nombres de la forme

$$2^p - 1$$

(où  $p$  est premier, comme dans tout le chapitre), sont appelés les *nombres de Mersenne*, du nom du Père Marin Mersenne qui a étudié certaines de leurs propriétés au dix-neuvième siècle [269]. Les nombres de Mersenne que l'on sait premiers à ce jour (avant 1997) sont ceux pour lesquels  $p = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497, 86243, 110503, 132049, 216091, 756839, 859433, 1257787 ou 1398269.$

Si  $n$  est composite, le nombre  $2^n - 1$  ne peut pas être premier, car  $2^{km} - 1$  admet  $2^m - 1$  pour facteur :

$$2^{km} - 1 = (2^m - 1)(2^{m(k-1)} + 2^{m(k-2)} + \dots + 1).$$

Si  $p$  est premier,  $2^p - 1$  n'est pas toujours premier :  $2^{11} - 1 = 2047 = 23 \cdot 89$  est le plus petit de ces nombres non premiers (Mersenne savait cela).

La factorisation et le test de primalité de grands entiers sont des sujets extrêmement étudiés actuellement. On peut trouver dans la Section 4.5.4 de [208] un compte-rendu de l'état de l'art en 1981 ; beaucoup de résultats nouveaux ont été découverts depuis lors. Les pages 391 à 394 de [208] décrivent une méthode de test de primalité spécifique aux nombres de Mersenne.

Depuis presque deux cents ans, le plus grand nombre premier connu a toujours été un nombre de Mersenne, bien qu'on n'en connaisse que quelques

*Probablement plus encore au moment où vous lisez ces lignes.*

dizaines. Beaucoup de gens essaient d'en trouver d'autres, mais c'est une tâche qui devient de plus en plus difficile. Ceux que la gloire (sinon la fortune) intéresse et qui veulent être cités dans le *Livre Guiness des Records* devraient plutôt essayer des nombres de la forme  $2^n k + 1$ , pour des petites valeurs de  $k$  comme 3 ou 5. On peut tester la primalité de ces nombres presque aussi vite que pour les nombres de Mersenne. On trouvera les détails dans l'exercice 4.5.4–27 de [208].

Nous n'avons pas tout à fait répondu à notre question : combien y a-t-il de nombres premiers ? Nous savons maintenant qu'il y en a une infinité, mais certains ensembles infinis sont plus "denses" que d'autres. Par exemple, il y a parmi les entiers positifs une infinité de nombres pairs et une infinité de carrés parfaits ; pourtant, selon certains critères, les entiers pairs sont plus nombreux que les carrés parfaits. Un de ces critères est la valeur du  $n$ ème nombre de chacun des deux ensembles. Le  $n$ ème entier pair est  $2n$  et le  $n$ ème carré parfait est  $n^2$  ; comme  $2n$  est bien plus petit que  $n^2$  quand  $n$  est grand, le premier apparaît bien plus tôt que le second dans l'ensemble des entiers positifs ; on peut donc dire qu'il y a beaucoup plus d'entiers pairs que de carrés parfaits. Un critère similaire est le nombre d'éléments de chacun des deux ensembles qui sont inférieurs à un réel  $x$  donné. Il y a  $\lfloor x/2 \rfloor$  entiers pairs inférieurs à  $x$ , et  $\lfloor \sqrt{x} \rfloor$  carrés parfaits ; comme  $x/2$  est bien plus grand que  $\sqrt{x}$  lorsque  $x$  est grand, on déduit encore une fois que les entiers pairs sont les plus nombreux.

*Bizarre. Je croyais, du fait qu'il existe une bijection entre ces deux ensembles, qu'il y avait exactement autant de nombres pairs que de carrés parfaits.*

Que pouvons-nous dire des nombres premiers selon ces deux critères ? Nous savons que le  $n$ ème nombre premier,  $P_n$ , est à peu près égal à  $n$  fois le logarithme népérien de  $n$  :

$$P_n \sim n \ln n.$$

(Le symbole " $\sim$ " peut se lire "tend asymptotiquement vers" ; cela signifie que le quotient  $P_n/n \ln n$  tend vers 1 lorsque  $n$  tend vers l'infini). On connaît aussi le résultat suivant, appelé le théorème des nombres premiers : le nombre  $\pi(x)$  de nombres premiers inférieurs à  $x$  satisfait

$$\pi(x) \sim \frac{x}{\ln x}.$$

Les preuves de ces deux résultats sont bien au-delà du propos de ce livre ; il est cependant facile de montrer qu'ils sont équivalents. Dans le chapitre 9, nous nous intéresserons à la vitesse de croissance des fonctions vers l'infini ; nous verrons alors que la fonction  $n \ln n$ , notre approximation de  $P_n$ , se trouve asymptotiquement entre  $2n$  et  $n^2$ . Il y a donc moins de nombres premiers que d'entiers pairs, mais plus que de carrés parfaits.

Ces formules ne sont valables que lorsque  $n$  ou  $x \rightarrow \infty$ . On connaît des approximations plus précises. Par exemple, Rosser et Schoenfeld [312]

ont établi les bornes suivantes :

$$\ln x - \frac{3}{2} < \frac{x}{\pi(x)} < \ln x - \frac{1}{2} \quad \text{si } x \geq 67; \quad (4.19)$$

$$n(\ln n + \ln \ln n - \frac{3}{2}) < P_n < n(\ln n + \ln \ln n - \frac{1}{2}), \quad \text{si } n \geq 20. \quad (4.20)$$

Si on tire un entier  $n$  "au hasard", il a, en gros, une chance sur  $\ln n$  d'être premier. Par exemple, si on considère des nombres proches de  $10^{16}$ , il faudra en tirer à peu près  $16 \ln 10 \approx 36,8$  pour en trouver un qui soit premier. (Il se trouve justement qu'il y a exactement 10 nombres premiers entre  $10^{16} - 370$  et  $10^{16} - 1$ ). Il y a cependant beaucoup d'irrégularités dans la distribution des nombres premiers. Par exemple, tous les nombres compris entre  $P_1 P_2 \dots P_n + 2$  et  $P_1 P_2 \dots P_n + P_{n+1} - 1$  sont composites. On connaît d'autre part beaucoup d'exemples de "nombres premiers jumeaux"  $p$  et  $p + 2$  (5 et 7, 11 et 13, 17 et 19, 29 et 31, ..., 9999999999999641 et 9999999999999643, ...); cependant personne ne sait s'il existe ou non une infinité de telles paires (voir Hardy et Wright [181, §1.4 et §2.8]).

Il existe une façon simple de trouver tous les  $\pi(x)$  nombres premiers jusqu'à un entier donné  $x$ : c'est le crible d'Eratosthène. D'abord, on écrit tous les entiers de 2 à  $x$ . Puis on entoure 2 pour indiquer qu'il est premier et on barre tous ses multiples. Ensuite, on entoure le premier nombre non barré et non entouré et on barre tous ses multiples. On répète l'opération jusqu'à avoir soit entouré, soit barré chaque nombre de la liste. A ce moment là, les nombres entourés sont tous les nombres premiers entre 2 et  $x$ . Prenons par exemple  $n = 10$ . On commence par écrire les nombres de 2 à 10, puis on entoure 2 et on barre ses multiples 4, 6, 8 et 10. Le nombre 3 est le premier qui n'est ni entouré ni barré, donc on l'entoure et on barre 6 et 9. Puis on entoure 5 et on barre 10. Pour finir, on entoure 7. Les nombres entourés sont 2, 3, 5 et 7 ; ce sont donc les  $\pi(10) = 4$  nombres premiers inférieurs ou égaux à 10.

#### 4.4 FACTEURS FACTORIELS

Jetons maintenant un coup d'œil sur la factorisation de certains nombres hautement composites, qu'on appelle factorielles :

$$n! = 1 \cdot 2 \cdot \dots \cdot n = \prod_{k=1}^n k, \quad n \geq 0 \text{ entier.} \quad (4.21)$$

Selon notre convention concernant les produits vides,  $0!$  est égal à 1. Donc  $n! = (n-1)!n$  pour tout entier  $n$  strictement positif. La factorielle de  $n$  est le nombre de permutations de  $n$  objets distincts, c'est-à-dire le nombre de façons d'ordonner  $n$  objets sur une ligne. Il y a  $n$  choix possibles pour le premier objet ; pour chacun de ces  $n$  choix, il y a  $n-1$  choix possibles pour

*"Je me sers de la notation très simple  $n!$  pour désigner le produit de nombres décroissants depuis  $n$  jusqu'à l'unité, savoir  $n(n-1)(n-2)\dots 3.2.1$ . L'emploi continuel de l'analyse combinatoire que je fais dans la plupart de mes démonstrations, a rendu cette notation indispensable."*  
*— Ch. Kramp [228]*

le deuxième objet ; pour chacune de ces  $n(n - 1)$  possibilités, le troisième objet peut être choisi de  $(n - 2)$  façons différentes ; et ainsi de suite, ce qui fait  $n(n - 1)(n - 2) \dots (1)$  possibilités au total. Voici les premières valeurs de la fonction factorielle :

n	0	1	2	3	4	5	6	7	8	9	10
n!	1	1	2	6	24	120	720	5040	40320	362880	3628800

Il est bon de connaître les 5 ou 6 premières valeurs de cette table, ainsi que le fait que  $10!$  vaut 3,5 millions et des poussières ; retenez aussi que le nombre de chiffres de  $n!$  est plus grand que  $n$  dès que  $n \geq 25$ .

On peut montrer que  $n!$  est extrêmement grand avec une méthode proche de l'astuce de Gauss que nous avons vue au chapitre 1 :

$$n!^2 = (1 \cdot 2 \cdot \dots \cdot n)(n \cdot \dots \cdot 2 \cdot 1) = \prod_{k=1}^n k(n+1-k).$$

Nous savons que  $n \leq k(n+1-k) \leq \frac{1}{4}(n+1)^2$ , puisque le polynôme du second degré  $k(n+1-k) = \frac{1}{4}(n+1)^2 - (k - \frac{1}{2}(n+1))^2$  a son **minimum** en  $k = 1$  et son **maximum** en  $k = \frac{1}{2}(n+1)$ . Ainsi,

$$\prod_{k=1}^n n \leq n!^2 \leq \prod_{k=1}^n \frac{(n+1)^2}{4},$$

ce qui nous donne

$$n^{n/2} \leq n! \leq \frac{(n+1)^n}{2^n}. \quad (4.22)$$

Cette dernière relation nous indique que la fonction factorielle croît exponentiellement !

La formule de Stirling, que nous démontrerons au chapitre 9, donne une meilleure approximation de  $n!$  lorsque  $n$  est grand :

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \quad (4.23)$$

En utilisant une approximation encore plus précise, on obtient l'erreur relative de cette formule : la formule de Stirling sous-estime  $n!$  d'un facteur  $1/(12n)$  environ. Cette estimation est plutôt précise, même lorsque  $n$  est assez petit. Par exemple, (4.23) donne une valeur à peu près égale à 3598696 pour  $n = 10$ , ce qui est, en gros,  $0,83\% \approx 1/120$  trop petit. Pas si mal, le calcul asymptotique !

Revenons à nos nombres premiers. Nous aimerais connaître, pour tout nombre premier  $p$  donné, la plus grande puissance de  $p$  qui divise  $n!$ .

Autrement dit, nous cherchons la valeur de l'exposant de  $p$  dans l'unique décomposition de  $n!$  en facteurs premiers. Nous noterons cet exposant  $\epsilon_p(n!)$ . Regardons d'abord un petit exemple,  $p = 2$  et  $n = 10$ . Comme  $10!$  est un produit de dix nombres, on peut trouver  $\epsilon_2(10!)$  en additionnant les exposants de 2 qui contribuent à la décomposition de chacun de ces dix nombres ; cela revient à additionner les colonnes du tableau suivant :

	1	2	3	4	5	6	7	8	9	10	puissances de 2
divisible par 2	x	x	x	x	x						$5 = \lfloor 10/2 \rfloor$
divisible par 4		x		x							$2 = \lfloor 10/4 \rfloor$
divisible par 8				x							$1 = \lfloor 10/8 \rfloor$
puissances de 2	0	1	0	2	0	1	0	3	0	1	8

Les sommes des colonnes forment ce qu'on appelle parfois la *fonction de la règle  $\rho(k)$* , en raison de leur ressemblance avec les longueurs successives des graduations d'une règle comme "████████████████████████". La somme de ces huit colonnes est égale à 8 ; donc  $2^8$  est un diviseur de  $10!$ , mais  $2^9$  ne l'est pas.

C'est une règle à calcul ?

On peut aussi opérer différemment, en additionnant les lignes. La première ligne marque les nombres qui contiennent 2 dans leur décomposition (donc qui sont divisibles par 2) ; il y en a  $\lfloor 10/2 \rfloor = 5$ . La deuxième ligne marque ceux dont la décomposition contient un 2 de plus ; il y en a  $\lfloor 10/4 \rfloor = 2$ . Enfin, la troisième marque les nombres dont la décomposition contient encore un 2 de plus ; il y en a  $\lfloor 10/8 \rfloor = 1$ . Toutes les contributions de 2 ont été comptées, donc  $\epsilon_2(10!) = 5 + 2 + 1 = 8$ .

Pour  $n$  quelconque, on obtient par cette méthode

$$\epsilon_2(n!) = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{n}{4} \right\rfloor + \left\lfloor \frac{n}{8} \right\rfloor + \cdots = \sum_{k \geq 1} \left\lfloor \frac{n}{2^k} \right\rfloor.$$

En réalité, la somme est finie car le terme général est nul quand  $2^k > n$ . Elle contient donc seulement  $\lfloor \lg n \rfloor$  termes non nuls et est très facile à calculer. Par exemple, pour  $n = 100$  on a

$$\epsilon_2(100!) = 50 + 25 + 12 + 6 + 3 + 1 = 97.$$

Chaque terme est égal à la partie entière inférieure de la moitié du terme précédent. Ceci est vrai pour tout  $n$  car on déduit de (3.11) que  $\lfloor n/2^{k+1} \rfloor = \lfloor \lfloor n/2^k \rfloor / 2 \rfloor$ . On voit très bien ce qui se passe si on écrit les nombres en binaire :

$$100 = (1100100)_2 = 100$$

$$\lfloor 100/2 \rfloor = (110010)_2 = 50$$

$$\begin{aligned}
 \lfloor 100/4 \rfloor &= (11001)_2 = 25 \\
 \lfloor 100/8 \rfloor &= (1100)_2 = 12 \\
 \lfloor 100/16 \rfloor &= (110)_2 = 6 \\
 \lfloor 100/32 \rfloor &= (11)_2 = 3 \\
 \lfloor 100/64 \rfloor &= (1)_2 = 1
 \end{aligned}$$

Pour obtenir un terme, il suffit de supprimer le bit de poids faible (c'est-à-dire le bit de droite) du terme précédent.

Cette représentation binaire nous permet aussi de voir que

$$\epsilon_2(n!) = n - v_2(n), \quad (4.24)$$

où  $v_2(n)$  est le nombre de 1 dans l'écriture binaire de  $n$ . Cela est vrai car chaque 1 qui contribue pour  $2^m$  à la valeur de  $n$  contribue pour  $2^{m-1} + 2^{m-2} + \dots + 2^0 = 2^m - 1$  à la valeur de  $\epsilon_2(n!)$ .

Avec un raisonnement similaire, on peut généraliser ce qui précède pour tout entier  $p$  premier :

$$\epsilon_p(n!) = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \dots = \sum_{k \geq 1} \left\lfloor \frac{n}{p^k} \right\rfloor. \quad (4.25)$$

Quel est l'ordre de grandeur de  $\epsilon_p(n!)$ ? On trouve tout simplement un bon majorant en supprimant les crochets de partie entière du terme général ; on n'a plus alors qu'à sommer une progression géométrique infinie :

$$\begin{aligned}
 \epsilon_p(n!) &< \frac{n}{p} + \frac{n}{p^2} + \frac{n}{p^3} + \dots \\
 &= \frac{n}{p} \left( 1 + \frac{1}{p} + \frac{1}{p^2} + \dots \right) \\
 &= \frac{n}{p} \left( \frac{p}{p-1} \right) \\
 &= \frac{n}{p-1}.
 \end{aligned}$$

Pour  $p = 2$  et  $n = 100$ , cette inégalité donne  $97 < 100$ . Ainsi, non seulement le majorant 100 est correct, mais de plus il est proche de la valeur exacte 97. En fait, la valeur exacte  $n - v_2(n)$  est de façon générale très proche de  $n$  car  $v_2(n) \leq \lceil \lg n \rceil$  est asymptotiquement beaucoup plus petit que  $n$ .

En appliquant nos formules à  $p = 2$  et  $p = 3$ , on trouve  $\epsilon_2(n!) \sim n$  et  $\epsilon_3(n!) \sim n/2$ . Il semblerait donc raisonnable que, de temps en temps,  $\epsilon_3(n!)$  soit exactement égal à la moitié de  $\epsilon_2(n!)$ . Cela arrive par exemple quand  $n = 6$  ou  $n = 7$ , car  $6! = 2^4 \cdot 3^2 \cdot 5 = 7!/7$ . Mais personne n'a pu prouver jusqu'à présent qu'il existe un nombre infini de telles coïncidences.

On déduit du majorant de  $\epsilon_p(n!)$  un majorant pour  $p^{\epsilon_p(n!)}$ , qui représente la contribution de  $p$  à  $n!$  :

$$p^{\epsilon_p(n!)} < p^{n/(p-1)}.$$

On peut simplifier cette formule (au risque d'éloigner considérablement le majorant) en remarquant que  $p \leq 2^{p-1}$ . Par conséquent,  $p^{n/(p-1)} \leq (2^{p-1})^{n/(p-1)} = 2^n$ . En d'autres termes, la contribution de  $n$ 'importe quel nombre premier à  $n!$  est plus petite que  $2^n$ .

Cette observation permet d'écrire une nouvelle preuve du fait qu'il existe une infinité de nombres premiers. En effet, s'il existait seulement  $k$  nombres premiers  $2, 3, \dots, P_k$ , alors on aurait  $n! < (2^n)^k = 2^{nk}$  pour tout  $n > 1$ , car chaque nombre premier ne peut contribuer que pour un facteur  $2^n - 1$  au maximum. Toutefois, l'inégalité  $n! < 2^{nk}$  est facile à contredire si on prend un  $n$  assez grand, par exemple  $n = 2^{2^k}$ . Dans ce cas,

$$n! < 2^{nk} = 2^{2^{2^k}} = n^{n/2},$$

ce qui contredit l'inégalité  $n! \geq n^{n/2}$  de (4.22). Il existe donc toujours une infinité de nombres premiers.

On peut même utiliser cet argument pour obtenir un minorant grossier de  $\pi(n)$ , le nombre de nombres premiers inférieurs ou égaux à  $n$ . Chacun de ces nombres contribue à  $n!$  pour un facteur plus petit que  $2^n$ , donc

$$n! < 2^{n\pi(n)}.$$

En remplaçant  $n!$  par son approximation selon la formule de Stirling (4.23) qui est un minorant, puis en prenant le logarithme, on obtient

$$n\pi(n) > n \lg(n/e) + \frac{1}{2} \lg(2\pi n);$$

et donc

$$\pi(n) > \lg(n/e).$$

Ce minorant est plutôt faible en comparaison de la valeur effective  $\pi(n) \sim n/\ln n$ , car  $\log n$  est bien plus petit que  $n/\log n$  lorsque  $n$  est grand. Cependant, nous l'avons trouvé sans trop d'efforts ; et après tout, un minorant est un minorant.

## 4.5 PRIMALITÉ RELATIVE

Lorsque  $\text{pgcd}(m, n) = 1$ , les entiers  $m$  et  $n$  n'ont aucun facteur premier en commun. Dans ce cas, on dit qu'ils sont *premiers entre eux*.

Ce concept est si important qu'il nous faut une notation spéciale pour le désigner. Hélas, les théoriciens des nombres n'ont pas encore pu se mettre d'accord sur une notation vraiment adéquate. C'est pourquoi nous lançons un appel : ECOUTEZ-NOUS, O MATHÉMATICIENS DU MONDE ! NE NOUS FAITES PAS LANGUIR DAVANTAGE ! NOUS POUVONS RENDRE PLUS CLAIRES BIEN DES FORMULES EN ADOPTANT UNE NOUVELLE NOTATION DÈS MAIN-TENANT ! PERMETTEZ-NOUS D'ÉCRIRE " $m \perp n$ " ET DE DIRE " $m$  EST PREMIER PAR RAPPORT À  $n$ " SI  $m$  ET  $n$  SONT PREMIERS ENTRE EUX. Autrement dit, déclarons que

$$m \perp n \iff m, n \text{ sont des entiers et } \text{pgcd}(m, n) = 1. \quad (4.26)$$

Une fraction  $m/n$  est réduite si et seulement si  $m \perp n$ . Pour réduire une fraction, il faut diviser son numérateur et son dénominateur par leur plus grand facteur commun. On peut donc raisonnablement penser que

$$\frac{m}{\text{pgcd}(m, n)} \perp \frac{n}{\text{pgcd}(m, n)}; \quad (4.27)$$

et c'est effectivement vrai. C'est une conséquence d'une règle plus générale selon laquelle  $\text{pgcd}(km, kn) = k \text{pgcd}(m, n)$  ; on la prouve dans l'exercice 14.

La relation  $\perp$  s'exprime simplement si on travaille sur les représentations en puissances premières des nombres, grâce à la règle (4.14) :

$$m \perp n \iff \min(m_p, n_p) = 0 \text{ pour tout } p. \quad (4.28)$$

*Le produit terme à terme est nul, exactement comme pour deux vecteurs orthogonaux.*

Comme  $m_p$  et  $n_p$  sont positifs ou nuls, on peut l'écrire

$$m \perp n \iff m_p n_p = 0 \text{ pour tout } p. \quad (4.29)$$

Nous avons maintenant les moyens de démontrer une règle importante que nous pourrons utiliser pour combiner deux relations  $\perp$  ayant le même membre gauche :

$$k \perp m \text{ et } k \perp n \iff k \perp mn. \quad (4.30)$$

Cela revient à dire que, lorsque  $m_p$  et  $n_p$  sont positifs ou nuls,  $k_p m_p = 0$  et  $k_p n_p = 0$  si et seulement si  $k_p(m_p + n_p) = 0$ .

*Il est intéressant de noter que les mathématiciens disent "découvert", alors que n'importe qui d'autre dirait "inventé".*

Il existe une manière élégante de construire l'ensemble de toutes les fractions positives ou nulles  $m/n$  avec  $m \perp n$  : c'est l'*arbre de Stern-Brocot*, ainsi appelé parce qu'il a été découvert indépendamment par Moritz Stern [339], un mathématicien allemand, et Achille Brocot [40], un horloger français. On part des deux fractions  $(\frac{0}{1}, \frac{1}{0})$ , puis on répète l'opération suivante autant de fois qu'on veut :

$$\text{insérer } \frac{m+m'}{n+n'} \text{ entre deux fractions adjacentes } \frac{m}{n} \text{ et } \frac{m'}{n'}.$$

La nouvelle fraction  $(m + m')/(n + n')$  est appelée le *médiant* de  $m/n$  et  $m'/n'$ . Par exemple, la première étape consiste à ajouter une nouvelle fraction entre  $\frac{0}{1}$  et  $\frac{1}{0}$ ,

$$\frac{0}{1}, \frac{1}{1}, \frac{1}{0};$$

et l'étape suivante nous en donne deux de plus :

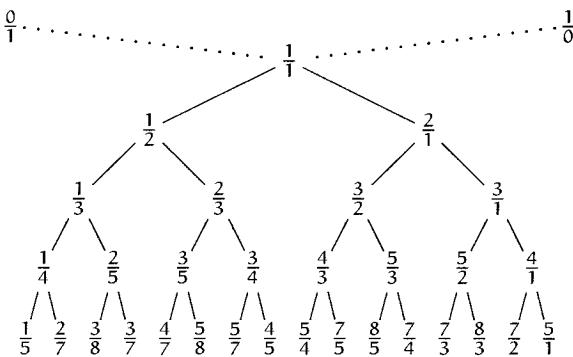
$$\frac{0}{1}, \frac{1}{2}, \frac{1}{1}, \frac{2}{1}, \frac{1}{0}.$$

On en trouve quatre autres à la troisième étape,

$$\frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}, \frac{3}{2}, \frac{2}{1}, \frac{3}{1}, \frac{1}{0};$$

puis 8, 16 et ainsi de suite. L'ensemble de ces fractions peut être vu comme un arbre binaire infini dont voici les 5 premiers niveaux :

*J'imagine que la  
“réduite” de 1/0  
est l'infini.*



Chaque fraction est de la forme  $\frac{m+m'}{n+n'}$ , où  $\frac{m}{n}$  est le plus proche ancêtre de  $\frac{m+m'}{n+n'}$  situé à sa gauche, et  $\frac{m'}{n'}$  est son plus proche ancêtre de droite (un "ancêtre" est une fraction qu'on peut atteindre en remontant le long des branches de l'arbre).

Pourquoi cette construction marche-t-elle si bien ? Pourquoi, par exemple, chacun des médiants  $(m + m')/(n + n')$  apparaît-il dans l'arbre sous sa forme réduite ? (Si  $m$ ,  $m'$ ,  $n$  et  $n'$  étaient impairs, on obtiendrait des fractions de type pair/pair ; la construction semble garantir que deux fractions ayant un numérateur et un dénominateur impair n'apparaissent jamais côté à côté). Pourquoi toutes les fractions possibles  $m/n$  apparaissent-elles chacune une fois exactement ? Qu'est-ce qui les empêche d'apparaître deux fois ou pas du tout ?

Etonnamment, il existe des réponses simples à toutes ces questions ; elles sont basées sur le fait crucial que voici : si  $m/n$  et  $m'/n'$  sont deux fractions consécutives sur un même niveau de l'arbre, alors

$$m'n - mn' = 1. \quad (4.31)$$

Cette relation est vraie au départ ( $1 \cdot 1 - 0 \cdot 0 = 1$ ) ; lorsqu'on insère un nouveau médiant  $(m + m')/(n + n')$ , il suffit de vérifier les formules suivantes :

$$(m + m')n - m(n + n') = 1;$$

$$m'(n + n') - (m + m')n' = 1.$$

Chacune de ces équations est équivalente à la condition d'origine (4.31) qu'elles sont sensées remplacer. La formule (4.31) est donc un invariant, quel que soit l'avancement de la construction de l'arbre.

De plus, si  $m/n < m'/n'$  et si toutes les valeurs sont positives ou nulles, on peut vérifier facilement que

$$m/n < (m + m')/(n + n') < m'/n'.$$

Bien qu'un médiant ne se situe pas toujours exactement au milieu des deux fractions qui lui ont donné naissance, il se trouve quand même quelque part entre les deux. L'ordre est donc préservé au cours de la construction ; c'est pourquoi il n'est pas possible de trouver une même fraction en deux endroits différents.

Il reste encore une question non résolue. Existe-t-il une fraction  $a/b$  telle que  $a \perp b$  qui n'est pas dans l'arbre ? La réponse est non. Pour prouver cela, on peut se contenter d'étudier la construction au voisinage immédiat de  $a/b$ . Au départ, on a

$$\frac{m}{n} = \frac{0}{1} < \left(\frac{a}{b}\right) < \frac{1}{0} = \frac{m'}{n'}.$$

Les parenthèses encadrant  $\frac{a}{b}$  indiquent que cette fraction n'est pas encore présente dans la construction. Si, à une étape donnée, on a la configuration

$$\frac{m}{n} < \left(\frac{a}{b}\right) < \frac{m'}{n'},$$

alors on engendre  $(m + m')/(n + n')$  et trois cas se présentent : soit  $(m + m')/(n + n') = a/b$  et nous avons gagné ; soit  $(m + m')/(n + n') < a/b$  et on peut poser  $m \leftarrow m + m'$ ,  $n \leftarrow n + n'$  ; soit  $(m + m')/(n + n') > a/b$  et on pose  $m' \leftarrow m + m'$ ,  $n' \leftarrow n + n'$ . Ce processus ne peut pas durer indéfiniment, car les conditions

$$\frac{a}{b} - \frac{m}{n} > 0 \quad \text{et} \quad \frac{m'}{n'} - \frac{a}{b} > 0$$

entraînent que

$$an - bm \geq 1 \quad \text{et} \quad bm' - an' \geq 1.$$

Par conséquent

$$(m' + n')(an - bm) + (m + n)(bm' - an') \geq m' + n' + m + n,$$

*C'est vrai ; mais en cas de fracture multiple, il vaut mieux aller chez un médecin.*

ce qui, en vertu de (4.31), est équivalent à  $a+b \geq m'+n'+m+n$ . Comme il existe à chaque étape une valeur parmi  $m$ ,  $n$ ,  $m'$  et  $n'$  qui est augmentée, on gagne forcément en au plus  $a+b$  étapes.

La *suite de Farey* d'ordre  $N$ , que l'on note  $\mathcal{F}_N$ , est la suite croissante des fractions réduites comprises entre 0 et 1 dont le dénominateur est inférieur ou égal à  $N$ . Par exemple,

$$\mathcal{F}_6 = \frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1}.$$

De façon générale,  $\mathcal{F}_N$  s'obtient en engendrant, à partir  $\mathcal{F}_1 = \frac{0}{1}, \frac{1}{1}$ , tous les médians dont le dénominateur ne dépasse pas  $N$ . Ainsi on est assuré de n'oublier aucune fraction, du fait que, dans la construction de Stern–Brocot, un médiant de dénominateur  $\leq N$  n'est jamais formé à partir d'une fraction de dénominateur  $> N$ . (En d'autres termes,  $\mathcal{F}_N$  détermine un *sous-arbre* de l'arbre de Stern–Brocot, obtenu en coupant les branches non désirées). Il s'ensuit que, si  $m/n$  et  $m'/n'$  sont deux éléments consécutifs d'une suite de Farey, alors  $m'n - mn' = 1$ .

D'après cette méthode de construction, il est facile de construire  $\mathcal{F}_N$  à partir de  $\mathcal{F}_{N-1}$  : il suffit d'insérer la fraction  $(m+m')/N$  chaque fois qu'il existe dans  $\mathcal{F}_{N-1}$  deux fractions successives  $m/n$  et  $m'/n'$  dont la somme des dénominateurs est égale à  $N$ . Par exemple, on passe de  $\mathcal{F}_6$  à  $\mathcal{F}_7$  en insérant  $\frac{1}{7}, \frac{2}{7}, \dots, \frac{6}{7}$  selon la règle donnée :

$$\mathcal{F}_7 = \frac{0}{1}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \frac{1}{1}.$$

Si  $N$  est premier, on ajoutera  $N-1$  nouvelles fractions ; sinon il y en aura moins, car le processus ne crée que des fractions dont les numérateurs sont premiers par rapport à  $N$ .

Il y a un certain temps, nous avons prouvé dans (4.5) que, pour tous  $m \perp n$  et  $0 < m \leq n$ , on peut trouver des entiers  $a$  et  $b$  tels que

$$ma - nb = 1. \tag{4.32}$$

(Pour être exacts, nous avions écrit  $m'm + n'n = \text{pgcd}(m, n)$  ; mais on peut remplacer  $\text{pgcd}(m, n)$  par 1,  $m'$  par  $a$  et  $-n'$  par  $b$ ). Si on suppose que  $b/a$  est la fraction qui précède  $m/n$  dans  $\mathcal{F}_n$ , on obtient une autre preuve de (4.32). En fait, (4.5) est simplement une autre forme de (4.31). Par exemple  $a=5$  et  $b=2$  satisfont  $3a - 7b = 1$  car  $\frac{2}{5}$  précède  $\frac{3}{7}$  dans  $\mathcal{F}_7$ . De par cette construction, si  $0 < m \leq n$ , on est assuré de toujours pouvoir trouver une solution de (4.32) telle que  $0 \leq b < a < n$ . Similairement, si  $0 \leq n < m$  et  $m \perp n$ , on peut résoudre (4.32) de façon que  $0 < a \leq b \leq m$  en supposant que  $a/b$  est la fraction qui suit  $n/m$  dans  $\mathcal{F}_m$ .

Les suites de trois termes consécutifs d'une suite de Farey possèdent une étonnante propriété ; on la découvrira dans l'exercice 61. Cessons

Garez Farey.

maintenant de regarder les suites de Farey et penchons-nous plutôt sur l'arbre de Stern–Brocot dans son entier ; cela va s'avérer plus intéressant encore.

Toute fraction positive réduite apparaît une et une seule fois dans l'arbre de Stern–Brocot. On peut donc le considérer comme un *système de numération* pour les nombres rationnels. Pour cela, considérons la suite des pas “vers la gauche” ou “vers la droite” à faire pour parcourir l'arbre en descendant de la racine vers une fraction donnée, et convenons d'écrire la lettre L ou R respectivement pour indiquer que l'on fait un pas vers la gauche ou vers la droite en parcourant. Ainsi, tout mot formé d'une suite de L et de R détermine un unique emplacement dans l'arbre. Par exemple, le mot LRRL s'interprète ainsi : on descend à gauche de  $\frac{1}{1}$  vers  $\frac{1}{2}$ , puis à droite vers  $\frac{2}{3}$ , puis encore à droite vers  $\frac{3}{4}$ , enfin à gauche pour arriver à  $\frac{5}{7}$ . En ce sens, LRRL constitue une représentation de  $\frac{5}{7}$ . Toute fraction strictement positive peut ainsi être représentée par une unique suite de L et de R.

Il y a cependant un petit problème : à la fraction  $\frac{1}{1}$  correspond le mot *vide* ; il nous faut donc une notation pour cela. Convenons de l'appeler I, parce que cela ressemble à 1 et que cela représente “l'identité”.

Cette représentation soulève naturellement deux questions : (1) Etant donnés deux entiers strictement positifs m et n tels que  $m \perp n$ , quel est le mot qui correspond à  $m/n$  ? (2) Etant donné un mot formé de lettres L et R, quelle fraction représente-t-il ? La question 2 semble plus facile ; commençons donc par elle. Si S est un mot de lettres L et R, soit

$$f(S) = \text{la fraction correspondant à } S.$$

$$\text{Par exemple, } f(\text{LRRL}) = \frac{5}{7}.$$

D'après la construction de l'arbre,  $f(S) = (m + m')/(n + n')$  si  $m/n$  et  $m'/n'$  sont les deux fractions les plus proches de S qui se trouvent dans un niveau supérieur. Au départ,  $m/n = 0/1$  et  $m'/n' = 1/0$  ; puis, chaque fois qu'on fait un pas en descendant vers la droite ou vers la gauche, on remplace  $m/n$  ou  $m'/n'$  respectivement par le médiant  $(m + m')/(n + n')$ .

Ce serait plus pratique si nous pouvions décrire cela par des formules mathématiques. Après quelques essais, il semble que la meilleure manière consiste à définir une matrice  $2 \times 2$

$$M(S) = \begin{pmatrix} n & n' \\ m & m' \end{pmatrix}$$

qui représente les quatre nombres présents dans les fractions  $m/n$  et  $m'/n'$  qui engendent  $f(S)$ . On aurait pu mettre les m en haut et les n en bas, comme dans les fractions ; mais il s'avère plus profitable d'inverser les choses, car, au tout début du processus, on a  $M(I) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ , que l'on appelle la matrice identité, et que l'on note traditionnellement I.

En faisant un pas vers la gauche, on remplace  $n'$  par  $n + n'$  et  $m'$  par  $m + m'$ ; donc

$$\begin{aligned} M(SL) &= \begin{pmatrix} n & n+n' \\ m & m+m' \end{pmatrix} = \begin{pmatrix} n & n' \\ m & m' \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &= M(S) \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

(C'est un cas particulier de la règle de multiplication

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} w & x \\ y & z \end{pmatrix} = \begin{pmatrix} aw+by & ax+bz \\ cw+dy & cx+dz \end{pmatrix}$$

pour les matrices  $2 \times 2$ ). De façon similaire, on trouve que

$$M(SR) = \begin{pmatrix} n+n' & n' \\ m+m' & m' \end{pmatrix} = M(S) \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Donc, si on définit L et R comme des matrices  $2 \times 2$ ,

$$L = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad (4.33)$$

on obtient, par induction sur la longueur de S, la formule toute simple :  $M(S) = S$ . N'est-ce pas agréable ? (Les lettres L et R représentent à la fois des matrices et des lettres d'un mot). Par exemple,

$$M(LRRL) = LRRL = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix};$$

les fractions qui encadrent et engendent  $LRRL = \frac{5}{7}$  sont  $\frac{2}{3}$  et  $\frac{3}{4}$ . Nous pouvons maintenant répondre à la question 2 :

$$f(S) = f\left(\begin{pmatrix} n & n' \\ m & m' \end{pmatrix}\right) = \frac{m+m'}{n+n'}. \quad (4.34)$$

Qu'en est-il de la question 1 ? La réponse vient facilement, maintenant que nous avons compris le lien essentiel entre les sommets de l'arbre et les matrices. Etant donné un couple d'entiers strictement positifs  $m$  et  $n$  tels que  $m \perp n$ , on trouve l'emplacement de  $m/n$  dans l'arbre de Stern–Brocot en effectuant une "recherche binaire" de la façon suivante :

```

S := I;
tant que m/n ≠ f(S) faire
    si m/n < f(S) alors (écrire(L); S := SL)
    sinon (écrire(R); S := SR).

```

*Si vous ne connaissez rien aux matrices, pas de panique. C'est le seul endroit où on en trouve dans ce livre.*

Cet algorithme écrit la suite de L et de R recherchée.

On peut aussi opérer différemment, en modifiant directement  $m$  et  $n$  au lieu de travailler sur  $S$ . Pour toute matrice  $2 \times 2 S$ , on a

$$f(RS) = f(S) + 1$$

car  $RS$  s'obtient en ajoutant à la deuxième ligne de  $S$  sa première ligne. (Voyons cela au ralenti :

$$S = \begin{pmatrix} n & n' \\ m & m' \end{pmatrix}; \quad RS = \begin{pmatrix} n & n' \\ m+n & m'+n' \end{pmatrix};$$

donc  $f(S) = (m+m')/(n+n')$  et  $f(RS) = ((m+n)+(m'+n'))/(n+n')$ . Si on applique l'algorithme de recherche binaire à une fraction  $m/n$  telle que  $m > n$ , la première lettre affichée sera  $R$  ; par conséquent, on trouvera avec l'algorithme exactement 1 de plus que ce qu'on aurait trouvé en prenant  $(m-n)/n$  au lieu de  $m/n$ . En faisant un raisonnement similaire pour  $L$ , on obtient

$$\frac{m}{n} = f(RS) \iff \frac{m-n}{n} = f(S), \quad \text{si } m > n;$$

$$\frac{m}{n} = f(LS) \iff \frac{m}{n-m} = f(S), \quad \text{si } m < n.$$

On peut alors écrire l'algorithme de recherche binaire sans faire appel aux matrices :

```
tant que  $m \neq n$  faire
    si  $m < n$  alors (écrire(L);  $n := n - m$ )
        sinon (écrire(R);  $m := m - n$ ) .
```

Par exemple, en partant de  $m/n = 5/7$ , on obtient successivement

$$m = 5 \quad 5 \quad 3 \quad 1 \quad 1$$

$$n = 7 \quad 2 \quad 2 \quad 2 \quad 1$$

affichage L R R L

avec cet algorithme.

L'arbre de Stern-Brocot ne contient évidemment aucun nombre irrationnel ; on y trouve cependant tous les rationnels “proches” de tout irrationnel donné. Par exemple, si on lance l'algorithme à la recherche du nombre  $e = 2.71828\dots$ , on obtient une suite infinie de L et de R qui commence par

RRLRRRLRLLLRLRRRRRLRLLLLLRLR ... .

On peut considérer que ce mot infini constitue la représentation de  $e$  dans le système de numération de Stern–Brocot. Cette représentation est aussi pertinente que sa représentation par un nombre décimal ( $2.718281828459\dots$ ) ou par un nombre binaire ( $(10.10110111110\dots)_2$ ) infinis. De plus, il se trouve que la représentation de  $e$  dans le système de Stern–Brocot présente une grande régularité :

$$e = RL^0 RLR^2 LRL^4 RLR^6 LRL^8 RLR^{10} LRL^{12} RL \dots ;$$

cette propriété se déduit d'un cas particulier d'un résultat établi par Euler [105] alors qu'il avait 24 ans.

On déduit de cette représentation que les fractions

$$\begin{array}{ccccccccccccccccccccc} R & R & L & R & R & L & R & L & L & L & L & L & R & L & R & R & R & R \\ \frac{1}{1}, \frac{2}{1}, \frac{3}{1}, \frac{5}{2}, \frac{8}{3}, \frac{11}{4}, \frac{19}{7}, \frac{30}{11}, \frac{49}{18}, \frac{68}{25}, \frac{87}{32}, \frac{106}{39}, \frac{193}{71}, \frac{299}{110}, \frac{492}{181}, \frac{685}{252}, \frac{878}{323}, \dots \end{array}$$

constituent les approximations rationnelles les plus simples de  $e$  (c'est-à-dire dont les numérateurs et dénominateurs sont les plus petits possibles). En effet, si une fraction donnée  $m/n$  n'est pas dans cette liste, alors il y existe forcément une fraction, de numérateur  $\leq m$  et de dénominateur  $\leq n$ , qui se trouve dans l'intervalle délimité par  $m/n$  et  $e$ . Par exemple,  $\frac{27}{10}$ , qui n'est pas dans la liste, est une approximation moins simple et moins proche de  $e$  que  $\frac{19}{7} = 2.714\dots$ , qui y apparaît. Cette propriété s'explique par le fait que l'arbre de Stern–Brocot ne se contente pas de contenir tous les rationnels : ils y apparaissent dans l'ordre, et les fractions situées sur les niveaux les plus hauts sont celles qui ont les plus petits numérateurs et dénominateurs. Ainsi, la fraction  $\frac{27}{10} = RRLRRL$  est plus petite que  $\frac{19}{7} = RRLRRL$ , elle-même inférieure à  $e = RRLRRL\dots$ . On obtient de cette façon d'excellentes approximations. Par exemple,  $\frac{878}{323} \approx 2.718266 \approx .999994e$  ; ce résultat est obtenu en prenant les 16 premières lettres de la représentation de  $e$  ; sa précision est similaire à celle que l'on obtiendrait en prenant les 16 premiers bits de la représentation binaire de  $e$ .

Une simple modification de l'algorithme de recherche binaire suffit pour trouver la représentation infinie de n'importe quel nombre irrationnel  $\alpha$  :

**si**  $\alpha < 1$  **alors** (écrire(L);  $\alpha := \alpha/(1 - \alpha)$ )

**sinon** (écrire(R);  $\alpha := \alpha - 1$ ) .

(Ce traitement doit être effectué un nombre infini de fois, ou, à défaut, jusqu'à ce qu'on soit trop fatigué). Si  $\alpha$  est rationnel, la représentation obtenue est la même que précédemment, à ceci près qu'on y ajoute la suite  $RL^\infty$ . Par exemple, si  $\alpha = 1$ , on obtient  $RLLL\dots$ , ce qui correspond à la suite infinie de fractions  $\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \dots$ , dont la limite tend vers 1 à l'infini. C'est tout à fait similaire à ce qui se passe pour la représentation

Hermann Minkowski a présenté cette remarquable représentation binaire au Congrès International de Mathématiques de Heidelberg en 1904.

binaire habituelle. On sait que tout nombre réel  $x \in [0..1)$  possède une représentation binaire infinie  $(.b_1 b_2 b_3 \dots)_2$  qui ne se termine pas par une suite infinie de 1 ; de même, tout réel  $\alpha \in [0..\infty)$  possède une représentation de Stern–Brocot infinie  $B_1 B_2 B_3 \dots$  qui ne se termine pas par une suite infinie de R. Il existe donc une bijection, qui préserve l'ordre, entre  $[0..1)$  et  $[0..\infty)$  : il suffit de poser  $0 \leftrightarrow L$  et  $1 \leftrightarrow R$ .

La représentation de Stern–Brocot est étroitement liée à l'algorithme d'Euclide. Si  $\alpha = m/n$ , sa représentation est formée de  $\lfloor m/n \rfloor$  lettres R, suivies de  $\lfloor n/(m \bmod n) \rfloor$  lettres L, puis de  $\lfloor (m \bmod n)/(n \bmod (m \bmod n)) \rfloor$  lettres R, etc. Ces nombres,  $m \bmod n$ ,  $n \bmod (m \bmod n)$ , ... sont exactement les valeurs calculées au cours de l'algorithme d'Euclide (il y a juste un petit traitement particulier à faire à la fin pour éviter d'avoir un nombre infini de lettres R). Nous nous intéresserons de près à ce lien au cours du chapitre 6.

## 4.6 “MOD” : LA CONGRUENCE

L'arithmétique modulaire est l'un des outils essentiels que nous offre la théorie des nombres. Au chapitre 3, nous avons utilisé l'opération binaire “mod” dans des expressions, tout comme n'importe quelle opération usuelle. Dans le présent chapitre, “mod” va aussi nous servir à écrire des équations. Pour cela, nous allons convenir d'une notation plus pratique :

$$a \equiv b \pmod{m} \iff a \bmod m = b \bmod m. \quad (4.35)$$

Par exemple,  $9 \equiv -16 \pmod{5}$  car  $9 \bmod 5 = 4 = (-16) \bmod 5$ . L'expression “ $a \equiv b \pmod{m}$ ” se dit “ $a$  est congru à  $b$  modulo  $m$ ”. Bien que cette définition soit valable pour tous réels  $a$ ,  $b$ , et  $m$ , nous ne considérerons que des nombres entiers.

Voici une autre définition, équivalente à la première, de la congruence :

$$a \equiv b \pmod{m} \iff a - b \text{ est un multiple de } m. \quad (4.36)$$

En effet, si  $a \bmod m = b \bmod m$ , alors il découle de la définition de “mod” en (3.21) qu'il existe deux entiers  $k$  et  $l$  tels que  $a - b = a \bmod m + km - (b \bmod m + lm) = (k - l)m$ . Inversement, si  $a - b = km$ , alors  $a = b$  si  $m = 0$ , et sinon,

$$\begin{aligned} a \bmod m &= a - \lfloor a/m \rfloor m = b + km - \lfloor (b + km)/m \rfloor m \\ &= b - \lfloor b/m \rfloor m = b \bmod m. \end{aligned}$$

Il est souvent plus pratique d'appliquer la définition de (4.36) plutôt que celle de (4.35). Par exemple,  $8 \equiv 23 \pmod{5}$  du fait que  $8 - 23 = -15$  est un multiple de 5 ; pas besoin de calculer  $8 \bmod 5$  et  $23 \bmod 5$ .

“Numerorum  
congruentiam hoc si-  
gno,  $\equiv$ , in posterum  
denotabimus, mo-  
dulum ubi opus erit  
in clausulis adiun-  
gentes,  $-16 \equiv 9$   
(mod. 5),  $-7 \equiv$   
 $15 \pmod{11}$ .”

—C.F. Gauss [142]

Si le symbole de congruence, “ $\equiv$ ”, ressemble à “ $=$ ”, ce n'est pas par hasard : c'est parce que les congruences sont presque semblables à des équations. C'est ainsi que toute congruence est une *relation d'équivalence*, c'est-à-dire qu'elle est réflexive ( $a \equiv a$ ), symétrique ( $a \equiv b \Rightarrow b \equiv a$ ) et transitive ( $a \equiv b \equiv c \Rightarrow a \equiv c$ ). Ces propriétés sont faciles à prouver, car toute relation “ $\equiv$ ” satisfaisant “ $a \equiv b \iff f(a) = f(b)$ ” pour une fonction  $f$  donnée est une relation d'équivalence (dans notre cas,  $f(x) = x \bmod m$ ). En outre, l'addition et la soustraction d'éléments congrus n'affecte pas la congruence :

$$\begin{aligned} a \equiv b \text{ et } c \equiv d &\implies a + c \equiv b + d \pmod{m}; \\ a \equiv b \text{ et } c \equiv d &\implies a - c \equiv b - d \pmod{m}. \end{aligned}$$

Ceci est vrai parce que, si  $a - b$  et  $c - d$  sont deux multiples de  $m$ , les nombres  $(a+c)-(b+d) = (a-b)+(c-d)$  et  $(a-c)-(b-d) = (a-b)-(c-d)$  le sont aussi. Remarquons qu'il n'est pas nécessaire de préciser “ $(\bmod m)$ ” chaque fois qu'on écrit “ $\equiv$ ” ; si le module est constant, il suffit de l'écrire une seule fois pour préciser le contexte. C'est un des attraits de la notation “ $\equiv$ ”.

Il y a une propriété similaire pour la multiplication, pourvu que l'on travaille avec des entiers :

$$a \equiv b \text{ et } c \equiv d \implies ac \equiv bd \pmod{m},$$

$a, b, c$  entiers.

Preuve :  $ac - bd = (a - b)c + b(c - d)$ . On en déduit, en appliquant plusieurs fois cette propriété, que

$$a \equiv b \implies a^n \equiv b^n \pmod{m},$$

$a, b$  entiers ;  
 $n \geq 0$  entier.

Par exemple, comme  $2 \equiv -1 \pmod{3}$ , on a  $2^n \equiv (-1)^n \pmod{3}$  ; cela signifie que  $2^n - 1$  est un multiple de 3 si et seulement si  $n$  est pair.

Ainsi, presque toutes les opérations algébriques habituellement utilisées pour manipuler des équations conviennent aussi aux congruences. Pas toutes cependant : la division en est une exception notable. Si  $ad \equiv bd \pmod{m}$ , il n'est pas toujours vrai que  $a \equiv b$ . Par exemple,  $3 \cdot 2 \equiv 5 \cdot 2 \pmod{4}$  et  $3 \not\equiv 5$ .

Toutefois, la propriété est vraie si on ajoute une hypothèse supplémentaire :  $d$  et  $m$  doivent être premiers entre eux.

$$ad \equiv bd \iff a \equiv b \pmod{m},$$

$a, b, d, m$  entiers et  $d \perp m$ . (4.37)

Par exemple, du fait que  $15 \equiv 35 \pmod{m}$ , on peut déduire que  $3 \equiv 7 \pmod{m}$ , sauf si  $m$  est un multiple de 5.

*“I feel fine today modulo a slight headache.”*

— The Hacker's Dictionary [337]

Pour démontrer ceci, on utilise encore une fois la règle (4.5) pour trouver deux entiers  $d'$  et  $m'$  tels que  $d'd + m'm = 1$ . Alors, si  $ad \equiv bd$ , on peut multiplier les deux côtés de la congruence par  $d'$ , obtenant ainsi  $ad'd \equiv bd'd$ . Comme  $d'd \equiv 1$ , on a  $ad'd \equiv a$  et  $bd'd \equiv b$  ; donc  $a \equiv b$ . On voit dans cette preuve que le nombre  $d'$  se comporte presque comme  $1/d$  ; pour cette raison,  $d'$  est appelé “l’inverse de  $d$  modulo  $m$ ”.

Une autre façon d’appliquer la division aux congruences est de diviser aussi le module :

$$ad \equiv bd \pmod{md} \iff a \equiv b \pmod{m}, \text{ pour } d \neq 0. \quad (4.38)$$

Cette règle est valable pour tous réels  $a, b, d$  et  $m$ , car elle se démontre en utilisant uniquement la règle de distributivité :  $(a \pmod{m})d = ad \pmod{md}$ . En effet,  $a \pmod{m} = b \pmod{m} \iff (a \pmod{m})d = (b \pmod{m})d \iff ad \pmod{md} = bd \pmod{md}$ . Ainsi, par exemple, on déduit du fait que  $3 \cdot 2 \equiv 5 \cdot 2 \pmod{4}$  que  $3 \equiv 5 \pmod{2}$ .

Le propriétés (4.37) et (4.38) peuvent être combinées pour obtenir une règle qui diminue autant que possible la valeur du module :

$$\begin{aligned} ad &\equiv bd \pmod{m} \\ \iff a &\equiv b \pmod{\frac{m}{\text{pgcd}(d, m)}}, \quad a, b, d, m \text{ entiers.} \end{aligned} \quad (4.39)$$

Pour prouver cela, on multiplie  $ad \equiv bd$  par le nombre  $d'$  tel que  $d'd + m'm = \text{pgcd}(d, m)$  ; cela nous donne  $a \cdot \text{pgcd}(d, m) \equiv b \cdot \text{pgcd}(d, m) \pmod{m}$ , et il ne reste plus qu’à diviser le tout par  $\text{pgcd}(d, m)$ .

Examinons d’un peu plus près cette idée de modifier le module. Si on  $a \equiv b \pmod{100}$ , alors nécessairement  $a \equiv b \pmod{10}$ , ou modulo n’importe quel diviseur de 100. Le fait que  $a - b$  est un multiple de 100 est plus fort que le fait qu’il est multiple de 10. Comme tout multiple de  $md$  est un multiple de  $d$ , on a

$$a \equiv b \pmod{md} \implies a \equiv b \pmod{m}, \quad d \text{ entier.} \quad (4.40)$$

Inversement, si on sait que  $a \equiv b$  modulo deux petits nombres, peut-on en déduire que  $a \equiv b$  modulo un nombre plus grand ? La réponse est oui :

$$\begin{aligned} a &\equiv b \pmod{m} \quad \text{et} \quad a \equiv b \pmod{n} \\ \iff a &\equiv b \pmod{\text{ppcm}(m, n)}, \quad m, n > 0 \text{ entiers.} \end{aligned} \quad (4.41)$$

Par exemple, si on sait que  $a \equiv b$  modulo 12 et 18, on peut en conclure avec certitude que  $a \equiv b \pmod{36}$ . Cette propriété se déduit du fait que si  $a - b$  est un multiple commun de  $m$  et  $n$ , c’est aussi un multiple de  $\text{ppcm}(m, n)$  ; c’est une conséquence du principe d’unique décomposition en facteurs premiers de tout entier.

Il y a un cas particulier très important pour cette dernière règle : si  $m \perp n$ , alors  $\text{ppcm}(m, n) = mn$  et on peut écrire

$$\begin{aligned} a &\equiv b \pmod{mn} \\ \iff a &\equiv b \pmod{m} \text{ et } a \equiv b \pmod{n}, \quad \text{si } m \perp n. \end{aligned} \quad (4.42)$$

Par exemple,  $a \equiv b \pmod{100}$  si et seulement si  $a \equiv b \pmod{25}$  et  $a \equiv b \pmod{4}$ . En d'autres termes, il suffit de connaître  $x \pmod{25}$  et  $x \pmod{4}$  pour déterminer la valeur de  $x \pmod{100}$ . C'est un cas particulier du *Théorème des Restes Chinois* (voir l'exercice 30), ainsi nommé car il est dû à un Chinois, Sun Tzu, qui l'a découvert vers l'an 350.

On déduit aisément de (4.42) que si la décomposition de  $m$  en facteurs premiers (4.11) est  $\prod_p p^{m_p}$ , alors

$$a \equiv b \pmod{m} \iff a \equiv b \pmod{p^{m_p}} \text{ pour tout } p,$$

Ainsi, toute congruence modulo un nombre entier peut s'exprimer avec des congruences modulo des nombres premiers.

## 4.7 RÉSIDUS INDÉPENDANTS

On peut construire avec les congruences un outil important : un *système de numération par résidus*. Dans ce système, chaque entier  $x$  est représenté par une séquence de résidus (ou restes) selon des modules premier entre eux :

$$\begin{aligned} \text{Res}(x) &= (x \pmod{m_1}, \dots, x \pmod{m_r}), \\ \text{avec } m_j &\perp m_k \text{ pour } 1 \leq j < k \leq r. \end{aligned}$$

Si on connaît  $x \pmod{m_1}, \dots, x \pmod{m_r}$ , on ne peut pas en déduire  $x$  ; on peut cependant déterminer  $x \pmod{m}$ , où  $m$  est le produit  $m_1 \dots m_r$ . Si on connaît un intervalle qui contient  $x$ , ce qui arrive souvent en pratique, et si  $m$  est assez grand, on pourra alors trouver exactement  $x$ .

Regardons par exemple un petit système de numération avec seulement deux modules, 3 et 5 :

$x \pmod{15}$	$x \pmod{3}$	$x \pmod{5}$
0	0	0
1	1	1
2	2	2
3	0	3
4	1	4
5	2	0
6	0	1

$x \bmod 15$	$x \bmod 3$	$x \bmod 5$
7	1	2
8	2	3
9	0	4
10	1	0
11	2	1
12	0	2
13	1	3
14	2	4

Tous les couples  $(x \bmod 3, x \bmod 5)$  sont distincts, car  $x \bmod 3 = y \bmod 3$  et  $x \bmod 5 = y \bmod 5$  si et seulement si  $x \bmod 15 = y \bmod 15$ .

Grâce aux règles qui régissent les congruences, on peut effectuer des additions, des soustractions et des multiplications *indépendamment* sur les deux composantes. Par exemple, pour multiplier  $7 = (1, 2)$  par  $13 = (1, 3)$  modulo 15, on calcule  $1 \cdot 1 \bmod 3 = 1$  et  $2 \cdot 3 \bmod 5 = 1$ , et on obtient  $(1, 1) = 1$ . Donc,  $7 \cdot 13 \bmod 15$  doit être égal à 1, ce qu'on peut vérifier facilement.

Cette indépendance des composantes est bien utile pour effectuer ces calculs par ordinateur : ils peuvent être exécutés en parallèle par plusieurs machines ou plusieurs processeurs. Supposons que chaque module  $m_k$  est un nombre premier  $p_k$ , distinct des autres modules, et proche de  $2^{31}$ , tout en étant plus petit. Dans ce cas, tout ordinateur capable d'effectuer les opérations arithmétiques de base sur les entiers de l'intervalle  $[-2^{31}..2^{31}]$  peut sans problème additionner, soustraire et multiplier modulo  $p_k$ . Si on prend un ensemble de  $r$  nombres premiers  $p_k$ , on peut additionner, soustraire et multiplier des nombres en multiple précision comprenant jusqu'à  $31r$  bits ; et cela plus rapidement que par n'importe quelle autre méthode.

Dans certains cas, on peut même faire des divisions. Supposons par exemple qu'on veuille calculer le déterminant  $D$  (entier) d'une matrice à coefficients entiers. On peut déduire des coefficients de la matrice un intervalle dans lequel se trouvera  $|D|$ . Hélas, pour calculer précisément et efficacement  $D$ , on ne peut éviter d'effectuer des divisions ; celles-ci donneront des fractions (et donc une perte de précision si on recourt aux approximations binaires). Pour remédier à cela, on peut calculer  $D \bmod p_k = D_k$ , en prenant des nombres premiers  $p_k$  suffisamment grands. On peut alors sans problème diviser modulo  $p_k$ , sauf si, par malchance, le diviseur est un multiple de  $p_k$ . Si cela arrive, il suffit de choisir un autre nombre premier. En calculant ainsi  $D_k$  pour un nombre suffisant de nombres  $p_k$ , on obtiendra assez d'informations pour en déduire  $D$ .

Nous n'avons pas encore expliqué comment retrouver  $x \bmod m$  à partir d'une suite de résidus  $(x \bmod m_1, \dots, x \bmod m_r)$ . Nous avons vu que

Le nombre de Mersenne

$2^{31} - 1$ ,  
par exemple,  
convient tout à fait.

c'est faisable en principe ; cependant, il se pourrait que les calculs soient si complexes qu'ils ne puissent pas être effectués en pratique. Il existe heureusement une façon simple de procéder. Illustrons la sur l'exemple  $(x \bmod 3, x \bmod 5)$  de notre tableau. L'idée consiste à résoudre le problème pour les deux cas  $(1, 0)$  et  $(0, 1)$  ; alors, si  $(1, 0) = a$  et  $(0, 1) = b$ , on a  $(x, y) = (ax + by) \bmod 15$ .

Dans notre cas, le tableau nous indique que  $a = 10$  et  $b = 6$ . Cependant, si les modules sont très grands, il nous faut une meilleure méthode que celle consistant à regarder le tableau. En d'autres termes, si  $m \perp n$ , il nous faut un moyen efficace de trouver des nombres  $a$  et  $b$  satisfaisant les équations

$$a \bmod m = 1, \quad a \bmod n = 0, \quad b \bmod m = 0, \quad b \bmod n = 1.$$

C'est encore une fois la propriété (4.5) qui vient à notre secours : l'algorithme d'Euclide nous donne deux nombres  $m'$  et  $n'$  tels que

$$m'm + n'n = 1.$$

On peut donc prendre  $a = n'n$  et  $b = m'm$ , ou, si on préfère, leur réduction modulo  $mn$ .

Pour minimiser la longueur des calculs lorsque les modules sont grands, on utilise d'autres astuces, dont on pourra trouver les détails dans [208, page 274]. Notons que la procédure de conversion des résidus pour retrouver les nombres originaux prend du temps. C'est pourquoi, en pratique, on ne gagne du temps que si on peut effectuer tout le calcul avec les résidus pour n'effectuer qu'une conversion tout à la fin.

Utilisons ce que nous venons d'apprendre pour résoudre un petit problème : combien y a-t-il de solutions à la congruence

$$x^2 \equiv 1 \pmod{m}, \tag{4.43}$$

si on considère que deux solutions  $x$  et  $x'$  sont identiques lorsque  $x \equiv x'$  ?

Suivant les principes généraux que nous avons vus, considérons d'abord le cas où  $m$  est égal à  $p^k$ , une puissance d'un nombre premier. Dans ce cas, on peut réécrire la congruence en

$$(x - 1)(x + 1) \equiv 0 \pmod{p^k}.$$

Ainsi,  $p$  doit diviser  $x - 1$ , ou  $x + 1$ , ou les deux ; mais il ne peut diviser les deux que s'il est égal à 2 ; laissons ce cas pour plus tard. Si  $p > 2$ , alors  $p^k \nmid (x - 1)(x + 1) \iff p^k \nmid (x - 1)$  ou  $p^k \nmid (x + 1)$  ; il y a donc exactement deux solutions,  $x \equiv +1$  et  $x \equiv -1$ .

Le cas  $p = 2$  est un peu différent. Si  $2^k \nmid (x - 1)(x + 1)$ , alors l'un des deux nombres  $x - 1$  ou  $x + 1$  est divisible par 2 mais pas par 4, et donc l'autre

est forcément divisible par  $2^{k-1}$ . Si  $k \geq 3$ , il y a donc quatre solutions, qui sont  $x \equiv \pm 1$  et  $x \equiv 2^{k-1} \pm 1$ . Par exemple, si  $p^k = 8$ , les solutions sont  $x \equiv 1, 3, 5, 7 \pmod{8}$ ; retenons donc que *le carré d'un entier impair est toujours de la forme  $8n + 1$* ; cela s'avère souvent utile.

Revenons maintenant au cas général :  $x^2 \equiv 1 \pmod{m}$  si et seulement si  $x^2 \equiv 1 \pmod{p^{m_p}}$  pour tout nombre premier  $p$  tel que  $m_p > 0$  dans la factorisation de  $m$ . Chaque nombre premier est indépendant des autres et, sauf si  $p = 2$ , il y a exactement deux solutions pour  $x \pmod{p^{m_p}}$ . Donc, si  $m$  a exactement  $r$  diviseurs premiers distincts, le nombre total de solutions à  $x^2 \equiv 1$  est égal à  $2^r$ , excepté si  $m$  est pair ; dans ce dernier cas, il y a une petite correction à faire. Voici le nombre exact de solutions pour tous les cas :

$$2^{r+[8|m|+[4|m|-[2|m|]}. \quad (4.44)$$

*Ne faisons pas d'impair : 2 n'est pas impair, tous les autres nombres premiers le sont.*

Par exemple, il existe quatre “racines carrées de 1 modulo 12” : les nombres 1, 5, 7 et 11. Si  $m = 15$ , les quatre solutions sont les nombres dont les résidus mod 3 et mod 5 sont  $\pm 1$  : (1, 1), (1, 4), (2, 1), et (2, 4) dans le système de numération par résidus. En système de numération habituel (décimal), ces solutions s'écrivent 1, 4, 11 et 14.

## 4.8 AUTRES APPLICATIONS.

Il y a une chose qu'il nous reste à faire depuis le chapitre 3 : prouver que les  $m$  nombres

$$0 \pmod{m}, \quad n \pmod{m}, \quad 2n \pmod{m}, \quad \dots, \quad (m-1)n \pmod{m} \quad (4.45)$$

sont exactement  $d$  copies des  $m/d$  nombres

$$0, \quad d, \quad 2d, \quad \dots, \quad m-d$$

avec  $d = \text{pgcd}(m, n)$ . Par exemple, si  $m = 12$  et  $n = 8$  on a  $d = 4$ , et on obtient les nombres 0, 8, 4, 0, 8, 4, 0, 8, 4, 0, 8, 4.

La première partie de la preuve — montrer qu'on obtient  $d$  copies des  $m/d$  premières valeurs — est maintenant triviale. D'après (4.38), on a

$$jn \equiv kn \pmod{m} \iff j(n/d) \equiv k(n/d) \pmod{m/d};$$

on se retrouve donc avec  $d$  copies des valeurs obtenues pour  $0 \leq k < m/d$ .

Il nous faut maintenant montrer que l'ensemble de ces  $m/d$  nombres est égal à  $\{0, d, 2d, \dots, m-d\}$ . Posons  $m = m'd$  et  $n = n'd$ . Alors, selon la règle de distributivité (3.23),  $kn \pmod{m} = d(kn' \pmod{m'})$ . Les valeurs obtenues pour  $0 \leq k < m'$  sont donc égales à  $d$  fois les nombres

$$0 \pmod{m'}, \quad n' \pmod{m'}, \quad 2n' \pmod{m'}, \quad \dots, \quad (m'-1)n' \pmod{m'}.$$

*Les mathématiciens aiment à dire que les choses sont triviales.*

Or, nous savons, d'après (4.27), que  $m' \perp n'$ , puisque ces deux nombres ont été obtenus en divisant  $m$  et  $n$  par leur pgcd. Il nous suffit donc de considérer le cas  $d = 1$ , c'est-à-dire le cas où  $m$  et  $n$  sont premiers entre eux.

Supposons donc que  $m \perp n$ . Dans ce cas, il est facile de voir que les nombres de (4.45) sont exactement ceux de l'ensemble  $\{0, 1, \dots, m - 1\}$ ; on utilise pour cela le "principe des tiroirs", qui dit que, si on range  $m$  objets dans  $m$  tiroirs, alors il y a un tiroir vide si et seulement s'il y a un tiroir contenant au moins deux objets (le principe des boîtes de Dirichlet, démontré dans l'exercice 3.8, est similaire à celui-ci). Nous savons que les nombres de (4.45) sont distincts car, d'après (4.37),

$$jn \equiv kn \pmod{m} \iff j \equiv k \pmod{m}$$

lorsque  $m \perp n$ . Par conséquent, nous avons  $m$  nombres distincts qui doivent remplir toutes les cases  $0, 1, \dots, m - 1$ . Le problème du chapitre 3 est donc clos.

Cette preuve est bien sûr complète ; mais, avec une méthode directe au lieu du principe des tiroirs, on peut prouver plus encore. Si  $m \perp n$  et si on se donne  $j \in [0..m]$ , on peut calculer  $k \in [0..m]$  tel que  $kn \pmod{m} = j$ . Pour cela, on résout la congruence

$$kn \equiv j \pmod{m}$$

en multipliant les deux membres par  $n'$  tel que  $m'm + n'n = 1$  pour obtenir

$$k \equiv jn' \pmod{m};$$

donc  $k = jn' \pmod{m}$ .

Avec ce que nous venons de prouver, nous pouvons établir un résultat important, découvert par Pierre de Fermat en 1640. Fermat était un grand mathématicien qui participa à la découverte de bien des domaines des mathématiques, dont le calcul infinitésimal. Il nous a laissé des carnets contenant des dizaines de théorèmes énoncés sans preuve. Chacun d'eux a pu être démontré par la suite, y compris le "Dernier Théorème de Fermat", devenu fameux car il a tenu tête aux meilleurs mathématiciens pendant 350 ans. Ce théorème dit que

$$a^n + b^n \neq c^n \tag{4.46}$$

pour tous entiers  $a, b, c$  et  $n$ , avec  $n > 2$  (il existe évidemment une infinité de solutions à  $a + b = c$  et  $a^2 + b^2 = c^2$ ). C'est Andrew Wiles qui a finalement résolu le problème. Sa preuve, extrêmement complexe, de (4.46), est parue dans *Annals of Mathematics* 142 (1995), 443–551.

#### FLASH INFO

Euler [115] conjecturait que  $a^4 + b^4 + c^4 \neq d^4$ , mais, en août 1987, Noam Elkies [92] a montré qu'il existe une infinité de solutions.

Depuis, Roger Frye a montré, après un calcul de 110 heures sur une Connection Machine, que la seule solution telle que  $d < 1000000$  est la suivante :

$$\begin{aligned} 95800^4 + 217519^4 \\ + 414560^4 \\ = 422481^4. \end{aligned}$$

Le théorème de Fermat de 1640 est bien plus facile à vérifier. On l'appelle le “petit théorème de Fermat” (ou tout simplement le “théorème de Fermat”). Il dit que

$$n^{p-1} \equiv 1 \pmod{p}, \quad \text{si } n \perp p. \quad (4.47)$$

Preuve : nous supposons comme d'habitude que  $p$  est un nombre premier. Nous savons que les  $p-1$  nombres  $n \pmod{p}$ ,  $2n \pmod{p}, \dots, (p-1)n \pmod{p}$  sont en fait les nombres  $1, 2, \dots, p-1$ . Par conséquent, si on effectue leur produit, on obtient

$$\begin{aligned} & n \cdot (2n) \cdot \dots \cdot ((p-1)n) \\ & \equiv (n \pmod{p}) \cdot (2n \pmod{p}) \cdot \dots \cdot ((p-1)n \pmod{p}) \\ & \equiv (p-1)!, \end{aligned}$$

les calculs étant effectués modulo  $p$ . Cela signifie que

$$(p-1)! n^{p-1} \equiv (p-1)! \pmod{p},$$

et on peut supprimer le facteur  $(p-1)!$  car il n'est pas divisible par  $p$ . CQFD.

Le petit théorème de Fermat peut être énoncé sous une autre forme, parfois plus pratique :

$$n^p \equiv n \pmod{p}, \quad n \text{ entier.} \quad (4.48)$$

Cette congruence est valable pour tout entier  $n$ . La preuve en est facile : si  $n \perp p$ , on multiplie simplement (4.47) par  $n$  ; sinon,  $p \mid n$ , donc  $n^p \equiv 0 \equiv n$ .

La même année où il découvrit (4.47), Fermat écrivit une lettre à Mersenne, dans laquelle il conjecturait que le nombre

$$f_n = 2^{2^n} + 1$$

“...laquelle proposition, si elle est vraie, est de très grand usage.”

était premier pour tout  $n \geq 0$ . Il savait que c'était vrai pour les cinq premières valeurs :

$$2^1 + 1 = 3; 2^2 + 1 = 5; 2^4 + 1 = 17; 2^8 + 1 = 257; 2^{16} + 1 = 65537;$$

mais ne voyait pas comment prouver que la suivante,  $2^{32} + 1 = 4294967297$ , était première.

Il est intéressant de remarquer que Fermat aurait pu, en utilisant le théorème qu'il avait récemment établi, prouver que  $2^{32} + 1$  n'est pas premier. Il lui aurait fallu pour cela faire quelques dizaines de multiplications : en prenant  $n = 3$  dans (4.47), on en déduit que

$$3^{2^{32}} \equiv 1 \pmod{2^{32} + 1}, \quad \text{si } 2^{32} + 1 \text{ est premier ;}$$

—P. de Fermat [121]

et il est possible de tester cette relation à la main, en commençant par 3 puis en élevant au carré 32 fois, ne conservant que les restes mod  $2^{32} + 1$ . On calcule donc  $3^2 = 9$ , puis  $3^{2^2} = 81$ , puis  $3^{2^3} = 6561$ , et ainsi de suite pour arriver à

$$3^{2^{32}} \equiv 3029026160 \pmod{2^{32} + 1}.$$

*Si celui-ci est le petit théorème de Fermat, l'autre est le dernier mais pas le moindre.*

Comme on n'obtient pas 1, le nombre  $2^{32} + 1$  n'est pas premier. Nous n'avons ainsi aucune information sur les facteurs de  $2^{32} + 1$ , mais nous savons qu'il en existe (ce sont les nombres 641 et 6700417, découverts par Euler en 1732 [102]).

Si  $3^{2^{32}}$  avait été congru à 1 modulo  $2^{32} + 1$ , cela n'aurait pas prouvé que  $2^{32} + 1$  est premier ; cela n'aurait rien prouvé du tout en fait. En revanche, l'exercice 47 concerne un résultat permettant de prouver que de grands nombres sont premiers, sans faire d'énormes calculs.

Pour prouver le petit théorème de Fermat, nous avons supprimé  $(p-1)!$  des deux côtés d'une congruence. Il se trouve que  $(p-1)!$  est toujours congru à  $-1$ , modulo  $p$  ; cette propriété est établie dans le théorème de Wilson :

$$(n-1)! \equiv -1 \pmod{n} \iff n \text{ est premier, } \text{ si } n > 1. \quad (4.49)$$

Il est très facile de démontrer une moitié de ce théorème : si  $n > 1$  n'est pas premier, alors il existe un nombre premier  $p$  qui le divise et qui est un facteur de  $(n-1)!$ , donc  $(n-1)!$  ne peut pas être congru à  $-1$  (si  $(n-1)!$  était congru à  $-1$  modulo  $n$ , il serait aussi congru à  $-1$  modulo  $p$ , mais il ne l'est pas).

Selon l'autre moitié du théorème de Wilson,  $(p-1)! \equiv -1 \pmod{p}$ . Ceci peut être démontré en appariant chaque nombre avec son inverse modulo  $p$ . Si  $n \perp p$ , nous savons qu'il existe  $n'$  tel que

$$n'n \equiv 1 \pmod{p};$$

le nombre  $n'$  est l'inverse de  $n$ , et  $n$  est aussi l'inverse de  $n'$ . Les inverses de  $n$  sont forcément congrus deux à deux, car si  $nn' \equiv nn''$ , alors  $n' \equiv n''$ .

*Entre deux nombres premiers, lequel est le premier ?*

Supposons maintenant qu'on apparie chaque nombre entre 1 et  $p-1$  avec son inverse. Comme le produit d'un nombre et de son inverse est congru à 1, le produit de tous les nombres de toutes les paires est aussi congru à 1 ; on peut donc raisonnablement penser que  $(p-1)!$  est congru à 1. Vérifions ceci pour  $p = 5$  par exemple. On obtient  $4! = 24$  ; mais ce nombre est congru à 4, pas à 1, modulo 5. Aïe ! Qu'est-ce qui cloche ? Regardons les inverses d'un peu plus près :

$$1' = 1, \quad 2' = 3, \quad 3' = 2, \quad 4' = 4.$$

Ah Ah ! Les nombres 2 et 3 sont appariés, mais pas les autres, car chacun des nombres 1 et 4 est son propre inverse.

Pour corriger notre première analyse, nous devons déterminer quels sont les nombres qui sont inverses d'eux-mêmes. Si  $x$  est son propre inverse, alors  $x^2 \equiv 1 \pmod{p}$  ; et nous avons déjà prouvé que cette congruence a exactement deux racines lorsque  $p > 2$  (si  $p = 2$ , il n'y a aucun problème car il est évident que  $(p-1)! \equiv -1$ ). Les racines sont 1 et  $p-1$ , et tous les autres nombres (entre 1 et  $p-1$ ) sont appariés ; donc nous trouvons bien

$$(p-1)! \equiv 1 \cdot (p-1) \equiv -1.$$

Il n'est malheureusement pas possible de calculer efficacement les factorielles. Le théorème de Wilson n'est donc, en pratique, d'aucune utilité pour tester la primalité. C'est juste un théorème.

## 4.9 PHI ET MU

Combien d'entiers, parmi  $\{0, 1, \dots, m-1\}$ , sont premiers par rapport à  $m$ ? Le nombre qui répond à cette question, noté  $\varphi(m)$ , est appelé le "totient" de  $m$  (ainsi nommé par J. J. Sylvester [347], un mathématicien britannique qui aimait inventer des mots nouveaux). Ainsi,  $\varphi(1) = 1$ ,  $\varphi(p) = p-1$  et  $\varphi(m) < m-1$  pour tout nombre composite  $m$ .

La fonction  $\varphi$  est appelée *fonction d'Euler*, car Euler a été le premier à l'étudier. Il découvrit, par exemple, que le petit théorème de Fermat peut être généralisé aux modules non premiers :

$$n^{\varphi(m)} \equiv 1 \pmod{m}, \quad \text{si } n \perp m. \quad (4.50)$$

(L'exercice 32 consiste à démontrer le théorème d'Euler).

Si  $m$  est une puissance d'un nombre premier  $p^k$ , on peut facilement calculer  $\varphi(m)$ , car  $n \perp p^k \iff p \nmid n$ . Les multiples de  $p$  qui appartiennent à  $\{0, 1, \dots, p^k-1\}$  sont  $\{0, p, 2p, \dots, p^k-p\}$  ; il y en a donc  $p^{k-1}$ , et tous les autres sont comptés par  $\varphi(p^k)$  :

$$\varphi(p^k) = p^k - p^{k-1}.$$

Notons que cette formule donne bien  $\varphi(p) = p-1$  pour  $k=1$ .

Si  $m > 1$  n'est pas une puissance d'un nombre premier, on peut poser  $m = m_1 m_2$ , avec  $m_1 \perp m_2$ . Alors les nombres  $0 \leq n < m$  peuvent être représentés dans un système de numération par résidus par le couple  $(n \bmod m_1, n \bmod m_2)$ . D'après (4.30) et (4.4), on a

$$n \perp m \iff n \bmod m_1 \perp m_1 \text{ et } n \bmod m_2 \perp m_2.$$

"Si fuerit N ad x  
numerus primus et  
n numerus partium  
ad N primarum,  
tum potestas x<sup>n</sup>  
unitate minuta sem-  
per per numerum N  
erit divisibilis."

—L. Euler [111]

Par conséquent,  $n \bmod m$  est "bon" si et seulement si  $n \bmod m_1$  et  $n \bmod m_2$  sont "bons" tous les deux, si on considère la primalité relative comme une vertu. On peut maintenant calculer récursivement le nombre de bonnes valeurs modulo  $m$  : c'est  $\varphi(m_1)\varphi(m_2)$ , car il y a  $\varphi(m_1)$  bonnes façons de choisir la première composante  $n \bmod m_1$  et  $\varphi(m_2)$  bonnes façons de choisir la seconde composante  $n \bmod m_2$  dans la représentation par résidus.

Par exemple,  $\varphi(12) = \varphi(4)\varphi(3) = 2 \cdot 2 = 4$ , car  $n$  est premier à 12 si et seulement si  $n \bmod 4 = (1 \text{ ou } 3)$  et  $n \bmod 3 = (1 \text{ ou } 2)$ . Les quatre nombres premiers par rapport à 12 sont  $(1, 1)$ ,  $(1, 2)$ ,  $(3, 1)$ ,  $(3, 2)$  dans le système de numération par résidus ; ils s'écrivent 1, 5, 7, 11 en notation décimale. Donc, d'après le théorème d'Euler,  $n^4 \equiv 1 \pmod{12}$  si  $n \perp 12$ .

On dit qu'une fonction  $f(m)$  sur des entiers strictement positifs est *multiplicative* si  $f(1) = 1$  et

$$f(m_1 m_2) = f(m_1)f(m_2) \quad \text{si } m_1 \perp m_2. \quad (4.51)$$

Nous venons de prouver que  $\varphi(m)$  est multiplicative. Nous avons aussi vu une autre fonction multiplicative dans ce chapitre : le nombre de solutions non congrues deux à deux de  $x^2 \equiv 1 \pmod{m}$  est une fonction multiplicative. La fonction  $f(m) = m^\alpha$  en est un autre exemple, pour tout  $\alpha$ .

Une fonction multiplicative est totalement définie par ses valeur sur les puissances des nombres premiers. En effet, tout entier strictement positif  $m$  peut être décomposé en facteurs premiers, chacun de ces facteurs étant premier à tous les autres. La formule

$$f(m) = \prod_p f(p^{m_p}), \quad \text{si } m = \prod_p p^{m_p} \quad (4.52)$$

est valable si et seulement si  $f$  est multiplicative.

En particulier, cette formule nous donne la valeur de la fonction d'Euler pour tout  $m$  :

$$\varphi(m) = \prod_{p \nmid m} (p^{m_p} - p^{m_p-1}) = m \prod_{p \nmid m} \left(1 - \frac{1}{p}\right). \quad (4.53)$$

Par exemple,  $\varphi(12) = (4-2)(3-1) = 12(1-\frac{1}{2})(1-\frac{1}{3})$ .

Etudions maintenant une application de la fonction  $\varphi$  à l'étude des nombres rationnels mod 1. On dit que la fraction  $m/n$  est *basique* si  $0 \leq m < n$ . Par conséquent,  $\varphi(n)$  désigne le nombre de fractions basiques réduites ayant  $n$  pour dénominateur. D'autre part, la suite de Farey  $\mathcal{F}_n$  contient toutes les fractions basiques réduites ayant un dénominateur inférieur ou égal à  $n$ , ainsi que la fraction  $\frac{1}{1}$ .

"Si sint A et B numeri inter se primi et numerus partium ad A primarum sit = a, numerus vero partium ad B primarum sit = b, tum numerus partium ad productum AB primarum erit = ab."

— L. Euler [111]

L'ensemble de *toutes* les fractions basiques de dénominateur 12, avant réduction, est

$$\frac{0}{12}, \frac{1}{12}, \frac{2}{12}, \frac{3}{12}, \frac{4}{12}, \frac{5}{12}, \frac{6}{12}, \frac{7}{12}, \frac{8}{12}, \frac{9}{12}, \frac{10}{12}, \frac{11}{12}.$$

Après réduction, cela devient

$$\frac{0}{1}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{5}{12}, \frac{1}{2}, \frac{7}{12}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}, \frac{11}{12},$$

et on peut grouper ces fractions selon leurs dénominateurs :

$$\frac{0}{1}; \quad \frac{1}{2}; \quad \frac{1}{3}, \frac{2}{3}; \quad \frac{1}{4}, \frac{3}{4}; \quad \frac{1}{6}, \frac{5}{6}; \quad \frac{1}{12}, \frac{5}{12}, \frac{7}{12}, \frac{11}{12}.$$

Maintenant, que pouvons-nous faire de cela ? Eh bien, nous voyons que chaque diviseur  $d$  de 12 apparaît comme dénominateur ; les numérateurs correspondants sont les nombres comptés par  $\varphi(d)$  ; et les seuls dénominateurs qui apparaissent sont des diviseurs de 12. Donc

$$\varphi(1) + \varphi(2) + \varphi(3) + \varphi(4) + \varphi(6) + \varphi(12) = 12.$$

On obtient bien sûr un résultat similaire si on part des fractions  $\frac{0}{m}, \frac{1}{m}, \dots, \frac{m-1}{m}$  pour n'importe quel  $m$  donné. Par conséquent,

$$\sum_{d|m} \varphi(d) = m. \tag{4.54}$$

Vers le début de ce chapitre, nous disions que les problèmes de théorie des nombres faisaient souvent apparaître des sommes sur les diviseurs d'un nombre. Nous ne mentionnons pas, car (4.54) est bien une telle somme (et nous en verrons d'autres).

Voici maintenant une curieuse propriété : si  $f$  est une fonction telle que la somme  $g(m) = \sum_{d|m} f(d)$  est multiplicative, alors  $f$  est multiplicative (ce résultat, combiné avec (4.54) et le fait que  $g(m) = m$  est trivialement multiplicative, donne une autre raison pour laquelle  $\varphi(m)$  est multiplicative). Cette propriété peut être prouvée par induction sur  $m$ : la base est facile car  $f(1) = g(1) = 1$ . Soit maintenant  $m > 1$ , et supposons que, si  $m_1 \perp m_2$  et  $m_1 m_2 < m$ , alors  $f(m_1 m_2) = f(m_1)f(m_2)$ . Si  $m = m_1 m_2$  et  $m_1 \perp m_2$ , alors on a

$$g(m_1 m_2) = \sum_{d|m_1 m_2} f(d) = \sum_{d_1|m_1} \sum_{d_2|m_2} f(d_1 d_2),$$

et  $d_1 \perp d_2$  puisque tous les diviseurs de  $m_1$  sont premiers par rapport à tous les diviseurs de  $m_2$ . D'après l'hypothèse d'induction,  $f(d_1 d_2) = f(d_1)f(d_2)$ ,

sauf peut être si  $d_1 = m_1$  et  $d_2 = m_2$ . On obtient donc

$$\begin{aligned} & \left( \sum_{d_1 \mid m_1} f(d_1) \sum_{d_2 \mid m_2} f(d_2) \right) - f(m_1)f(m_2) + f(m_1m_2) \\ &= g(m_1)g(m_2) - f(m_1)f(m_2) + f(m_1m_2). \end{aligned}$$

C'est égal à  $g(m_1m_2) = g(m_1)g(m_2)$ , donc  $f(m_1m_2) = f(m_1)f(m_2)$ .

Réiproquement, si  $f(m)$  est multiplicative, alors la fonction  $g(m) = \sum_{d \mid m} f(d)$  est aussi multiplicative. L'exercice 33 montre même plus encore.

La *fonction de Möbius*  $\mu(m)$ , qui doit son nom au mathématicien du dix-neuvième siècle August Möbius, connu aussi par son fameux ruban, peut être définie pour tout entier  $m \geq 1$  par l'équation

$$\sum_{d \mid m} \mu(d) = [m=1]. \quad (4.55)$$

Cette équation est en réalité une récurrence, car le membre gauche est une somme contenant  $\mu(m)$  et certaines valeurs de  $\mu(d)$  avec  $d < m$ . Par exemple, en posant successivement  $m = 1, 2, \dots, 12$ , on peut calculer ses douze premières valeurs :

$m$	1	2	3	4	5	6	7	8	9	10	11	12
$\mu(m)$	1	-1	-1	0	-1	1	-1	0	0	1	-1	0

Richard Dedekind [77] et Joseph Liouville [251] ont découvert en 1857 l'important principe d'inversion que voici :

$$g(m) = \sum_{d \mid m} f(d) \iff f(m) = \sum_{d \mid m} \mu(d)g\left(\frac{m}{d}\right). \quad (4.56)$$

Selon ce principe, la fonction  $\mu$  nous fournit une nouvelle façon de se représenter une fonction  $f(m)$  pour laquelle on connaît  $\sum_{d \mid m} f(d)$ .

La preuve de (4.56) fait appel aux deux règles (4.7) et (4.9) que nous avons vues vers le début du chapitre : si  $g(m) = \sum_{d \mid m} f(d)$ , alors

*C'est le moment de tenter l'exercice d'échauffement 11.*

$$\begin{aligned} \sum_{d \mid m} \mu(d)g\left(\frac{m}{d}\right) &= \sum_{d \mid m} \mu\left(\frac{m}{d}\right)g(d) = \sum_{d \mid m} \mu\left(\frac{m}{d}\right) \sum_{k \mid d} f(k) \\ &= \sum_{k \mid m} \sum_{d \mid (m/k)} \mu\left(\frac{m}{kd}\right) f(k) = \sum_{k \mid m} \sum_{d \mid (m/k)} \mu(d) f(k) \\ &= \sum_{k \mid m} [m/k=1] f(k) = f(m). \end{aligned}$$

La deuxième partie de l'expression (4.56) se démontre de manière similaire (voir l'exercice 12).

La relation (4.56) décrit une propriété de la fonction de Möbius qui est bien utile. Nous en avons calculé les douze premières valeurs ; mais que vaut  $\mu(m)$  lorsque  $m$  est grand ? Comment résoudre la récurrence (4.55) ? La fonction  $g(m) = [m=1]$  est évidemment multiplicative ; donc la fonction de Möbius définie par (4.55) doit aussi être multiplicative, d'après la curieuse propriété que nous avons démontrée il y a une ou deux minutes. Nous pouvons donc trouver  $\mu(m)$  si nous connaissons les  $\mu(p^k)$ .

Si  $m = p^k$ , on a, d'après (4.55),

$$\mu(1) + \mu(p) + \mu(p^2) + \cdots + \mu(p^k) = 0$$

pour tout  $k \geq 1$ , du fait que les diviseurs de  $p^k$  sont  $1, \dots, p^k$ . Il s'ensuit que

$$\mu(p) = -1; \quad \mu(p^k) = 0 \quad \text{pour } k > 1.$$

Par conséquent, d'après (4.52), on obtient la formule

$$\mu(m) = \prod_{p|m} \mu(p^{m_p}) = \begin{cases} (-1)^r, & \text{si } m = p_1 p_2 \dots p_r; \\ 0, & \text{s'il existe un } p^2 \text{ divisant } m. \end{cases} \quad (4.57)$$

Si on considère (4.54) comme une récurrence pour la fonction  $\varphi(m)$ , on peut la résoudre en utilisant la règle de Dedekind-Liouville (4.56). On obtient ainsi

$$\varphi(m) = \sum_{d|m} \mu(d) \frac{m}{d}. \quad (4.58)$$

Par exemple,

$$\begin{aligned} \varphi(12) &= \mu(1) \cdot 12 + \mu(2) \cdot 6 + \mu(3) \cdot 4 + \mu(4) \cdot 3 + \mu(6) \cdot 2 + \mu(12) \cdot 1 \\ &= 12 - 6 - 4 + 0 + 2 + 0 = 4. \end{aligned}$$

Si  $m$  est divisible par  $r$  nombres premiers différents  $\{p_1, \dots, p_r\}$ , alors la somme (4.58) ne contient que  $2^r$  termes non nuls, car la fonction  $\mu$  s'annule souvent. En ce sens, (4.58) permet d'éclairer davantage la formule (4.53), qui peut s'exprimer ainsi :

$$\varphi(m) = m \left(1 - \frac{1}{p_1}\right) \dots \left(1 - \frac{1}{p_r}\right).$$

Si on développe cette expression, on obtient exactement les  $2^r$  termes non nuls de (4.58). La fonction de Möbius a des applications dans beaucoup d'autres situations.

Par exemple, essayons de trouver le nombre de fractions qui constituent la suite de Farey  $\mathcal{F}_n$ . C'est le nombre de fractions réduites de l'intervalle  $[0..1]$  dont les dénominateurs ne dépassent pas  $n$  ; c'est donc 1 de plus que  $\Phi(n)$ , où la fonction  $\Phi$  est ainsi définie :

$$\Phi(x) = \sum_{1 \leq k \leq x} \varphi(k). \quad (4.59)$$

(On ajoute 1 à  $\Phi(n)$  pour tenir compte de la dernière fraction  $\frac{1}{1}$ ). La somme de (4.59) n'est pas particulièrement sympathique. On sait quand même déterminer  $\Phi(x)$ , d'une autre manière, en remarquant que

$$\sum_{d \geq 1} \Phi\left(\frac{x}{d}\right) = \frac{1}{2} \lfloor x \rfloor [1+x] \quad (4.60)$$

pour tout réel  $x \geq 0$ . En effet, il existe exactement  $\frac{1}{2} \lfloor x \rfloor [1+x]$  fractions  $m/n$  (réduites ou non) telles que  $0 \leq m < n \leq x$ , ce qui nous donne le membre droit de l'égalité. Le nombre de ces fractions qui satisfont  $\text{pgcd}(m, n) = d$  est égal à  $\Phi(x/d)$ , car chacune d'elle peut aussi s'écrire  $m'/n'$  avec  $0 \leq m' < n' \leq x/d$ , après avoir remplacé  $m$  par  $m'd$  et  $n$  par  $n'd$ . Ainsi, les deux membres comptent les mêmes objets.

Pour bien comprendre les équations (4.59) et (4.60), voyons tout cela de plus près. On déduit de la définition de  $\Phi(x)$  que  $\Phi(x) = \Phi(\lfloor x \rfloor)$  ; il s'avère cependant plus pratique de définir  $\Phi(x)$  pour tous les réels au lieu de se restreindre aux entiers. Voici la table des premières valeurs de  $\Phi$  et  $\varphi$  pour des arguments entiers :

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12
$\varphi(n)$	—	1	1	2	2	4	2	6	4	6	4	10	4
$\Phi(n)$	0	1	2	4	6	10	12	18	22	28	32	42	46

Vérifions (4.60) pour  $x = 12$  :

$$\begin{aligned} \Phi(12) + \Phi(6) + \Phi(4) + \Phi(3) + \Phi(2) + \Phi(1) \\ = 46 + 12 + 6 + 4 + 2 + 6 = 78 = \frac{1}{2} \cdot 12 \cdot 13. \end{aligned}$$

Impeccable.

L'identité (4.60) peut être vue comme une récurrence implicite pour  $\Phi(x)$ . En effet, on peut l'utiliser, comme nous venons de le voir, pour calculer par exemple  $\Phi(12)$  en fonction de  $\Phi(6)$ ,  $\Phi(4)$ ,  $\Phi(3)$ ,  $\Phi(2)$  et  $\Phi(1)$ . On sait résoudre ce genre de récurrence grâce à une autre belle propriété de la fonction de Möbius :

$$g(x) = \sum_{d \geq 1} f(x/d) \quad \Longleftrightarrow \quad f(x) = \sum_{d \geq 1} \mu(d) g(x/d). \quad (4.61)$$

*(Ce procédé d'extension aux valeurs réelles est très utile pour résoudre beaucoup de récurrences qui apparaissent en analyse d'algorithme).*

*En fait, Möbius [273] a introduit cette fonction pour (4.61), non pour (4.56).*

Toute fonction  $f$  telle que  $\sum_{k,d \geq 1} |f(x/kd)| < \infty$  satisfait cette règle d'inversion. Voici comment la démontrer : supposons que  $g(x) = \sum_{d \geq 1} f(x/d)$  ; alors

$$\begin{aligned}\sum_{d \geq 1} \mu(d) g(x/d) &= \sum_{d \geq 1} \mu(d) \sum_{k \geq 1} f(x/kd) \\&= \sum_{m \geq 1} f(x/m) \sum_{d,k \geq 1} \mu(d)[m = kd] \\&= \sum_{m \geq 1} f(x/m) \sum_{d \nmid m} \mu(d) \\&= \sum_{m \geq 1} f(x/m)[m = 1] = f(x).\end{aligned}$$

La preuve de l'autre implication est similaire.

Nous pouvons maintenant résoudre la récurrence (4.60) en  $\Phi(x)$  :

$$\Phi(x) = \frac{1}{2} \sum_{d \geq 1} \mu(d) \lfloor x/d \rfloor \lfloor 1 + x/d \rfloor. \quad (4.62)$$

Cette somme est toujours finie. Par exemple,

$$\begin{aligned}\Phi(12) &= \frac{1}{2}(12 \cdot 13 - 6 \cdot 7 - 4 \cdot 5 + 0 - 2 \cdot 3 + 2 \cdot 3 \\&\quad - 1 \cdot 2 + 0 + 0 + 1 \cdot 2 - 1 \cdot 2 + 0) \\&= 78 - 21 - 10 - 3 + 3 - 1 + 1 - 1 = 46.\end{aligned}$$

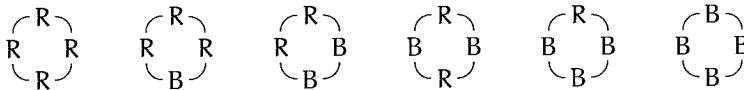
Nous verrons dans le chapitre 9 comment se servir de (4.62) pour obtenir une bonne approximation de  $\Phi(x)$  ; nous prouverons en fait un résultat établi par Mertens en 1874 [270] :

$$\Phi(x) = \frac{3}{\pi^2} x^2 + O(x \log x).$$

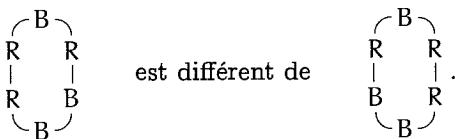
Ainsi, la fonction  $\Phi(x)$  croît "en douceur" ; elle aplaniit un peu le comportement plus brutal de  $\varphi(k)$ .

Pour suivre la tradition commencée au chapitre précédent, terminons ce chapitre-ci en étudiant un problème qui nous permettra d'appliquer beaucoup de ce que nous venons de voir ; ce sera aussi une transition vers le chapitre suivant. Supposons que nous ayons un certain nombre de perles de  $n$  couleurs différentes. Notre but est de compter le nombre de façons différentes de les enfiler pour en faire un collier fermé de longueur  $m$ . Pour commencer, essayons de nommer pour régner en notant  $N(m, n)$  le nombre de colliers différents possibles.

Par exemple, avec des perles de deux couleurs, R et B, il existe  $N(4, 2) = 6$  manières différentes de faire un collier de longueur 4 :



Chacune des autres configurations possibles est équivalente à l'une de celles-ci, car la rotation d'un collier ne le modifie pas. Par contre, nous considérons que deux colliers symétriques peuvent être différents : si  $m = 6$ ,



C'est P. A. MacMahon qui a, en 1892 [264], résolu le premier ce problème de dénombrement.

A première vue, il n'y a pas de récurrence évidente pour  $N(m, n)$ . Essayons de compter les colliers en faisant une petite manipulation : prenons  $m$  copies de chacun d'eux, puis coupons chaque copie en un point différent pour obtenir des suites linéaires de perles. Pour  $m = 4$  et  $n = 2$  on obtient ainsi

RRRR	RRRR	RRRR	RRRR
RRBR	RRRB	BRRR	RBRR
RBBR	RRBB	BRRB	BBRR
RBRB	BRBR	RBRB	BRBR
RBBB	BRBB	BBRB	BBBR
BBBB	BBBB	BBBB	BBBB

Chacune des  $n^m$  configurations possibles apparaît au moins une fois dans ce tableau de  $mN(m, n)$  mots ; certains motifs y sont plusieurs fois. Il n'est pas difficile de savoir combien de fois apparaît un motif donné  $a_0 \dots a_{m-1}$  : c'est le nombre de décalages circulaires  $a_k \dots a_{m-1} a_0 \dots a_{k-1}$  qui produisent le motif de départ  $a_0 \dots a_{m-1}$ . Ainsi, BRBR apparaît deux fois car les quatre façons de couper le collier fermé BRBR sont les quatre décalages circulaires (BRBR, RBRB, BRBR, RBRB) ; deux d'entre eux coïncident avec BRBR. On déduit de cet argument que

$$\begin{aligned} mN(m, n) &= \sum_{a_0, \dots, a_{m-1} \in S_n} \sum_{0 \leq k < m} [a_0 \dots a_{m-1} = a_k \dots a_{m-1} a_0 \dots a_{k-1}] \\ &= \sum_{0 \leq k < m} \sum_{a_0, \dots, a_{m-1} \in S_n} [a_0 \dots a_{m-1} = a_k \dots a_{m-1} a_0 \dots a_{k-1}] \end{aligned}$$

où  $S_n$  est un ensemble de  $n$  couleurs différentes.

Pour un  $k$  donné, voyons maintenant combien de motifs satisfont l'égalité  $a_0 \dots a_{m-1} = a_k \dots a_{m-1} a_0 \dots a_{k-1}$ . Par exemple, si  $m = 12$  et  $k = 8$ , il s'agit de compter le nombre de solutions de

$$a_0 a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} = a_8 a_9 a_{10} a_{11} a_0 a_1 a_2 a_3 a_4 a_5 a_6 a_7.$$

On trouve  $a_0 = a_8 = a_4$ ,  $a_1 = a_9 = a_5$ ,  $a_2 = a_{10} = a_6$  et  $a_3 = a_{11} = a_7$ . Les valeurs de  $a_0$ ,  $a_1$ ,  $a_2$  et  $a_3$  peuvent donc être choisies de  $n^4$  manières différentes, et les valeurs des  $a_k$  autres perles dépendant de celles-ci. Revenons au cas général. Il s'agit de résoudre

$$a_j = a_{(j+k) \bmod m}, \quad \text{pour } 0 \leq j < m,$$

ce qui revient à dire que  $a_j$  est égal à  $a_{(j+kl) \bmod m}$  pour  $l = 1, 2, \dots$ , etc. Nous savons que les multiples de  $k$  modulo  $m$  sont  $\{0, d, 2d, \dots, m-d\}$ , avec  $d = \text{pgcd}(k, m)$ . La solution consiste donc à choisir indépendamment chacun des  $a_0, \dots, a_{d-1}$ , puis poser  $a_j = a_{j-d}$  pour  $d \leq j < m$ . Il y a donc  $n^d$  solutions.

Nous venons de démontrer que

$$mN(m, n) = \sum_{0 \leq k < m} n^{\text{pgcd}(k, m)}.$$

Simplifions cette somme en posant  $d = \text{pgcd}(k, m)$  :

$$\begin{aligned} N(m, n) &= \frac{1}{m} \sum_{d \mid m} n^d \sum_{0 \leq k < m} [d = \text{pgcd}(k, m)] \\ &= \frac{1}{m} \sum_{d \mid m} n^d \sum_{0 \leq k < m} [k/d \perp m/d] \\ &= \frac{1}{m} \sum_{d \mid m} n^d \sum_{0 \leq k < m/d} [k \perp m/d]. \end{aligned}$$

(On peut remplacer  $k/d$  par  $k$  car  $k$  est forcément un multiple de  $d$ ). Par définition,  $\sum_{0 \leq k < m/d} [k \perp m/d] = \varphi(m/d)$  ; on obtient donc finalement la formule de MacMahon :

$$N(m, n) = \frac{1}{m} \sum_{d \mid m} n^d \varphi\left(\frac{m}{d}\right) = \frac{1}{m} \sum_{d \mid m} \varphi(d) n^{m/d}. \quad (4.63)$$

Pour  $m = 4$  et  $n = 2$  par exemple, il y a exactement  $\frac{1}{4}(1 \cdot 2^4 + 1 \cdot 2^2 + 2 \cdot 2^1) = 6$  colliers différents. C'est bien ce que nous avions trouvé.

Si on ne connaît pas son lien avec les colliers, il n'est pas évident à première vue que la valeur de  $N(m, n)$  donnée par la somme de MacMahon

est un entier ! Essayons de démontrer directement, c'est-à-dire sans profiter de ce lien, que

$$\sum_{d|m} \varphi(d) n^{m/d} \equiv 0 \pmod{m}. \quad (4.64)$$

Si  $m$  est premier, cette congruence se réduit à  $n^p + (p-1)n \equiv 0 \pmod{p}$ , ce qui donne  $n^p \equiv n$ . Nous avons vu dans (4.48) que cette dernière congruence est une autre forme du théorème de Fermat. On en déduit que (4.64) est vérifiée lorsque  $m = p$ . Ce résultat peut être vu comme une généralisation du théorème de Fermat au cas où le module n'est pas premier (la généralisation faite par Euler (4.50) est différente).

Nous avons donc démontré que (4.64) est vraie si le module est premier. Voyons maintenant les autres cas, en commençant par la plus petite valeur de possible de  $m$ , qui est 4. Il nous faut prouver que

$$n^4 + n^2 + 2n \equiv 0 \pmod{4}.$$

Si  $n$  est pair, chacun des trois termes de gauche est congru à 0 modulo 4, donc leur somme l'est aussi. Si  $n$  est impair,  $n^4$  et  $n^2$  sont congrus à 1 et  $2n$  est congru à 2 ; par conséquent, le membre gauche est congru à  $1+1+2$  et donc à 0 modulo 4. Voilà qui est fait.

Soyons un peu plus téméraires et tentons  $m = 12$ . Bien que cette valeur soit plutôt petite, elle risque d'être intéressante car elle a beaucoup de diviseurs, dont le carré d'un nombre premier (de plus, nous espérons bien pouvoir généraliser la preuve pour  $m = 12$  à une preuve pour tout  $m$ ). Voici la congruence que nous devons démontrer :

$$n^{12} + n^6 + 2n^4 + 2n^3 + 2n^2 + 4n \equiv 0 \pmod{12}.$$

D'après (4.42), cette congruence est vraie modulo 12 si et seulement si elle est vraie aussi modulo 3 et modulo 4. Commençons donc par la démontrer modulo 3. Comme 3 est premier, et comme nous avons déjà montré que (4.64) est valable pour les modules premiers, nous savons que  $n^3 + 2n \equiv 0 \pmod{3}$ . Profitons de ce fait pour regrouper intelligemment les termes de la somme qui nous intéresse :

$$\begin{aligned} n^{12} + n^6 + 2n^4 + 2n^3 + 2n^2 + 4n \\ &= (n^{12} + 2n^4) + (n^6 + 2n^2) + 2(n^3 + 2n) \\ &\equiv 0 + 0 + 2 \cdot 0 \\ &\equiv 0 \pmod{3}. \end{aligned}$$

Nous avons fait la moitié du travail. Nous pouvons utiliser exactement la même méthode pour la congruence modulo 4. Comme nous avons démontré que  $n^4 + n^2 + 2n \equiv 0 \pmod{4}$ , nous pouvons grouper les termes

ainsi :

$$\begin{aligned} n^{12} + n^6 + 2n^4 + 2n^3 + 2n^2 + 4n \\ = (n^{12} + n^6 + 2n^3) + 2(n^4 + n^2 + 2n) \\ \equiv 0 + 2 \cdot 0 \equiv 0 \pmod{4}. \end{aligned}$$

*CQFD : Ce Qui est Facile à Démontrer.*

CQFD pour le cas  $m = 12$ .

Jusqu'ici, nous avons prouvé notre congruence pour  $m$  premier,  $m = 4$  et  $m = 12$ . Essayons maintenant de la démontrer dans le cas où  $m$  est une puissance d'un nombre premier. Commençons par le cas  $m = p^3$ , où  $p$  est un nombre premier quelconque. Dans ce cas, le membre gauche de (4.64) devient

$$\begin{aligned} n^{p^3} + \varphi(p)n^{p^2} + \varphi(p^2)n^p + \varphi(p^3)n \\ = n^{p^3} + (p-1)n^{p^2} + (p^2-p)n^p + (p^3-p^2)n \\ = (n^{p^3} - n^{p^2}) + p(n^{p^2} - n^p) + p^2(n^p - n) + p^3n. \end{aligned}$$

Pour montrer que c'est congru à 0 modulo  $p^3$ , on peut faire comme suit : prouver que  $n^{p^3} - n^{p^2}$  est divisible par  $p^3$ , que  $n^{p^2} - n^p$  est divisible par  $p^2$ , et que  $n^p - n$  est divisible par  $p$  ; alors le tout sera divisible par  $p^3$ . D'après le théorème de Fermat (sa seconde forme), on a  $n^p \equiv n \pmod{p}$ , donc  $p$  divise  $n^p - n$  ; alors il existe un entier  $q$  tel que

$$n^p = n + pq.$$

Elevons le tout à la puissance  $p$ , puis développons le membre droit en utilisant la formule du binôme (que nous verrons au chapitre 5), enfin regroupons les termes. Nous obtenons ainsi

$$\begin{aligned} n^{p^2} &= (n + pq)^p = n^p + (pq)^1 n^{p-1} \binom{p}{1} + (pq)^2 n^{p-2} \binom{p}{2} + \dots \\ &= n^p + p^2 Q \end{aligned}$$

où  $Q$  est un certain nombre entier. Nous avons pu mettre  $p^2$  en facteur car  $\binom{p}{1} = p$  dans le deuxième terme et, dans tous les termes suivants, un facteur  $(pq)^2$  apparaît. Ainsi nous trouvons que  $p^2$  divise  $n^{p^2} - n^p$ .

Elevons encore une fois à la puissance  $p$ , puis développons et regroupons les termes, afin d'obtenir

$$\begin{aligned} n^{p^3} &= (n^p + p^2 Q)^p \\ &= n^{p^2} + (p^2 Q)^1 n^{p(p-1)} \binom{p}{1} + (p^2 Q)^2 n^{p(p-2)} \binom{p}{2} + \dots \\ &= n^{p^2} + p^3 Q \end{aligned}$$

pour un certain entier  $Q$  (différent du premier). Ainsi,  $p^3$  divise  $n^{p^3} - n^{p^2}$ . Cela termine la preuve pour  $m = p^3$ , car nous avons montré que  $p^3$  divise le membre gauche de (4.64).

Maintenant, on peut prouver par induction que

$$n^{p^k} = n^{p^{k-1}} + p^k Q$$

pour un dernier entier  $Q$  (dernier, car nous sommes à court de fontes). Par conséquent,

$$n^{p^k} \equiv n^{p^{k-1}} \pmod{p^k}, \quad \text{pour } k > 0. \quad (4.65)$$

Donc, le membre gauche de (4.64), qui s'écrit

$$(n^{p^k} - n^{p^{k-1}}) + p(n^{p^{k-1}} - n^{p^{k-2}}) + \cdots + p^{k-1}(n^p - n) + p^k n,$$

est divisible par  $p^k$ ; il est donc congru à 0 modulo  $p^k$ .

Nous y sommes presque. Maintenant que nous savons que (4.64) est vraie pour les puissances de nombres premiers, il ne nous reste plus qu'à prouver qu'elle est vraie lorsque  $m = m_1 m_2$ , avec  $m_1 \perp m_2$ , en supposant qu'elle est vraie pour  $m_1$  et  $m_2$ . Le cas  $m = 12$ , que nous avons déjà résolu, nous incite à être optimistes.

Comme la fonction  $\varphi$  est multiplicativa, on peut écrire

$$\begin{aligned} \sum_{d|m} \varphi(d) n^{m/d} &= \sum_{d_1|m_1, d_2|m_2} \varphi(d_1 d_2) n^{m_1 m_2 / d_1 d_2} \\ &= \sum_{d_1|m_1} \varphi(d_1) \left( \sum_{d_2|m_2} \varphi(d_2) (n^{m_1 / d_1})^{m_2 / d_2} \right). \end{aligned}$$

Or, la somme interne est congrue à 0 modulo  $m_2$ , puisque nous supposons que (4.64) est vraie pour  $m_2$ . Par conséquent, la somme entière est congrue à 0 modulo  $m_2$ . Par un même argument, on montre qu'elle est aussi congrue à 0 modulo  $m_1$ . D'après (4.42), nous pouvons donc conclure qu'elle est congrue à 0 modulo  $m$ . CQFD.

## Exercices

### Echauffements

- 1 Soit  $n_k$  le plus petit entier qui a exactement  $k$  diviseurs. Donnez  $n_k$  pour  $1 \leq k \leq 6$ .
- 2 Prouvez que  $\text{pgcd}(m, n) \cdot \text{ppcm}(m, n) = m \cdot n$ , puis utilisez cette identité pour exprimer  $\text{ppcm}(m, n)$  en fonction de  $\text{ppcm}(n \bmod m, m)$ , lorsque  $n \bmod m \neq 0$ . *Suggestion* : utilisez (4.12), (4.14) et (4.15).

- 3** Soit  $\pi(x)$  le nombre de nombres premiers inférieurs ou égaux à  $x$ . Prouvez ou réfutez la formule suivante :

$$\pi(x) - \pi(x-1) = [\text{x est premier}].$$

- 4** Que se passerait-il si la construction de Stern–Brocot partait des quatre fractions  $(\frac{0}{1}, \frac{1}{0}, \frac{0}{-1}, \frac{-1}{0})$  au lieu de  $(\frac{0}{1}, \frac{1}{0})$ ?
- 5** Trouvez des formules simples pour  $L^k$  et  $R^k$ , si  $L$  et  $R$  sont les matrices  $2 \times 2$  de (4.33).
- 6** Que signifie “ $a \equiv b \pmod{0}$ ” ?
- 7** Dix personnes numérotées de 1 à 10 forment un cercle, comme dans le problème de Josèphe, et chaque mième personne est exécutée (la valeur de  $m$  pouvant être bien plus grande que 10). Montrez que, pour tout  $k$ , les trois premières victimes ne peuvent pas être  $10, k$  et  $k+1$  (dans cet ordre).
- 8** Le système de numération par résidus  $(x \bmod 3, x \bmod 5)$  a une curieuse propriété : le nombre 13 s'écrit  $(1, 3)$ , c'est-à-dire comme dans le système décimal. Trouvez une façon de trouver toutes les instances d'une telle coïncidence sans calculer les quinze couples de résidus. En d'autres termes, donnez toutes les solutions des congruences

$$10x + y \equiv x \pmod{3}, \quad 10x + y \equiv y \pmod{5}.$$

*Suggestion :* utilisez le fait que  $10u+6v \equiv u \pmod{3}$  et que  $10u+6v \equiv v \pmod{5}$ .

- 9** Montrez que  $(3^{77} - 1)/2$  est un nombre impair et composite. *Suggestion :* regardez  $3^{77} \bmod 4$ .
- 10** Calculez  $\varphi(999)$ .
- 11** Trouvez une fonction  $\sigma(n)$  satisfaisant la propriété suivante :

$$g(n) = \sum_{0 \leq k \leq n} f(k) \iff f(n) = \sum_{0 \leq k \leq n} \sigma(k) g(n-k).$$

(C'est un analogue de la fonction de Möbius : voir (4.56)).

- 12** Simplifiez la formule  $\sum_{d|m} \sum_{k|d} \mu(k) g(d/k)$ .
- 13** Un entier strictement positif  $n$  est dit *sans carré* s'il n'existe aucun  $m > 1$  tel que  $m^2$  divise  $n$ . Trouvez une condition nécessaire et suffisante pour que  $n$  soit sans carré,
- a en fonction de la représentation en puissances premières (4.11) de  $n$ ;
  - b en fonction de  $\mu(n)$ .

## **Exercices de base**

- 14 Prouvez ou réfutez :

  - $\text{pgcd}(km, kn) = k \text{pgcd}(m, n)$  ;
  - $\text{ppcm}(km, kn) = k \text{ppcm}(m, n)$ .

15 Tout nombre premier est-il facteur d'au moins un nombre d'Euclide  $e_n$  ?

16 Que vaut la somme des inverses des  $n$  premiers nombres d'Euclide ?

17 Soit  $f_n$  le "nombre de Fermat"  $2^{2^n} + 1$ . Prouvez que  $f_m \perp f_n$  si  $m < n$ .

18 Montrez que si  $2^n + 1$  est premier, alors  $n$  est une puissance de 2.

19 Démontrez les identités suivantes, où  $n$  est un entier strictement positif :

$$\begin{aligned} \sum_{1 \leq k < n} \left\lfloor \frac{\varphi(k+1)}{k} \right\rfloor &= \sum_{1 < m \leq n} \left\lfloor \left( \sum_{1 \leq k < m} \left[ \left( \lfloor m/k \rfloor / \lceil m/k \rceil \right) \right]^{-1} \right) \right\rfloor \\ &= n - 1 - \sum_{k=1}^n \left\lceil \left\{ \frac{(k-1)! + 1}{k} \right\} \right\rceil. \end{aligned}$$

*Suggestion :* c'est une question d'astuce.

- 20 Pour tout entier  $n$  strictement positif, il existe un nombre premier  $p$  tel que  $n < p \leqslant 2n$ . (C'est le "postulat de Bertrand", que Joseph Bertrand a vérifié pour  $n < 3000000$  en 1845 et que Tchebychev a démontré pour tout  $n$  en 1850). Utilisez le postulat de Bertrand pour prouver qu'il existe une constante  $b \approx 1,25$  telle que les nombres

$$\lfloor 2^b \rfloor, \lfloor 2^{2^b} \rfloor, \lfloor 2^{2^{2^b}} \rfloor, \dots$$

sont tous premiers.

- 21 Soit  $P_n$  le  $n$ ième nombre premier. Trouvez une constante  $K$  telle que

$$\lfloor (10^{n^2} K) \bmod 10^n \rfloor = p_n.$$

- 22 Le nombre 111111111111111111 est premier. Montrer que, pour toute base  $b$ ,  $(11\dots1)_b$  ne peut être premier que si le nombre de 1 qu'il contient est premier.

23 Trouvez une récurrence pour  $\rho(k)$ , la fonction de la règle qui apparaît dans le chapitre. Montrez qu'il y a un lien entre  $\rho(k)$  et le disque déplacé à la  $k$ ième étape lorsqu'une Tour de Hanoi de taille  $n$  est déplacée en  $2^n - 1$  étapes, pour  $1 \leq k \leq 2^n - 1$ .

*C'est un test pour la vue ?*

*Y aurait pas de l'addition latérale là-dedans ?*

- 24 Généralisez (4.24) en exprimant  $\epsilon_p(n!)$  en fonction de  $v_p(n)$ , la somme des chiffres de la représentation de  $n$  en base  $p$ .
- 25 On dit que  $m$  *divise exactement*  $n$ , et on écrit  $m \mid n$ , si  $m \nmid n$  et  $m \perp n/m$ . Par exemple, dans le passage concernant les facteurs factoriels  $p^{\epsilon_p(n!)} \mid n!$ . Prouvez ou réfutez ce qui suit :
- $k \mid n$  et  $m \mid n \iff km \mid n$ , si  $k \perp m$ .
  - Pour tous  $m, n > 0$ , soit  $\text{pgcd}(m, n) \mid m$ , soit  $\text{pgcd}(m, n) \mid n$ .
- 26 Soit  $\mathcal{G}_N$  la suite de toutes les fractions positives ou nulles réduites  $m/n$  telles que  $mn \leq N$ . Par exemple,

$$\mathcal{G}_{10} = \frac{0}{1}, \frac{1}{10}, \frac{1}{9}, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}, \frac{3}{2}, \frac{2}{1}, \frac{3}{2}, \frac{4}{1}, \frac{5}{1}, \frac{6}{1}, \frac{7}{1}, \frac{8}{1}, \frac{9}{1}, \frac{10}{1}.$$

Est-il vrai que  $m'n - mn' = 1$  lorsque  $m/n$  précède immédiatement  $m'/n'$  dans  $\mathcal{G}_N$  ?

- 27 Donnez une règle simple pour comparer les nombres rationnels en se basant sur leurs représentations par des L et des R dans le système de numération de Stern–Brocot.
- 28 Voici la représentation de Stern–Brocot du nombre  $\pi$  :

$$\pi = R^3 L^7 R^{15} L R^{292} L R L R^2 L R^3 L R^{14} L^2 R \dots ;$$

servez-vous-en pour trouver toutes les approximations les plus simples de  $\pi$  dont les dénominateurs sont inférieurs à 50. La fraction  $\frac{22}{7}$  en fait-elle partie ?

- 29 On décrit dans le chapitre une correspondance entre les nombres réels en représentation binaire  $x = (.b_1 b_2 b_3 \dots)_2$  de l'intervalle  $[0..1]$  et les nombres réels en représentation de Stern–Brocot  $\alpha = B_1 B_2 B_3 \dots$  de l'intervalle  $[0..\infty)$ . Si  $x$  correspond à  $\alpha$  et  $x \neq 0$ , quel nombre correspond à  $1 - x$  ?
- 30 Démontrez la proposition suivante (le Théorème des Restes Chinois) : Soient  $m_1, \dots, m_r$  des entiers strictement positifs tels que  $m_j \perp m_k$  pour  $1 \leq j < k \leq r$ ; soit  $m = m_1 \dots m_r$ ; et soient  $a_1, \dots, a_r, A$  des entiers. Alors il existe exactement un entier  $a$  tel que

$$a \equiv a_k \pmod{m_k} \text{ pour } 1 \leq k \leq r \quad \text{et} \quad A \leq a < A + m.$$

- 31 En notation en base 10, un nombre est divisible par 3 si et seulement si la somme de ses chiffres est divisible par 3. Démontrez cette règle bien connue, puis généralisez-la.

- 32 Prouvez le théorème d'Euler (4.50) en généralisant la preuve de (4.47).
- 33 Montrez que, si  $f(m)$  et  $g(m)$  sont des fonctions multiplicatives, alors  $h(m) = \sum_{d \mid m} f(d) g(m/d)$  l'est aussi.

*Faut-il prononcer “œufs-l’air” ou “oye-l’air” ? Faut-il prononcer “œufs-clide” ou “oye-clide” ?*

**34** Démontrez que (4.56) est un cas particulier de (4.61).

**Devoirs à la maison**

**35** Soit  $I(m, n)$  une fonction qui satisfait la relation

$$I(m, n)m + I(n, m)n = \text{pgcd}(m, n),$$

si  $m$  et  $n$  sont des entiers positifs ou nuls tels que  $m \neq n$ . Ainsi,  $I(m, n) = m'$  et  $I(n, m) = n'$  dans (4.5) ; la valeur de  $I(m, n)$  est un *inverse* de  $m$  par rapport à  $n$ . Trouvez une récurrence qui définit  $I(m, n)$ .

**36** Considérez l'ensemble  $Z(\sqrt{10}) = \{ m + n\sqrt{10} \mid m, n \text{ entiers} \}$ . On dit que le nombre  $m + n\sqrt{10}$  est une *unité* si  $m^2 - 10n^2 = \pm 1$ , car il a un inverse (c'est-à-dire car  $(m + n\sqrt{10}) \cdot \pm(m - n\sqrt{10}) = 1$ ). Par exemple,  $3 + \sqrt{10}$  est une unité, tout comme  $19 - 6\sqrt{10}$ . Dans toute factorisation, on peut insérer des couples d'unités qui se neutralisent mutuellement ; nous ignorerons donc ces cas. Les nombres de  $Z(\sqrt{10})$  qui ne sont pas des unités sont dits premiers s'ils ne peuvent pas s'écrire comme un produit de deux nombres qui ne sont pas des unités. Montrez que 2, 3 et  $4 \pm \sqrt{10}$  sont premiers dans  $Z(\sqrt{10})$ . *Suggestion* : si  $2 = (k + l\sqrt{10}) \times (m + n\sqrt{10})$  alors  $4 = (k^2 - 10l^2)(m^2 - 10n^2)$ . D'autre part, le carré de tout entier mod 10 est 0, 1, 4, 5, 6 ou 9.

**37** Prouvez (4.17). *Suggestion* : montrez que  $e_n - \frac{1}{2} = (e_{n-1} - \frac{1}{2})^2 + \frac{1}{4}$ , puis considérez  $2^{-n} \log(e_n - \frac{1}{2})$ .

**38** Démontrez que si  $a \perp b$  et  $a > b$ , alors

$$\text{pgcd}(a^m - b^m, a^n - b^n) = a^{\text{pgcd}(m, n)} - b^{\text{pgcd}(m, n)}, \quad 0 \leq m < n.$$

(Toutes les variables sont entières). *Suggestion* : utilisez l'algorithme d'Euclide.

**39** Soit  $S(m)$  le plus petit entier positif  $n$  pour lequel il existe une suite croissante d'entiers

$$m = a_1 < a_2 < \dots < a_t = n$$

telle que  $a_1 a_2 \dots a_t$  soit un carré parfait. (Si  $m$  est un carré parfait, on peut poser  $t = 1$  et  $n = m$ ). Par exemple,  $S(2) = 6$  car la meilleure suite de ce type est  $a_1 = 2, a_2 = 3, a_3 = 6$ . Voici les premières valeurs de  $S$  :

$n$	1	2	3	4	5	6	7	8	9	10	11	12
$S(n)$	1	6	8	4	10	12	14	15	9	18	22	20

Montrez que  $S(m) \neq S(m')$  pour tous  $0 < m < m'$ .

**40** Si la représentation en base  $p$  de  $n$  est  $(a_m \dots a_1 a_0)_p$ , montrez que

$$n!/p^{\epsilon_p(n!)} \equiv (-1)^{\epsilon_p(n!)} a_m! \dots a_1! a_0! \pmod{p}.$$

(Le membre gauche est égal à la factorielle de  $n$  dans laquelle on a supprimé le facteur  $p$ . Lorsque  $n = p$ , on retrouve le théorème de Wilson).

- 41** a Montrez que si  $p \bmod 4 = 3$ , il n'existe pas d'entier  $n$  tel que  $p$  divise  $n^2 + 1$ . *Suggestion :* utilisez le théorème de Fermat.  
 b Montrez que si  $p \bmod 4 = 1$ , alors, par contre, cet entier existe. *Suggestion :* réécrivez  $(p-1)!$  en  $(\prod_{k=1}^{(p-1)/2} k(p-k))$  et pensez au théorème de Wilson.
- 42** Soient deux fractions réduites  $m/n$  et  $m'/n'$ . Considérez la fraction réduite de la somme  $m/n + m'/n'$ , et montrez que son dénominateur est égal à  $nn'$  si et seulement si  $n \perp n'$ . (En d'autres termes,  $(mn' + m'n)/nn'$  est déjà réduite si et seulement si  $n$  et  $n'$  n'ont pas de facteur commun).
- 43** Il y a  $2^k$  noeuds situés au niveau  $k$  de l'arbre de Stern–Brocot. Ils correspondent aux matrices  $L^k, L^{k-1}R, \dots, R^k$ . Montrez que cette suite peut être obtenue en partant de  $L^k$  puis en multipliant successivement par

$$\begin{pmatrix} 0 & -1 \\ 1 & 2\rho(n)+1 \end{pmatrix}$$

pour  $1 \leq n < 2^k$ , où  $\rho(n)$  est la fonction de la règle.

- 44** Montrez qu'un joueur de base-ball qui a une moyenne au bâton de 0,316 a forcément frappé la balle au moins 19 fois. (S'il a réussi  $m$  fois sur  $n$ , alors  $m/n \in [0,3155 \dots 0,3165]$ ).
- 45** Le nombre 9376 présente une propriété bien particulière d'auto-reproduction :

$$9376^2 = 87909376.$$

Combien existe-t-il de nombres à 4 chiffres  $x$  qui satisfont l'équation  $x^2 \bmod 10000 = x$ ? Combien existe-t-il de nombres à  $n$  chiffres  $x$  qui satisfont  $x^2 \bmod 10^n = x$ ?

- 46** a Prouvez que si  $n^j \equiv 1$  et  $n^k \equiv 1 \pmod{m}$ , alors  $n^{\text{pgcd}(j,k)} \equiv 1$ .  
 b Montrez que si  $n > 1$ , alors  $2^n \not\equiv 1 \pmod{n}$ . *Suggestion :* considérez le plus petit facteur premier de  $n$ .

- 47** Montrez que si  $n^{m-1} \equiv 1 \pmod{m}$  et  $n^{(m-1)/p} \not\equiv 1 \pmod{m}$  pour tout nombre premier tel que  $p \nmid (m-1)$ , alors  $m$  est premier. *Suggestion :* montrez que, si la condition est vraie, alors les nombres  $n^k \pmod{m}$  sont distincts, pour  $1 \leq k < m$ .
- 48** Généralisez le théorème de Wilson (4.49) en déterminant la valeur de l'expression  $(\prod_{1 \leq n < m, n \perp m} n) \pmod{m}$ , pour  $m > 1$ .
- 49** Soit  $R(N)$  le nombre de couples d'entiers  $(m, n)$  tels que  $1 \leq m \leq N$ ,  $1 \leq n \leq N$  et  $m \perp n$ .
- Exprimez  $R(N)$  avec la fonction  $\Phi$ .
  - Prouvez que  $R(N) = \sum_{d \geq 1} [N/d]^2 \mu(d)$ .
- 50** Soient  $m$  un entier strictement positif et

$$\omega = e^{2\pi i/m} = \cos(2\pi/m) + i \sin(2\pi/m).$$

On dit que  $\omega$  est une *racine mième de l'unité*, car  $\omega^m = e^{2\pi i} = 1$ . En fait, chacun des nombres complexes  $\omega^0, \omega^1, \dots, \omega^{m-1}$  est une racine mième de l'unité, car  $(\omega^k)^m = e^{2\pi k i} = 1$ ; donc  $z - \omega^k$  est un facteur du polynôme  $z^m - 1$ , pour  $0 \leq k < m$ . Le nombre  $z^m - 1$  se décompose donc en  $m+1$  facteurs distincts complexes

$$z^m - 1 = \prod_{0 \leq k < m} (z - \omega^k).$$

- a Soit  $\Psi_m(z) = \prod_{0 \leq k < m, k \perp m} (z - \omega^k)$ . (Ce polynôme, de degré  $\varphi(m)$ , est appelé le *polynôme cyclotomique d'ordre m*). Prouvez que

$$z^m - 1 = \prod_{d \mid m} \Psi_d(z).$$

- b Montrez que  $\Psi_m(z) = \prod_{d \mid m} (z^d - 1)^{\mu(m/d)}$ .

### Problèmes d'examen

- 51** Démontrez le théorème de Fermat (4.48) en développant  $(1+1+\cdots+1)^p$ .
- 52** Soient  $n$  et  $x$  des entiers strictement positifs tels que  $x$  n'a pas de diviseur  $\leq n$  (à part 1), et soit  $p$  un nombre premier. Montrez qu'il existe au moins  $\lfloor n/p \rfloor$  multiples de  $p$  parmi les nombres  $\{x-1, x^2 - 1, \dots, x^{n-1} - 1\}$ .
- 53** Trouvez tous les entiers strictement positifs tels que  $n \mid \lceil (n-1)!/(n+1) \rceil$ .
- 54** Calculez  $1000! \pmod{10^{250}}$  à la main.

55 Soit  $P_n = \prod_{k=1}^n k!$  le produit des  $n$  premières factorielles. Montrez que  $P_{2n}/P_n^4$  est un entier, pour tout entier positif  $n$ .

56 Montrez que

$$\left( \prod_{k=1}^{2n-1} k^{\min(k, 2n-k)} \right) / \left( \prod_{k=1}^{n-1} (2k+1)^{2n-2k-1} \right)$$

est une puissance de 2.

57 Soit  $S(m, n)$  l'ensemble des entiers  $k$  tels que

$$m \bmod k + n \bmod k \geq k.$$

Par exemple,  $S(7, 9) = \{2, 4, 5, 8, 10, 11, 12, 13, 14, 15, 16\}$ . Prouvez que

$$\sum_{k \in S(m, n)} \varphi(k) = mn.$$

*Suggestion :* montrez que  $\sum_{1 \leq m \leq n} \sum_{d|m} \varphi(d) = \sum_{d \geq 1} \varphi(d) \lfloor n/d \rfloor$ , considérez ensuite  $\lfloor (m+n)/d \rfloor - \lfloor m/d \rfloor - \lfloor n/d \rfloor$ .

58 Soit  $f(m) = \sum_{d|m} d$ . Trouvez une condition nécessaire et suffisante pour que  $f(m)$  soit une puissance de 2.

### Questions subsidiaires

59 Prouvez que si  $x_1, \dots, x_n$  sont des entiers strictement positifs tels que  $1/x_1 + \dots + 1/x_n = 1$ , alors  $\max(x_1, \dots, x_n) < e_n$ . *Suggestion :* démontrez par induction ce résultat plus fort : "si  $1/x_1 + \dots + 1/x_n + 1/\alpha = 1$ , où  $x_1, \dots, x_n$  sont des entiers strictement positifs et  $\alpha$  est un nombre rationnel  $\geq \max(x_1, \dots, x_n)$ , alors  $\alpha+1 \leq e_{n+1}$  et  $x_1 \dots x_n (\alpha+1) \leq e_1 \dots e_n e_{n+1}$ ." (La preuve n'est pas triviale).

60 Prouvez qu'il existe une constante  $P$  telle que la formule (4.18) ne donne que des nombres premiers. Vous pouvez utiliser le fait suivant (éminemment non trivial) : si  $p$  est suffisamment grand et si  $\theta > \frac{6}{11}$ , alors il existe un nombre premier entre  $p$  et  $p + p^\theta$ .

61 Démontrez que si  $m/n, m'/n'$  et  $m''/n''$  sont des éléments consécutifs de  $\mathcal{F}_N$ , alors

$$\begin{aligned} m'' &= \lfloor (n+N)/n' \rfloor m' - m, \\ n'' &= \lfloor (n+N)/n' \rfloor n' - n. \end{aligned}$$

(Cette récurrence permet de calculer les éléments de  $\mathcal{F}_N$  dans l'ordre, en partant de  $\frac{0}{1}$  et  $\frac{1}{N}$ ).

- 62 Quel est le nombre binaire qui correspond à  $e$ , dans la correspondance entre la représentation binaire et la représentation de Stern-Brocot ? (On ne demande pas une formule close ; donnez votre réponse sous forme de somme infinie).

- 63 En n'utilisant que des méthodes présentées dans le chapitre, montrez que si le Dernier Théorème de Fermat (4.46) était faux, le plus petit nombre  $n$  pour lequel il faillirait serait forcément premier. (Vous pouvez supposer que (4.46) est vrai pour  $n = 4$ ). Montrez en outre que si  $a^p + b^p = c^p$  est le plus petit contre-exemple, alors il existe un entier  $m$  tel que

$$a + b = \begin{cases} m^p, & \text{si } p \nmid c, \\ p^{p-1}m^p, & \text{si } p \mid c. \end{cases}$$

Par conséquent,  $c \geq m^p/2$  doit être extrêmement grand. *Suggestion :* posez  $x = a + b$ , et remarquez que  $\text{pgcd}(x, (a^p + (x-a)^p)/x) = \text{pgcd}(x, pa^{p-1})$ .

- 64 La suite de Peirce  $\mathcal{P}_N$  d'ordre  $N$  est une suite infinie de fractions, séparées par des signes “<” ou “=”, qui contient toutes les fractions positives ou nulles  $m/n$  telles que  $m \geq 0$  et  $n \leq N$  (y compris les fractions non réduites). On la construit récursivement en partant de

$$\mathcal{P}_1 = \frac{0}{1} < \frac{1}{1} < \frac{2}{1} < \frac{3}{1} < \frac{4}{1} < \frac{5}{1} < \frac{6}{1} < \frac{7}{1} < \frac{8}{1} < \frac{9}{1} < \frac{10}{1} < \dots$$

Pour  $N \geq 1$ , on construit  $\mathcal{P}_{N+1}$  en insérant deux symboles juste avant le  $kN$ ème symbole de  $\mathcal{P}_N$ , pour tout  $k > 0$ . Ces deux symboles sont

$$\frac{k-1}{N+1} = , \quad \text{si } kN \text{ est impair;} \\ \mathcal{P}_{N,kN} \quad \frac{k-1}{N+1} , \quad \text{si } kN \text{ est pair.}$$

La notation  $\mathcal{P}_{N,j}$  désigne le  $j$ ème symbole de  $\mathcal{P}_N$ , qui sera “<” ou “=” si  $j$  est pair, et une fraction si  $j$  est impair. Par exemple,

$$\begin{aligned} \mathcal{P}_2 &= \frac{0}{2} = \frac{0}{1} < \frac{1}{2} < \frac{2}{1} = \frac{1}{1} < \frac{3}{2} < \frac{4}{1} = \frac{2}{1} < \frac{5}{2} < \frac{6}{1} = \frac{3}{1} < \frac{7}{2} < \frac{8}{1} = \frac{4}{1} < \frac{9}{2} < \frac{10}{1} = \frac{5}{1} < \dots ; \\ \mathcal{P}_3 &= \frac{0}{2} = \frac{0}{3} = \frac{0}{1} < \frac{1}{3} < \frac{2}{2} = \frac{3}{1} = \frac{1}{1} < \frac{4}{3} < \frac{3}{2} < \frac{5}{3} < \frac{4}{2} = \frac{6}{3} = \frac{2}{1} < \frac{7}{3} < \frac{5}{2} < \dots ; \\ \mathcal{P}_4 &= \frac{0}{2} = \frac{0}{4} = \frac{0}{3} = \frac{0}{1} < \frac{1}{4} < \frac{1}{3} < \frac{2}{4} = \frac{1}{2} < \frac{2}{3} < \frac{3}{4} < \frac{2}{1} = \frac{4}{3} = \frac{3}{1} < \frac{5}{4} < \frac{4}{3} < \frac{6}{4} = \dots ; \\ \mathcal{P}_5 &= \frac{0}{2} = \frac{0}{4} = \frac{0}{5} = \frac{0}{3} = \frac{0}{1} < \frac{1}{4} < \frac{1}{3} < \frac{2}{5} < \frac{3}{4} = \frac{1}{2} < \frac{2}{5} < \frac{3}{4} < \frac{4}{5} < \frac{2}{1} = \frac{4}{3} = \dots ; \\ \mathcal{P}_6 &= \frac{0}{2} = \frac{0}{4} = \frac{0}{6} = \frac{0}{5} = \frac{0}{3} = \frac{0}{1} < \frac{1}{6} < \frac{1}{5} < \frac{1}{4} < \frac{2}{6} = \frac{1}{3} < \frac{2}{5} < \frac{2}{4} = \frac{3}{6} = \frac{1}{2} < \frac{3}{5} < \frac{4}{6} = \dots . \end{aligned}$$

(Les éléments qui sont égaux apparaissent dans un ordre un peu bizarre). Prouvez que les signes “<” et “=” que l'on écrit en suivant les

règles ci-dessus décrivent bien les relations entre les fractions adjacentes de la suite de Peirce.

### Sujets de recherche

- 65 Tous les nombres d'Euclide  $e_n$  sont-ils sans carré ?
- 66 Tous les nombres de Mersenne  $2^p - 1$  sont-ils sans carré ?
- 67 Prouvez ou réfutez le fait que  $\max_{1 \leq j < k \leq n} a_k / \text{pgcd}(a_j, a_k) \geq n$ , pour toute suite d'entiers  $0 < a_1 < \dots < a_n$ .
- 68 Existe-t-il une constante  $Q$  telle que  $\lfloor Q^{2^n} \rfloor$  est premier pour tout  $n$  positif ou nul ?
- 69 Soit  $P_n$  le  $n$ ième nombre premier. Prouvez ou réfutez le fait que  $P_{n+1} - P_n = O(\log P_n)^2$ .
- 70 Est-il vrai que  $\epsilon_3(n!) = \epsilon_2(n!)/2$  pour une infinité de  $n$  ?
- 71 Prouvez ou réfutez : si  $k \neq 1$ , il existe  $n > 1$  tel que  $2^n \equiv k \pmod{n}$ . Y a-t-il une infinité de ces nombres  $n$  ?
- 72 Prouvez ou réfutez : pour tout entier  $a$ , il existe une infinité de  $n$  tels que  $\varphi(n) \mid (n + a)$ .
- 73 Si les  $\Phi(n) + 1$  termes de la suite de Farey

$$\mathcal{F}_n = \langle \mathcal{F}_n(0), \mathcal{F}_n(1), \dots, \mathcal{F}_n(\Phi(n)) \rangle$$

étaient uniformément distribués, on aurait  $\mathcal{F}_n(k) \approx k/\Phi(n)$ . Par conséquent, la somme  $D(n) = \sum_{k=0}^{\Phi(n)} |\mathcal{F}_n(k) - k/\Phi(n)|$  donne la "déviation de  $\mathcal{F}_n$  par rapport à l'uniformité". Est-il vrai que  $D(n) = O(n^{1/2+\epsilon})$  pour tout  $\epsilon > 0$  ?

- 74 Quel est le nombre approximatif de valeurs distinctes dans l'ensemble  $\{0! \bmod p, 1! \bmod p, \dots, (p-1)! \bmod p\}$ , lorsque  $p \rightarrow \infty$  ?

# 5

## Coefficients binomiaux

RESPIRONS UN PEU. Nous avons eu notre lot de complications dans les chapitres précédents, avec des sommes, des parties entières, des fonctions mod, phi, mu... Stop ! Maintenant, nous allons étudier les coefficients binomiaux, qui ont deux qualités non négligeables : d'une part ils apparaissent très souvent dans diverses applications ; d'autre part, ils sont plus faciles à manipuler.

Veinards que nous sommes !

### 5.1 IDENTITÉS DE BASE

Le symbole  $\binom{n}{k}$  est un coefficient du binôme ou coefficient binomial, ou encore tout simplement "binomial". On l'appelle ainsi car il est lié à une propriété importante que nous verrons un peu plus loin : la formule du binôme. Le symbole se lit "binomial de  $n$  et  $k$ ".

Le coefficient du binôme est attaché à une interprétation combinatoire très classique : c'est le nombre de façons de choisir un sous-ensemble à  $k$  éléments dans un ensemble à  $n$  éléments. Par exemple, il y a six façons de choisir deux éléments dans l'ensemble  $\{1, 2, 3, 4\}$  :

$$\{1, 2\}, \quad \{1, 3\}, \quad \{1, 4\}, \quad \{2, 3\}, \quad \{2, 4\}, \quad \{3, 4\}.$$

On l'appelle aussi le "nombre de combinaisons d'ordre  $k$  de  $n$  éléments".

Par conséquent,  $\binom{4}{2} = 6$ .

Pour exprimer le nombre  $\binom{n}{k}$  de façon plus pratique, commençons par compter le nombre de *suites* de  $k$  éléments parmi  $n$ . Dans une suite, contrairement à un ensemble, l'ordre des éléments compte. Nous allons utiliser un argument qui nous déjà servi pour compter le nombre de permutations de  $n$  éléments au chapitre 4. Il y a  $n$  façons de choisir le premier élément de la suite ; puis  $n - 1$  façons de choisir le second, et ainsi de suite jusqu'au  $k$ ième élément qui peut être choisi de  $n - k + 1$  manières différentes. Cela donne donc  $n(n - 1) \dots (n - k + 1) = n^k$  possibilités en tout. Comme il y a  $k!$  façons d'ordonner les éléments d'un sous-ensemble de cardinal  $k$ , on en déduit qu'il y a  $k!$  fois plus de *suites* à  $k$  éléments que de *sous-ensembles*

à  $k$  éléments. Par conséquent,

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k(k-1)\dots(1)}.$$

En particulier,

$$\binom{4}{2} = \frac{4 \cdot 3}{2 \cdot 1} = 6,$$

ce qui confirme notre précédent résultat.

Nous dirons que  $n$  est l'*indice du haut* et  $k$  l'*indice du bas*. Si on se base sur l'interprétation combinatoire du binomial, les indices doivent forcément être des entiers positifs ou nuls, tout simplement parce qu'il n'existe pas d'ensembles de cardinal négatif ou non entier. Toutefois, il s'avère que les coefficients binomiaux peuvent être utilisés tout à fait indépendamment de cette interprétation. Ainsi, nous permettrons à l'indice du haut d'être un nombre réel quelconque (ou même un complexe), et à l'indice du bas d'être un entier quelconque. Voici donc la définition formelle de ce coefficient binomial plus général :

$$\binom{r}{k} = \begin{cases} \frac{r(r-1)\dots(r-k+1)}{k(k-1)\dots(1)} = \frac{r^k}{k!}, & k \geq 0 \text{ entier;} \\ 0, & k < 0 \text{ entier.} \end{cases} \quad (5.1)$$

Cette définition appelle plusieurs remarques importantes. Premièrement, notez que l'indice du haut est noté  $r$  au lieu de  $n$ , pour bien insister sur le fait que n'importe quel réel peut tenir ce rôle. Par exemple,  $\binom{-1}{3} = (-1)(-2)(-3)/(3 \cdot 2 \cdot 1) = -1$ . Ici, il n'y a pas d'interprétation combinatoire à ce résultat. Le cas particulier  $r = -1$  s'avérera particulièrement important, tout comme le cas  $r = -1/2$ .

Deuxièmement,  $\binom{r}{k}$  peut être vu comme un polynôme de degré  $k$  en  $r$ . Nous verrons que ce point de vue peut être très utile.

Troisièmement, l'indice du bas doit obligatoirement être un entier. Il est possible de généraliser la définition pour que cet indice puisse être un réel quelconque. C'est cependant inutile dans la plupart des applications ; c'est pourquoi nous ne nous intéresserons à cette généralisation que plus loin dans ce chapitre.

Enfin, remarquez les restrictions " $k \geq 0$  entier" et " $k < 0$  entier" à droite de la définition. Dans toutes les identités que nous étudierons, nous préciserons toujours les restrictions de ce type. Ainsi, leurs champs d'application seront bien clairs. Cependant, quand nous manipulerons des coefficients binomiaux, pour ne pas alourdir les calculs, il pourra nous arriver, temporairement, de ne pas nous préoccuper de certaines restrictions.

Dans ces cas-là, nous vérifierons a posteriori qu'aucune restriction n'a été violée. Cette vérification est bien entendu obligatoire.

En voici une illustration. Dans l'immense majorité des cas rencontrés, le binomial  $\binom{n}{n}$  vaut 1. Si on ne fait pas attention, on peut donc croire qu'il est effectivement toujours égal à 1. Cependant, si on regarde attentivement la définition (5.1), on voit que  $\binom{n}{n}$  ne vaut 1 que lorsque  $n \geq 0$  (en supposant que  $n$  est un entier). Si  $n < 0$ , alors  $\binom{n}{n} = 0$ .

Avant d'en arriver aux identités qui nous permettront de dompter les coefficients binomiaux, jetons un coup d'œil sur quelques premières valeurs. La table 166 constitue le début du *triangle de Pascal*. Elle est ainsi nommée

**Table 166** Le triangle de Pascal.

$n$	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

car les coefficients binomiaux ont fait l'objet d'un très important traité rédigé par Blaise Pascal (1623–1662) [285]. Les cases vides correspondent en réalité à des valeurs nulles, à cause du zéro dans le numérateur de (5.1). Par exemple,  $\binom{1}{2} = (1 \cdot 0)/(2 \cdot 1) = 0$ . On les a laissées en blanc pour mieux voir le reste de la table.

Il est bon de connaître les formules pour les trois premières colonnes :

$$\binom{r}{0} = 1, \quad \binom{r}{1} = r, \quad \binom{r}{2} = \frac{r(r-1)}{2}, \quad (5.2)$$

pour n'importe quel réel  $r$ . (Souvenez-vous : la formule  $\binom{n+1}{2} = \frac{1}{2}n(n+1)$  est celle des nombres triangulaires que nous avons rencontrés au chapitre 1 ; ils constituent la colonne  $\binom{n}{2}$  de la table 166). Il peut être bon aussi de mémoriser les cinq ou six premières lignes du triangle de Pascal, de sorte que, si un motif comme 1, 4, 6, 4, 1 apparaît dans un problème, nous puissions immédiatement penser aux coefficients binomiaux.

*Les coefficients étaient déjà connus en Asie bien des siècles avant, mais Pascal ne pouvait pas le savoir [90].*

*En Italie, on l'appelle le triangle de Tartaglia.*

*"C'est une chose  
étrange combien  
il est fertile en  
propriétés."*

—B. Pascal [285]

Les nombres du triangle de Pascal satisfont pratiquement une infinité d'identités. Il n'est donc pas étonnant d'y trouver des relations surprenantes dès qu'on le regarde un peu attentivement. Par exemple, il existe une curieuse "propriété de l'hexagone", illustrée par les nombres 56, 28, 36, 120, 210, 126 qui entourent 84, vers le bas de la table. Leurs deux "produits alternés" sont égaux :  $56 \cdot 36 \cdot 210 = 28 \cdot 120 \cdot 126 = 423360$ . Cette propriété reste vraie pour n'importe quel hexagone similaire pris dans le triangle de Pascal.

Passons maintenant aux identités. Le but de cette section est de nous apprendre quelque règles simples qui nous permettront de résoudre la grande majorité des problèmes où apparaissent les coefficients binomiaux.

Dans le cas où l'indice du haut est un entier  $n$  et l'indice du bas  $k$  lui est inférieur ou égal, la définition (5.1) peut être reformulée en termes de factorielles :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad n \geq k \geq 0 \text{ entiers.} \quad (5.3)$$

Pour obtenir cette formule, il suffit de multiplier le numérateur et le dénominateur de (5.1) par  $(n-k)!$ . Cette expression du binomial est parfois utile pour résoudre des problèmes (en particulier pour démontrer la propriété de l'hexagone). D'une autre côté, on est souvent amené à faire l'inverse, c'est-à-dire traduire des factorielles en binomiaux.

Cette formulation par des factorielles fait apparaître une symétrie ; chaque ligne du triangle de Pascal peut se lire indifféremment dans un sens ou dans l'autre :

$$\binom{n}{k} = \binom{n}{n-k}, \quad \begin{matrix} n \geq 0 \text{ entier,} \\ k \text{ entier.} \end{matrix} \quad (5.4)$$

Du point de vue combinatoire, cette égalité a un sens : choisir les  $k$  éléments d'un sous-ensemble revient à choisir les  $n - k$  qui n'y seront pas.

Dans la formule (5.4), il est naturel d'imposer à  $n$  et  $k$  d'être entiers car les indices du bas doivent être entiers. La raison pour laquelle  $n$  ne peut pas être négatif est moins évidente. Supposons par exemple que  $n = -1$ . L'équation

$$\binom{-1}{k} \stackrel{?}{=} \binom{-1}{-1-k}$$

est-elle valide ? La réponse est non. Si  $k = 0$ , par exemple, on trouve 0 à gauche et 1 à droite. Pour tout entier  $k \geq 0$ , le membre gauche est égal à

$$\binom{-1}{k} = \frac{(-1)(-2)\dots(-k)}{k!} = (-1)^k,$$

ce qui donne 1 ou  $-1$  selon la parité de  $k$ . Le membre droit, lui, est égal à 0, du fait que l'indice du bas est négatif. En revanche, si  $k$  est négatif, le membre gauche est nul et le membre droit est égal à

$$\binom{-1}{-1-k} = (-1)^{-1-k}.$$

Par conséquent, l'équation " $\binom{-1}{k} = \binom{-1}{-1-k}$ " est toujours fausse !

L'identité de symétrie est fausse aussi pour tout autre  $n$  négatif. Il est malheureusement facile d'oublier cette restriction, surtout si l'indice du haut est une expression qui ne peut être négative que pour des valeurs très particulières de certaines variables. Quiconque a beaucoup manipulé des coefficients binomiaux est tombé au moins trois fois dans ce piège.

D'un autre côté, l'identité de symétrie se rachète en quelque sorte par le fait qu'elle est valable même si  $k < 0$  ou  $k > n$  (dans ce cas, les deux membres sont nuls). Sinon, si  $0 \leq k \leq n$ , la symétrie se déduit immédiatement de (5.3) :

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\ &= \frac{n!}{(n-(n-k))! (n-k)!} = \binom{n}{n-k}.\end{aligned}$$

*Tout ce que j'espère,  
c'est ne pas y  
tomber le jour de  
l'examen.*

Voici une autre identité importante, qui permet de faire entrer ou sortir des choses d'un coefficient binomial :

$$\binom{r}{k} = \frac{r}{k} \binom{r-1}{k-1}, \quad k \neq 0 \text{ entier.} \quad (5.5)$$

La restriction sur  $k$  sert simplement à interdire la division par 0. On dit que l'équation (5.5) est une *identité d'absorption*, car, grâce à elle, un coefficient binomial peut absorber une variable qui serait nuisible à l'extérieur. Cette identité se déduit de la définition (5.1), car  $r^k = r(r-1)^{k-1}$  et  $k! = k(k-1)!$  lorsque  $k > 0$  ; et si  $k < 0$ , on obtient 0 des deux côtés.

En multipliant les deux membres de (5.5) par  $k$ , on obtient une autre identité d'absorption, valable même si  $k = 0$  :

$$k \binom{r}{k} = r \binom{r-1}{k-1}, \quad k \text{ entier.} \quad (5.6)$$

Elle a aussi un petit camarade qui ne modifie pas l'indice du bas :

$$(r-k) \binom{r}{k} = r \binom{r-1}{k}, \quad k \text{ entier.} \quad (5.7)$$

On peut obtenir (5.7) en faisant un sandwich avec une tranche de (5.6) entre deux symétries :

$$\begin{aligned} (r-k)\binom{r}{k} &= (r-k)\binom{r}{r-k} \quad (\text{par symétrie}) \\ &= r\binom{r-1}{r-k-1} \quad (\text{d'après (5.6)}) \\ &= r\binom{r-1}{k}. \quad (\text{par symétrie}) \end{aligned}$$

Attendez une minute ! Nous avons dit que l'identité était valable pour *tout* réel  $r$  ; or, le calcul que nous venons de faire n'est valable que si  $r$  est un entier strictement positif. En effet, l'indice du haut  $r-1$  doit être un entier positif ou nul pour qu'on puisse appliquer la propriété de symétrie (5.4). Aurions-nous triché ? Bien sûr que non. Il est vrai que le calcul n'est correct que si  $r$  est un entier strictement positif, mais l'identité est valable pour tout  $r$ , car les deux membres de (5.7) sont des polynômes en  $r$  de degré  $k+1$ . Tout polynôme de degré inférieur ou égal à  $d$  a au plus  $d$  zéros distincts. Par conséquent, la différence de deux tels polynômes, qui est aussi de degré inférieur ou égal à  $d$ , ne peut pas s'annuler en plus de  $d$  points, sauf si elle est identiquement nulle. En d'autres termes, si deux polynômes de degré inférieur ou égal à  $d$  ont même valeur en plus de  $d$  points, alors ils sont égaux. Nous avons démontré que  $(r-k)\binom{r}{k} = r\binom{r-1}{k}$  pour tout entier strictement positif  $r$  ; donc ces deux polynômes ont même valeur en une infinité de points, donc ils sont égaux.

Cette technique de preuve, que l'on appelle l'*argument polynomial*, est très souvent utilisée pour étendre des identités des entiers vers les réels. Il y a bien sûr des équations qui ne mettent pas en jeu des polynômes, comme l'identité de symétrie (5.4). Dans ces cas là, on ne peut pas utiliser cette méthode. Toutefois, nous rencontrerons beaucoup d'identités qui ont la forme adéquate.

Voici par exemple une autre identité polynomiale, appelée la *formule d'addition*. C'est peut-être la plus importante des identités sur les coefficients binomiaux :

$$\binom{r}{k} = \binom{r-1}{k} + \binom{r-1}{k-1}, \quad k \text{ entier.} \quad (5.8)$$

Cette formule nous dit que chaque nombre du triangle de Pascal est égal à la somme du nombre situé juste au-dessus et du nombre situé juste à gauche de ce dernier. Elle est valable aussi lorsque  $r$  est négatif, réel, ou même complexe. La seule restriction est que  $k$  doit être un entier, tout simplement pour que les coefficients binomiaux soient définis.

*(Enfin, pas ici en tout cas).*

Une façon de prouver la formule d'addition consiste à supposer que  $r$  est un entier positif et à utiliser l'interprétation combinatoire. Souvenez-vous que  $\binom{r}{k}$  représente le nombre de sous-ensembles à  $k$  éléments d'un ensemble de cardinal  $r$ . Supposons que nous ayons un ensemble de  $r$  œufs parmi lesquels il y a exactement un œuf pourri. Il y a  $\binom{r}{k}$  façons de choisir  $k$  de ces œufs. Exactement  $\binom{r-1}{k}$  de ces possibilités ne concernent que des bons œufs, tandis que  $\binom{r-1}{k-1}$  d'entre elles contiennent le mauvais, car elles contiennent  $k-1$  des  $r-1$  bons œufs. En additionnant ces deux nombres, on tombe sur (5.8). Dans ce raisonnement, nous avons supposé que  $r$  est un entier strictement positif, et que  $k \geq 0$ . Toutefois, comme les deux membres de l'identité sont nuls lorsque  $k < 0$ , l'argument polynomial permet d'établir que (5.8) est valable pour tous les autres cas.

On peut aussi démontrer la formule (5.8) en additionnant les deux identités d'absorption (5.7) et (5.6) :

$$(r-k)\binom{r}{k} + k\binom{r}{k} = r\binom{r-1}{k} + r\binom{r-1}{k-1}.$$

Le membre gauche est égal à  $r\binom{r}{k}$ , et on peut tout diviser par  $r$ . Ce calcul est correct pour tout  $r$  non nul, et on vérifie facilement que le résultat est vrai aussi si  $r = 0$ .

Si on n'a pas trouvé une des idées astucieuses ci-dessus, on peut toujours établir (5.8) en utilisant tout simplement la définition du binomial : si  $k > 0$ ,

$$\begin{aligned}\binom{r-1}{k} + \binom{r-1}{k-1} &= \frac{(r-1)^k}{k!} + \frac{(r-1)^{k-1}}{(k-1)!} \\ &= \frac{(r-1)^{k-1}(r-k)}{k!} + \frac{(r-1)^{k-1}k}{k!} \\ &= \frac{(r-1)^{k-1}r}{k!} \\ &= \frac{r^k}{k!} = \binom{r}{k}.\end{aligned}$$

Comme précédemment, le cas  $k \leq 0$  est facile à vérifier.

Nous venons de voir trois preuves tout à fait différentes de la formule d'addition. Cette diversité n'est pas surprenante : comme les coefficients binomiaux satisfont énormément d'identités utiles, une même formule peut être naturellement prouvée de différentes manières.

La formule d'addition constitue en fait une récurrence pour les nombres du triangle de Pascal. Nous verrons qu'elle est particulièrement utile pour démontrer d'autres identités par induction. En attendant, contentons-nous

de la déplier pour trouver un nouveau résultat :

$$\begin{aligned}
 \binom{5}{3} &= \binom{4}{3} + \binom{4}{2} \\
 &= \binom{4}{3} + \binom{3}{2} + \binom{3}{1} \\
 &= \binom{4}{3} + \binom{3}{2} + \binom{2}{1} + \binom{2}{0} \\
 &= \binom{4}{3} + \binom{3}{2} + \binom{2}{1} + \binom{1}{0} + \binom{1}{-1}.
 \end{aligned}$$

Comme  $\binom{1}{-1} = 0$ , ce terme disparaît et on peut s'arrêter. En généralisant, on arrive à

$$\begin{aligned}
 \sum_{k \leq n} \binom{r+k}{k} &= \binom{r}{0} + \binom{r+1}{1} + \cdots + \binom{r+n}{n} \\
 &= \binom{r+n+1}{n}, \quad n \text{ entier.}
 \end{aligned} \tag{5.9}$$

Remarquez qu'on n'a pas besoin d'imposer  $k \geq 0$  pour l'indice de sommation, puisque les termes tels que  $k < 0$  sont nuls.

Cette formule donne l'expression d'un binomial comme une somme de binomiaux dont la différence des indices du haut et du bas est constante. Nous l'avons trouvée en développant, à chaque étape, le coefficient binomial qui avait le plus petit indice du bas : d'abord  $\binom{5}{3}$ , puis  $\binom{4}{2}$ , puis  $\binom{3}{1}$ , enfin  $\binom{2}{0}$ . Que se passe-t-il si on opère dans l'autre sens, c'est-à-dire en développant le binomial de plus grand indice du haut ? Voici ce qu'on obtient :

$$\begin{aligned}
 \binom{5}{3} &= \binom{4}{3} + \binom{4}{2} \\
 &= \binom{3}{3} + \binom{3}{2} + \binom{4}{2} \\
 &= \binom{2}{3} + \binom{2}{2} + \binom{3}{2} + \binom{4}{2} \\
 &= \binom{1}{3} + \binom{1}{2} + \binom{2}{2} + \binom{3}{2} + \binom{4}{2} \\
 &= \binom{0}{3} + \binom{0}{2} + \binom{1}{2} + \binom{2}{2} + \binom{3}{2} + \binom{4}{2}.
 \end{aligned}$$

Le binomial  $\binom{0}{3}$  est nul, donc on peut s'arrêter (en fait,  $\binom{0}{2}$  et  $\binom{1}{2}$  sont nuls aussi, mais l'identité est plus jolie si on les y laisse). Voici ce que cela donne

de façon générale :

$$\begin{aligned} \sum_{0 \leq k \leq n} \binom{k}{m} &= \binom{0}{m} + \binom{1}{m} + \cdots + \binom{n}{m} \\ &= \binom{n+1}{m+1}, \quad m, n \geq 0 \text{ entiers.} \end{aligned} \tag{5.10}$$

Cette identité, que nous appellerons *sommation sur l'indice du haut*, donne l'expression d'un binomial comme une somme de binomiaux dont l'indice du bas est constant. Dans ce cas, il est nécessaire de préciser que  $k \geq 0$  dans la somme, car les termes correspondant à  $k < 0$  ne sont pas nuls. En revanche,  $m$  et  $n$  peuvent être négatifs.

Voici une intéressante interprétation combinatoire de l'identité (5.10) : il y a  $\binom{k}{m}$  façons de choisir  $m+1$  tickets parmi  $n+1$  tickets numérotés de 0 à  $n$ , si on veut que le plus grand numéro de ticket choisi soit  $k$ .

Les formules (5.9) et (5.10) peuvent toutes deux être prouvées par induction en utilisant la formule d'addition. On peut aussi prouver chacune d'elles en utilisant l'autre. Démontrons par exemple (5.9) à partir de (5.10). Nous en profiterons pour passer en revue quelque manipulations classiques de coefficients binomiaux. Notre stratégie générale va être la suivante : modeler doucement le membre gauche  $\sum \binom{r+k}{k}$  de (5.9) jusqu'à ce qu'il ressemble au membre gauche  $\sum \binom{k}{m}$  de (5.10) ; puis invoquer cette identité pour remplacer la somme par un seul coefficient binomial ; enfin, transformer ce coefficient en membre droit de (5.9).

Nous pouvons supposer que  $r$  et  $n$  sont des entiers positifs ou nuls. L'argument polynomial nous permettra de généraliser sans problème. Remplaçons donc  $r$  par  $m$ , pour qu'il ait davantage l'air d'un entier positif ou nul. Appliquons alors notre plan :

$$\begin{aligned} \sum_{k \leq n} \binom{m+k}{k} &= \sum_{-m \leq k \leq n} \binom{m+k}{k} \\ &= \sum_{-m \leq k \leq n} \binom{m+k}{m} \\ &= \sum_{0 \leq k \leq m+n} \binom{k}{m} \\ &= \binom{m+n+1}{m+1} = \binom{m+n+1}{n}. \end{aligned}$$

Examinons ce calcul pas à pas. L'étape cruciale est celle de la seconde ligne, où on applique la règle de symétrie (5.4) pour remplacer  $\binom{m+k}{k}$  par  $\binom{m+k}{m}$ . On ne peut s'autoriser à le faire que si  $m+k \geq 0$  ; c'est pourquoi la première étape a consisté à supprimer les termes tels que  $k < -m$  (et c'est légal car

tous ces termes sont nuls). Nous sommes alors presque prêts à appliquer (5.10). La quatrième ligne achève notre préparation en remplaçant  $k$  par  $k-m$  et en mettant de l'ordre dans l'intervalle de sommation. Cette étape, tout comme la première, consiste simplement à titiller la notation  $\sum$  dans le bon sens. Maintenant,  $k$  apparaît tout seul dans l'indice du haut, et les limites de sommation sont sous la forme voulue pour pouvoir appliquer (5.10) dans la quatrième ligne. Une dernière symétrie pour finir, et voilà le travail.

Certaines sommes que nous avons calculées aux chapitres 1 et 2 étaient en fait des cas particuliers de l'identité (5.10). Par exemple, pour  $m=1$ , on retrouve la somme des entiers de 1 à  $n$  :

$$\binom{0}{1} + \binom{1}{1} + \cdots + \binom{n}{1} = 0 + 1 + \cdots + n = \frac{(n+1)n}{2} = \binom{n+1}{2}.$$

Nous avons aussi déjà démontré, au chapitre 2, un résultat équivalent à (5.10) :

$$\sum_{0 \leq k \leq n} k^m = \frac{(n+1)^{m+1}}{m+1}, \quad m, n \geq 0 \text{ entiers.}$$

Pour retrouver notre identité, il suffit de diviser les deux membres de la formule ci-dessus par  $m!$ . D'autre part, remarquons que la formule d'addition (5.8) nous dit que

$$\Delta \left( \binom{x}{m} \right) = \binom{x+1}{m} - \binom{x}{m} = \binom{x}{m-1},$$

si on remplace  $r$  et  $k$  respectivement par  $x+1$  et  $m$ . D'après ce que nous avons vu au chapitre 2, on peut en déduire une formule pratique de sommation indéfinie :

$$\sum \binom{x}{m} \delta x = \binom{x}{m+1} + C. \quad (5.11)$$

*"At the age of twenty-one he [Moriarty] wrote a treatise upon the Binomial Theorem, which has had a European vogue. On the strength of it, he won the Mathematical Chair at one of our smaller Universities."*

—S. Holmes [84]

Les coefficients binomiaux tirent leur nom de la *formule du binôme*, qui permet de calculer les puissances du binôme  $x+y$ . Regardons cette formule pour les premières puissances :

$$\begin{aligned} (x+y)^0 &= 1x^0y^0 \\ (x+y)^1 &= 1x^1y^0 + 1x^0y^1 \\ (x+y)^2 &= 1x^2y^0 + 2x^1y^1 + 1x^0y^2 \\ (x+y)^3 &= 1x^3y^0 + 3x^2y^1 + 3x^1y^2 + 1x^0y^3 \\ (x+y)^4 &= 1x^4y^0 + 4x^3y^1 + 6x^2y^2 + 4x^1y^3 + 1x^0y^4. \end{aligned}$$

Il n'est pas difficile de comprendre pourquoi ces coefficients sont les mêmes que ceux du triangle de Pascal : lorsqu'on développe le produit

$$(x+y)^n = \overbrace{(x+y)(x+y) \dots (x+y)}^{n \text{ facteurs}},$$

chacun des termes est un produit de  $n$  facteurs, chaque facteur étant soit  $x$ , soit  $y$ . Le coefficient de  $x^k y^{n-k}$  dans le développement est exactement le nombre de façons de choisir, parmi les  $n$  binômes, les  $k$  binômes dans lesquels  $x$  contribue au terme, soit  $\binom{n}{k}$ .

Dans certains ouvrages, on considère que la valeur  $0^0$  est indéfinie, du fait que les fonctions  $x^0$  et  $0^x$  ont des limites différentes lorsque  $x$  tend vers 0 en décroissant. C'est cependant un erreur : il faut définir

$$x^0 = 1, \quad \text{pour tout } x,$$

si on veut que la formule du binôme soit correcte lorsque  $x = 0$  et  $y = 0$ , et/ou lorsque  $x = -y$ . Cette formule est trop importante pour que l'on restreigne arbitrairement sa portée ! En comparaison, la fonction  $0^x$  est d'une importance négligeable. (Voir [220] pour d'autres précisions).

Voyons maintenant précisément ce que dit la formule du binôme. La voici dans toute sa gloire :

$$(x+y)^r = \sum_k \binom{r}{k} x^k y^{r-k}, \quad \begin{array}{l} r \geq 0 \text{ entier} \\ \text{ou } |x/y| < 1. \end{array} \quad (5.12)$$

La somme porte sur tous les entiers  $k$ , mais c'est en fait une somme finie si  $r$  est un entier positif ou nul, car tous les termes sont nuls sauf ceux pour lesquels  $0 \leq k \leq r$ . D'un autre côté, la formule est valable aussi lorsque  $r$  est négatif, ou même un quelconque nombre réel ou complexe. Dans ces cas-là, la somme est vraiment infinie, et il faut que  $|x/y| < 1$  pour garantir sa convergence absolue.

Voici deux cas particuliers de cette formule qui méritent toute notre attention, même s'ils sont très simples. Si  $x = y = 1$  et  $r = n$  est positif ou nul, on a

$$2^n = \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n}, \quad n \geq 0 \text{ entier.}$$

Ce que nous dit cette équation, c'est que la somme des éléments de la ligne  $n$  du triangle de Pascal vaut  $2^n$ . Voici l'autre cas particulier : si  $x$  vaut  $-1$  au lieu de 1, on obtient

$$0^n = \binom{n}{0} - \binom{n}{1} + \dots + (-1)^n \binom{n}{n}, \quad n \geq 0 \text{ entier.}$$

**Table 175** Le triangle de Pascal étendu vers le haut.

$n$	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
-4	1	-4	10	-20	35	-56	84	-120	165	-220	286
-3	1	-3	6	-10	15	-21	28	-36	45	-55	66
-2	1	-2	3	-4	5	-6	7	-8	9	-10	11
-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
0	1	0	0	0	0	0	0	0	0	0	0

Par exemple,  $1 - 4 + 6 - 4 + 1 = 0$ . La somme *alternée* des éléments de la ligne  $n$  du triangle de Pascal vaut 0, sauf pour la toute première, la ligne 0, pour laquelle on a  $0^0 = 1$ .

Lorsque  $r$  n'est pas un entier positif ou nul, on utilise très souvent la formule du binôme avec  $y = 1$ . Voyons ce cas particulier de plus près, en écrivant  $z$  à la place de  $x$  pour bien insister sur le fait que cela peut être un nombre complexe quelconque :

$$(1+z)^r = \sum_k \binom{r}{k} z^k, \quad |z| < 1. \quad (5.13)$$

La formule (5.12) se déduit de celle-ci en posant  $z = x/y$  et en multipliant les deux membres par  $y^r$ .

Jusqu'à présent, nous n'avons démontré la formule du binôme que lorsque  $r$  est un entier positif ou nul. De plus, comme la somme est infinie dans le cas général, l'argument polynomial ne nous est d'aucune utilité pour généraliser à  $r$  quelconque. Toutefois, il nous est permis d'utiliser les développements de Taylor et la théorie des variables complexes :

$$\begin{aligned} f(z) &= \frac{f(0)}{0!} z^0 + \frac{f'(0)}{1!} z^1 + \frac{f''(0)}{2!} z^2 + \dots \\ &= \sum_{k \geq 0} \frac{f^{(k)}(0)}{k!} z^k. \end{aligned}$$

Les dérivées de la fonction  $f(z) = (1+z)^r$  sont faciles à calculer :  $f^{(k)}(z) = r^k (1+z)^{r-k}$ . En posant  $z = 0$ , on trouve (5.13).

Il nous faut aussi prouver que la somme infinie converge lorsque  $|z| < 1$ . Elle converge en effet, car  $\binom{r}{k} = O(k^{-1-r})$  d'après l'équation (5.83) ci-dessous.

Maintenant, examinons un peu les valeurs de  $\binom{n}{k}$  lorsque  $n$  est un entier négatif. On peut les obtenir avec la formule d'addition (5.8), en remplaçant les cases au dessus de celles de la table 166. On obtient ainsi la nouvelle

table 175. Par exemple,  $\binom{-1}{0} = 1$  car  $\binom{0}{0} = \binom{-1}{0} + \binom{-1}{-1}$  et  $\binom{-1}{-1} = 0$  ; alors on a  $\binom{-1}{1} = -1$ , puisque  $\binom{0}{1} = \binom{-1}{1} + \binom{-1}{0}$  ; etc.

Ces nombres ne nous sont pas inconnus : si on fait abstraction des signes moins, les lignes et les colonnes de la table 175 sont identiques aux colonnes de la table 166. Il y a donc forcément un lien entre les valeurs de  $\binom{n}{k}$  pour  $n$  négatif et pour  $n$  positif. Le voici :

$$\binom{r}{k} = (-1)^k \binom{k-r-1}{k}, \quad k \text{ entier.} \quad (5.14)$$

Cette formule est facile à prouver, car

$$\begin{aligned} r^k &= r(r-1)\dots(r-k+1) \\ &= (-1)^k(-r)(1-r)\dots(k-1-r) = (-1)^k(k-r-1)^k \end{aligned}$$

lorsque  $k \geq 0$ , et les deux membres sont nuls lorsque  $k < 0$ .

L'identité (5.14) est particulièrement intéressante car elle s'applique sans aucune restriction (sauf que, bien entendu, l'indice du bas doit être un entier pour que le binomial soit défini). Nous appellerons la transformation effectuée dans (5.14) le *changement de signe de l'indice du haut*.

Maintenant, le problème est de savoir comment nous allons bien pouvoir mémoriser cette formule. Elle est hélas bien moins simple que celles que nous avons vues jusqu'à présent : symétrie, absorption, addition etc. Voici toutefois un assez bon moyen mnémotechnique : pour effectuer la négation de l'indice du haut, on commence par écrire  $(-1)^k$ , où  $k$  est l'indice du bas ; puis on écrit encore  $k$ , deux fois, comme indice du haut et indice du bas du nouveau binomial ; puis on soustrait l'ancien index du haut du nouveau ; enfin, on lui soustrait encore 1 (bien sûr on ne fait que soustraire, puisqu'il s'agit d'un changement de signe).

Pour nous entraîner un peu, voyons ce que donnent deux changements de signe à la suite :

$$\begin{aligned} \binom{r}{k} &= (-1)^k \binom{k-r-1}{k} \\ &= (-1)^{2k} \binom{k-(k-r-1)-1}{k} = \binom{r}{k}, \end{aligned}$$

et nous sommes donc revenus au point de départ. Ce n'est peut-être pas la meilleure façon d'utiliser cette identité, mais c'est toujours rassurant de voir qu'on a fait un calcul correct.

On peut heureusement appliquer (5.14) plus utilement, par exemple pour permute des éléments du haut et du bas d'un binomial. Pour cela,

*Vous dites que c'est mnémotechnique ? Je dirais plutôt pneumatique. Du vent, quoi. Malgré tout, je dois avouer que ça m'aide à m'en souvenir.*

*(C'est le moment de faire l'exercice d'échauffement 4).*

*C'est aussi frustrant, si on veut aller ailleurs qu'à son point de départ.*

on utilise une expression symétrique de notre identité :

$$(-1)^m \binom{-n-1}{m} = (-1)^n \binom{-m-1}{n}, \quad m, n \geq 0 \text{ entiers}, \quad (5.15)$$

car les deux membres sont égaux à  $\binom{m+n}{n}$ .

On peut aussi l'utiliser pour calculer l'intéressante somme que voici :

$$\begin{aligned} \sum_{k \leq m} \binom{r}{k} (-1)^k &= \binom{r}{0} - \binom{r}{1} + \cdots + (-1)^m \binom{r}{m} \\ &= (-1)^m \binom{r-1}{m}, \quad m \text{ entier}. \end{aligned} \quad (5.16)$$

Pour arriver à cela, on change le signe de l'indice du haut, puis on applique (5.9), enfin on refait un changement de signe :

$$\begin{aligned} \sum_{k \leq m} \binom{r}{k} (-1)^k &= \sum_{k \leq m} \binom{k-r-1}{k} \\ &= \binom{-r+m}{m} \\ &= (-1)^m \binom{r-1}{m}. \end{aligned}$$

*(Maintenant le double changement de signe sert à quelque chose, parce qu'on en a fait un sandwich avec une autre opération à l'intérieur).*

Avec cette formule, on peut calculer une somme partielle alternée de la  $r$ ième ligne du triangle de Pascal. Par exemple, pour  $r = 5$  et  $m = 2$ , on trouve  $1 - 5 + 10 = 6 = (-1)^2 \binom{4}{2}$ .

Remarquez que si  $m \geq r$ , (5.16) représente la somme alternée d'une ligne entière, qui est nulle lorsque  $r$  est un entier strictement positif. Nous l'avons déjà démontré en développant  $(1-1)^r$  à l'aide de la formule du binôme. Il est intéressant de savoir que les sommes partielles de cette expression ont aussi une forme close.

Voyons maintenant la somme

$$\sum_{k \leq m} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{m}, \quad (5.17)$$

qui est plus simple que la précédente puisqu'elle n'est pas alternée. On doit donc bien pouvoir la calculer. Eh bien non. Il n'existe pas de forme close pour la somme partielle d'une ligne du triangle de Pascal. Les colonnes oui (c'est (5.10)), les lignes non. Curieusement, on peut quand même sommer partiellement les éléments d'une ligne, pourvu que chacun ait été multiplié

par sa distance au centre :

$$\sum_{k \leq m} \binom{r}{k} \left( \frac{r}{2} - k \right) = \frac{m+1}{2} \binom{r}{m+1}, \quad m \text{ entier.} \quad (5.18)$$

On démontre facilement cette formule par induction sur  $m$ . La relation entre ces deux sommes partielles (celle avec et celle sans le facteur  $(r/2 - k)$  dans la somme) est tout à fait analogue à la relation entre les intégrales

$$\int_{-\infty}^{\alpha} xe^{-x^2} dx = -\frac{1}{2} e^{-\alpha^2} \quad \text{et} \quad \int_{-\infty}^{\alpha} e^{-x^2} dx.$$

L'intégrale de gauche a une forme close, contrairement à celle de droite qui semble pourtant plus simple puisqu'elle n'a pas le facteur  $x$ . Ne nous fions pas aux apparences.

(En fait, l'intégrale de droite est égale à  $\frac{1}{2}\sqrt{\pi}(1 + \operatorname{erf} \alpha)$ , soit une constante plus un multiple de la valeur de la "fonction d'erreur" en  $\alpha$ . On pourrait considérer que c'est une formule close).

Vers la fin de ce chapitre, nous étudierons une méthode générique pour déterminer, étant donnée une série mettant en œuvre des coefficients binomiaux, s'il existe ou non une forme close pour la somme partielle de cette série. Cette méthode permet notamment de trouver les identités (5.16) et (5.18), et de savoir à coup sûr que (5.17) n'a pas de forme close.

Voici maintenant une curieuse relation entre des sommes partielles de séries binomiales :

$$\sum_{k \leq m} \binom{m+r}{k} x^k y^{m-k} = \sum_{k \leq m} \binom{-r}{k} (-x)^k (x+y)^{m-k}, \quad m \text{ entier.} \quad (5.19)$$

On démontre facilement cette identité par induction : les deux membres sont nuls lorsque  $m < 0$  et égaux à 1 lorsque  $m = 0$ . Appelons  $S_m$  la somme de gauche, et appliquons la formule d'addition (5.8) :

$$S_m = \sum_{k \leq m} \binom{m-1+r}{k} x^k y^{m-k} + \sum_{k \leq m} \binom{m-1+r}{k-1} x^k y^{m-k};$$

et comme on a

$$\sum_{k \leq m} \binom{m-1+r}{k} x^k y^{m-k} = y S_{m-1} + \binom{m-1+r}{m} x^m,$$

$$\sum_{k \leq m} \binom{m-1+r}{k-1} x^k y^{m-k} = x S_{m-1},$$

lorsque  $m > 0$ , on trouve

$$S_m = (x+y) S_{m-1} + \binom{-r}{m} (-x)^m.$$

Cette récurrence est satisfaite aussi par le membre droit de (5.19). On en déduit, par induction, que les deux membres sont égaux. CQFD.

On peut aussi faire une preuve plus concise. Si  $r$  est un entier tel que  $0 \geq r \geq -m$ , alors, d'après la formule du binôme, les deux membres de (5.19) sont égaux à  $(x + y)^{m+r}y^{-r}$ . Comme ces deux membres sont des polynômes en  $r$  de degré inférieur ou égal à  $m$ , il suffit qu'ils soient égaux en  $m + 1$  points pour qu'ils soient égaux en tout point.

A première vue, il ne semble pas très intéressant d'avoir une égalité entre deux sommes alors qu'aucune des deux n'est en forme close. Il arrive pourtant que l'une des deux soit plus facile à calculer que l'autre. Par exemple, si on prend  $x = -1$  et  $y = 1$ , on obtient une variante de l'identité (5.16) :

$$\sum_{k \leq m} \binom{m+r}{k} (-1)^k = \binom{-r}{m}, \quad m \geq 0 \text{ entier.}$$

Si on pose  $x = y = 1$  et  $r = m + 1$ , on trouve

$$\sum_{k \leq m} \binom{2m+1}{k} = \sum_{k \leq m} \binom{m+k}{k} 2^{m-k}.$$

Le membre gauche représente la somme de la moitié exactement des binomiaux dont l'indice du haut est  $2m + 1$ . Comme chaque ligne du triangle de Pascal est symétrique par rapport à son milieu, ce membre gauche est égal à  $\frac{1}{2}2^{2m+1} = 2^{2m}$ . On aboutit à une somme tout à fait inattendue :

$$\sum_{k \leq m} \binom{m+k}{k} 2^{-k} = 2^m, \quad m \geq 0 \text{ entier.} \quad (5.20)$$

Vérifions la pour  $m = 2$  :  $\binom{2}{0} + \frac{1}{2}\binom{3}{1} + \frac{1}{4}\binom{4}{2} = 1 + \frac{3}{2} + \frac{6}{4} = 4$ . Stupéfiant !

Jusqu'à présent, nous n'avons manipulé que des coefficients binomiaux isolés ou des sommes ne contenant qu'un binomial par terme. Or, nous allons bien sûr rencontrer tôt ou tard des problèmes dans lesquels il y aura des produits de plusieurs binomiaux. Nous allons donc passer le reste de cette section à étudier ce genre de cas de figure.

Voici une règle pratique pour simplifier le produit de deux binomiaux :

$$\binom{r}{m} \binom{m}{k} = \binom{r}{k} \binom{r-k}{m-k}, \quad m, k \text{ entiers.} \quad (5.21)$$

Nous connaissons déjà le cas  $k = 1$  : il s'agit de l'identité d'absorption (5.6). Bien que les deux membres de (5.21) soient des produits de binomiaux, il arrive souvent que l'un des deux soit plus facile à sommer que l'autre. En effet, son interaction avec le reste de la somme peut donner lieu à des

(On trouvera  
en [247] une jolie  
preuve combinatoire  
de cette formule).

simplifications. Remarquons par exemple que le membre gauche contient deux occurrences de  $m$  tandis que le membre droit n'en contient qu'une. Par conséquent, si on somme sur  $m$ , il vaut mieux remplacer  $\binom{r}{m} \binom{m}{k}$  par  $\binom{r}{k} \binom{r-k}{m-k}$ .

Si l'équation (5.21) marche, c'est principalement parce que les factorielles de  $m$  dans les binomiaux  $\binom{r}{m}$  et  $\binom{m}{k}$  se neutralisent mutuellement. En effet, si toutes les variables sont entières et si  $r \geq m \geq k \geq 0$ , on a

$$\begin{aligned}\binom{r}{m} \binom{m}{k} &= \frac{r!}{m! (r-m)!} \frac{m!}{k! (m-k)!} \\&= \frac{r!}{k! (m-k)! (r-m)!} \\&= \frac{r!}{k! (r-k)!} \frac{(r-k)!}{(m-k)! (r-m)!} \\&= \binom{r}{k} \binom{r-k}{m-k}.\end{aligned}$$

Facile. D'autre part, si  $m < k$  ou  $k < 0$ , les deux membres de (5.21) sont nuls. L'identité est donc vraie pour  $m$  and  $k$  entiers quelconques. Il ne reste plus qu'à invoquer l'argument polynomial pour l'étendre à tout  $r$ . *Ouais, super.*

Si on fait les changements de variables adéquats, tout binomial  $\binom{r}{k} = r!/(r-k)! k!$  peut s'écrire sous la forme  $(a+b)!/a! b!$ . De même, l'expression de la deuxième ligne du calcul ci-dessus,  $r!/k! (m-k)! (r-m)!$ , peut s'écrire  $(a + b + c)!/a! b! c!$ . On appelle ceci un "coefficient trinomial", ou tout simplement un trinomial. Et voici la formule du trinôme qui va avec :

$$\begin{aligned}(x+y+z)^n &= \sum_{\substack{0 \leq a, b, c \leq n \\ a+b+c=n}} \frac{(a+b+c)!}{a! b! c!} x^a y^b z^c \\&= \sum_{\substack{0 \leq a, b, c \leq n \\ a+b+c=n}} \binom{a+b+c}{b+c} \binom{b+c}{c} x^a y^b z^c.\end{aligned}$$

Ainsi,  $\binom{r}{m} \binom{m}{k}$  est en fait un coefficient trinomial déguisé en produit de binomiaux. Les trinomiaux s'écrivent avec la notation suivante,

$$\binom{a+b+c}{a, b, c} = \frac{(a+b+c)!}{a! b! c!}$$

qui offre l'avantage de mettre en évidence leur symétrie.

Les coefficients binomiaux et trinomiaux se généralisent en *coefficients multinomiaux*, qui peuvent toujours s'exprimer par un produit de coeffi-

"Excogitavi autem olim mirabilem regulam pro numeris coefficientibus potestatum, non tantum a binomio  $x+y$ , sed et a trinomio  $x+y+z$ , immo a polynomio quocunque, ut data potentia gradus cuiuscunq; v. gr. decimi, et potentia in ejus valore comprehensa, ut  $x^5 y^3 z^2$ , possim statim assignare numerum coefficientem, quem habere debet, sine ulla Tabula jam calculata."

— G. W. Leibniz [245]

**Table 181** Sommes de produits de coefficients binomiaux.

$$\sum_k \binom{r}{m+k} \binom{s}{n-k} = \binom{r+s}{m+n}, \quad m, n \text{ entiers. (5.22)}$$

$$\sum_k \binom{l}{m+k} \binom{s}{n+k} = \binom{l+s}{l-m+n}, \quad l \geq 0 \text{ entier, } m, n \text{ entiers. (5.23)}$$

$$\sum_k \binom{l}{m+k} \binom{s+k}{n} (-1)^k = (-1)^{l+m} \binom{s-m}{n-l}, \quad l \geq 0 \text{ entier, } m, n \text{ entiers. (5.24)}$$

$$\sum_{k \leq l} \binom{l-k}{m} \binom{s}{k-n} (-1)^k = (-1)^{l+m} \binom{s-m-1}{l-m-n}, \quad l, m, n \geq 0 \text{ entiers. (5.25)}$$

$$\sum_{0 \leq k \leq l} \binom{l-k}{m} \binom{q+k}{n} = \binom{l+q+1}{m+n+1}, \quad l, m \geq 0 \text{ entiers, } n \geq q \geq 0 \text{ entiers. (5.26)}$$

cients binomiaux :

$$\begin{aligned} \binom{a_1 + a_2 + \cdots + a_m}{a_1, a_2, \dots, a_m} &= \frac{(a_1 + a_2 + \cdots + a_m)!}{a_1! a_2! \dots a_m!} \\ &= \binom{a_1 + a_2 + \cdots + a_m}{a_2 + \cdots + a_m} \cdots \binom{a_{m-1} + a_m}{a_m}. \end{aligned}$$

Par conséquent, quand nous rentrerons une bête de cette espèce, nous pourrons appliquer nos techniques standard.

Justement, la table 181 recense quelques identités qui sont parmi les plus importantes de nos techniques standard. Ce sont celles qu'on utilise quand on est confronté à une somme contenant des produits de coefficients binomiaux. Dans chacune de ces identités, l'indice de sommation  $k$  apparaît dans les deux binomiaux ; il y a aussi quatre paramètres indépendants ou presque indépendants, notés  $m, n, r$  etc, chacun occupant une place différente. Plusieurs cas sont considérés, selon que  $k$  apparaît dans l'indice du haut ou l'indice du bas, avec un signe plus ou un signe moins. Pour que la somme puisse avoir une forme close, on a ajouté, dans deux cas, un facteur  $(-1)^k$ .

N'essayez pas de retenir la table 181, elle est bien trop compliquée pour que cela en vaille la peine. Elle nous servira simplement de référence. Toutefois, la première identité, qui est de loin la plus facile à mémoriser, vaut la peine d'être retenue. Elle dit que la somme (sur tous les entiers  $k$ ) du produit de deux binomiaux dont les indices du haut sont constants et

dont la somme des indices du bas est constante est égale au binomial obtenu en additionnant respectivement les indices du haut et les indices du bas. Cette identité doit son nom de *convolution de Vandermonde* au fait qu'Alexandre Vandermonde a écrit, à la fin du dix-huitième siècle, un important article la concernant [357]. Il s'avère toutefois qu'un Chinois, Chu Shih-Chieh, la connaissait bien avant, en 1303. Toutes les autres identités de la table 181 peuvent être obtenues à partir de la convolution de Vandermonde en appliquant avec soin quelques opérations comme le changement de signe de l'indice du haut, la règle de symétrie etc. La convolution de Vandermonde est donc la base de toute la table.

On peut démontrer cette convolution avec une jolie interprétation combinatoire. Si on remplace  $k$  par  $k - m$  et  $n$  par  $n - m$ , on peut supposer que  $m = 0$ . Il nous faut donc prouver que

$$\sum_k \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}, \quad n \text{ entier.} \quad (5.27)$$

Supposons que  $r$  et  $s$  sont deux entiers positifs ou nuls. L'argument polynomial nous permettra de généraliser. Dans le membre droit,  $\binom{r+s}{n}$  représente le nombre de façons de choisir  $n$  personnes parmi  $r$  hommes et  $s$  femmes. Dans le membre gauche, chaque terme de la somme est égal au nombre de façons de choisir  $k$  hommes et  $n - k$  femmes. En sommant sur  $k$ , on compte chaque possibilité une fois exactement.

*Phallocrates ! Vous mentionnez les hommes d'abord.*

Dans la grande majorité des cas, on utilise bien sûr ces identités de gauche à droite, dans un but de simplification. Il arrive cependant qu'on prenne l'autre direction pour obtenir, du moins temporairement, une expression plus compliquée. Par exemple, cela peut être fait pour créer une somme double dans laquelle on peut changer l'ordre de sommation puis simplifier.

Voyons maintenant les preuves des identités de la table 181. Il n'est pas difficile de prouver l'équation (5.23) : il suffit de remplacer le premier binomial par  $\binom{l}{l-m-k}$ , puis d'appliquer la convolution de Vandermonde (5.22).

La suivante, (5.24), est un peu plus difficile. En appliquant une suite de transformations, on peut se ramener à la convolution de Vandermonde. On peut tout aussi bien prendre la bonne vieille méthode de l'induction. C'est souvent la première chose à faire si aucune autre idée ne nous saute aux yeux. Il se trouve que l'induction sur  $l$  marche très bien dans notre cas.

Pour la base  $l = 0$ , tous les termes sont nuls sauf pour  $k = -m$ . Les deux membres de l'équation valent donc  $(-1)^m \binom{s-m}{n}$ . Supposons maintenant que l'identité est vraie pour toutes les valeurs strictement inférieures à un certain  $l > 0$  donné. Utilisons la formule d'addition pour remplacer  $\binom{l}{m+k}$  par  $\binom{l-1}{m+k} + \binom{l-1}{m+k-1}$ . La somme de départ se sépare en deux, et cha-

cune des nouvelles sommes peut être calculée avec l'hypothèse d'induction :

$$\begin{aligned} & \sum_k \binom{l-1}{m+k} \binom{s+k}{n} (-1)^k + \sum_k \binom{l-1}{m+k-1} \binom{s+k}{n} (-1)^k \\ &= (-1)^{l-1+m} \binom{s-m}{n-l+1} + (-1)^{l+m} \binom{s-m+1}{n-l+1}. \end{aligned}$$

Si on applique encore une fois la formule d'addition, tout ceci se simplifie pour donner le membre droit de (5.24).

Voici deux remarques importantes sur le calcul que nous venons de faire. Premièrement, nous voyons encore une fois l'intérêt de sommer sur tous les entiers  $k$  au lieu de se restreindre à un intervalle : ainsi on ne se pose pas de problèmes sur les conditions aux bornes. Deuxièmement, la formule d'addition se marie particulièrement bien avec l'induction mathématique, car c'est une récurrence sur les coefficients binomiaux : un binomial dont l'indice du haut est  $l$  s'exprime en fonction de deux binomiaux dont l'indice du haut est  $l$ , et c'est exactement ce dont on a besoin pour l'induction.

Voilà pour la table 181. Passons maintenant aux sommes de produits de trois binomiaux ou plus. Le problème, c'est que si l'indice de sommation apparaît dans tous les coefficients binomiaux, les chances de trouver une forme close sont minces. On connaît en effet très peu de sommes de ce type qui ont une forme close. Voici l'une de ces raretés, que l'on démontre dans l'exercice 43 :

$$\begin{aligned} & \sum_k \binom{m-r+s}{k} \binom{n+r-s}{n-k} \binom{r+k}{m+n} \\ &= \binom{r}{m} \binom{s}{n}, \quad m, n \geq 0 \text{ entiers.} \end{aligned} \tag{5.28}$$

En voici une autre plus symétrique :

$$\begin{aligned} & \sum_k \binom{a+b}{a+k} \binom{b+c}{b+k} \binom{c+a}{c+k} (-1)^k \\ &= \frac{(a+b+c)!}{a! b! c!}, \quad a, b, c \geq 0 \text{ entiers.} \end{aligned} \tag{5.29}$$

On connaît même une somme à deux coefficients qui lui est similaire,

$$\sum_k \binom{a+b}{a+k} \binom{b+a}{b+k} (-1)^k = \frac{(a+b)!}{a! b!}, \quad a, b \geq 0 \text{ entiers.} \tag{5.30}$$

et qui n'apparaît pas dans la table 181. La somme analogue à quatre coef-

ficients n'a pas de forme close, mais en voici une qui lui ressemble :

$$\sum_k (-1)^k \binom{a+b}{a+k} \binom{b+c}{b+k} \binom{c+d}{c+k} \binom{d+a}{d+k} / \binom{2a+2b+2c+2d}{a+b+c+d+k}$$

$$= \frac{(a+b+c+d)! (a+b+c)! (a+b+d)! (a+c+d)! (b+c+d)!}{(2a+2b+2c+2d)! (a+c)! (b+d)! a! b! c! d!},$$

$a, b, c, d \geq 0$  entiers.

Elle se déduit d'une identité à cinq paramètres découverte par John Dougall [82] au début du vingtième siècle. Voici maintenant la championne de ces identités, dans la catégorie "complexité" :

$$\sum_{k_{ij}} (-1)^{\sum_{i < j} k_{ij}} \left( \prod_{1 \leq i < j < n} \binom{a_i + a_j}{a_j + k_{ij}} \right) \left( \prod_{1 \leq j < n} \binom{a_j + a_n}{a_n + \sum_{i < j} k_{ij} - \sum_{i > j} k_{ji}} \right)$$

$$= \binom{a_1 + \cdots + a_n}{a_1, a_2, \dots, a_n}, \quad a_1, a_2, \dots, a_n \geq 0 \text{ entiers.} \quad (5.31)$$

Cette somme porte sur  $\binom{n-1}{2}$  variables d'indice  $k_{ij}$  pour  $1 \leq i < j < n$ . L'équation (5.29) est en fait le cas particulier pour  $n = 3$ , et voici le cas  $n = 4$ , en écrivant  $(a, b, c, d)$  pour  $(a_1, a_2, a_3, a_4)$  et  $(i, j, k)$  pour  $(k_{12}, k_{13}, k_{23})$ :

$$\sum_{i,j,k} (-1)^{i+j+k} \binom{a+b}{b+i} \binom{a+c}{c+j} \binom{b+c}{c+k} \binom{a+d}{d-i-j} \binom{b+d}{d+i-k} \binom{c+d}{d+j+k}$$

$$= \frac{(a+b+c+d)!}{a! b! c! d!}, \quad a, b, c, d \geq 0 \text{ entiers.}$$

Le membre gauche de (5.31) est le coefficient de  $z_1^0 z_2^0 \dots z_n^0$  dans le produit de  $n(n-1)$  fractions

$$\prod_{\substack{1 \leq i, j \leq n \\ i \neq j}} \left( 1 - \frac{z_i}{z_j} \right)^{a_i},$$

une fois qu'il a été développé en puissances positives et négatives des  $z$ . La formule du membre droit a été conjecturée par Freeman Dyson en 1962, puis prouvée peu de temps après par plusieurs personnes. L'exercice 89 fournit une preuve "simple" de (5.31). L'identité suivante, qu'on démontre dans l'exercice 83, vaut aussi le détour :

$$\sum_{j,k} (-1)^{j+k} \binom{j+k}{k+l} \binom{r}{j} \binom{n}{k} \binom{s+n-j-k}{m-j}$$

$$= (-1)^l \binom{n+r}{n+l} \binom{s-r}{m-n-l}, \quad l, m, n \text{ entiers, } n \geq 0. \quad (5.32)$$

Tout ceci nous éloigne des “identités de base” qui sont le thème de cette section. Arrêtons-nous donc pour faire le point sur ce que nous avons appris jusqu’ici. Nous avons vu que les coefficients binomiaux satisfont un très grand nombre d’identités. Par bonheur, un certain nombre d’entre elles sont faciles à retenir et permettent de retrouver la plupart des autres. La table 186 présente les dix identités les plus utiles, celles qu’il est bon de connaître par cœur.

## 5.2 PRATIQUE DE BASE

Les identités de la section précédente n’étaient pas vraiment difficiles à démontrer. En effet nous savions ce que nous devions prouver ; c’est ce qui nous permettait de concevoir un plan pour chaque preuve et de le suivre sans beaucoup de difficultés. En revanche, dans le monde réel, le problème n’est pas de démontrer des identités connues ; il s’agit de simplifier des sommes, sans savoir où il faut arriver, sans même savoir si ces sommes peuvent être simplifiées. Nous allons, dans cette section et la suivante, nous frotter à des problèmes de ce genre. Voyons pour commencer quelques sommes où n’apparaît qu’un coefficient binomial.

**Problème 1 : une somme de quotients.**

Nous aimerais trouver une forme close pour

$$\sum_{k=0}^m \binom{m}{k} / \binom{n}{k}, \quad n \geq m \geq 0 \text{ entiers.}$$

A première vue, cette somme peut sembler franchement effrayante ; d’abord parce que nous n’avons jamais vu d’identité avec des quotients de binomiaux ; ensuite parce qu’elle contient deux binomiaux, contrairement à ce que nous déclarions dans la phrase précédent le problème. Ne fuyez pas cependant. On peut écrire un quotient de binomiaux en un autre quotient de binomiaux, exactement comme pour un produit. C’est possible par exemple en écrivant les binomiaux sous forme de factorielles, comme nous l’avons fait pour trouver l’identité (5.21). Nous pouvons même éviter ce passage par les factorielles en posant  $r = n$  dans (5.21) et en divisant les deux membres par  $\binom{n}{k} \binom{n}{m}$ . On obtient ainsi

$$\binom{m}{k} / \binom{n}{k} = \binom{n-k}{m-k} / \binom{n}{m}.$$

Remplaçons dans notre somme le quotient de gauche par celui de droite :

$$\sum_{k=0}^m \binom{n-k}{m-k} / \binom{n}{m}.$$

*Algorithme  
auto-apprentissage:*  
1 lire problème  
2 trouver solution  
3 lire celle du livre  
4 si pas égales  
    goto 1  
    sinon goto pro-  
        blème suivant

*Malheureusement,  
cet algorithme peut  
boucler indéfiniment.*  
*Lignes à ajouter :*  
0     $c \leftarrow 0$   
3a     $c \leftarrow c + 1$   
3b    si  $c = N$   
      goto votre  
charge de TD



— E. W. Dijkstra

... Evidemment.  
Cette section  
s’appelle “pratique  
de base”, pas “pra-  
tique du basic”.

**Table 186** Les dix principales identités sur les coefficients binomiaux.

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ,	$n \geq k \geq 0$ entiers	développement factoriel
$\binom{n}{k} = \binom{n}{n-k}$ ,	$n \geq 0$ entier, $k$ entier.	symétrie
$\binom{r}{k} = \frac{r}{k} \binom{r-1}{k-1}$ ,	$k \neq 0$ entier.	absorption/extraction
$\binom{r}{k} = \binom{r-1}{k} + \binom{r-1}{k-1}$ ,	$k$ entier.	addition/induction
$\binom{r}{k} = (-1)^k \binom{k-r-1}{k}$ ,	$k$ entier.	changement de signe
$\binom{r}{m} \binom{m}{k} = \binom{r}{k} \binom{r-k}{m-k}$ ,	$m, k$ entiers.	transformation trinomiale
$\sum_k \binom{r}{k} x^k y^{r-k} = (x+y)^r$ ,	$r \geq 0$ entier ou $ x/y  < 1$ .	formule du binôme
$\sum_{k \leq n} \binom{r+k}{k} = \binom{r+n+1}{n}$ ,	$n$ entier.	sommation parallèle
$\sum_{0 \leq k \leq n} \binom{k}{m} = \binom{n+1}{m+1}$ ,	$m, n \geq 0$ entiers.	sommation du haut
$\sum_k \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}$ ,	$n$ entier.	convolution de Vandermonde

Nous avons toujours un quotient, mais le dénominateur ne dépend pas de l'indice de sommation  $k$ . Nous pouvons donc sortir momentanément ce dénominateur de la somme.

Nous pouvons aussi simplifier les conditions aux bornes en sommant pour tout  $k \geq 0$ , puisque les termes pour lesquels  $k > m$  sont nuls. La somme que nous obtenons n'est plus si terrible :

$$\sum_{k \geq 0} \binom{n-k}{m-k}.$$

Elle ressemble à celle de l'identité (5.9), car dans les deux l'indice  $k$  apparaît deux fois avec le même signe. Toutefois, ici on a un  $-k$  et là-bas (dans (5.9)) un  $+k$ . L'étape suivante est donc évidente ; il n'y a raisonnablement qu'une

seule chose à faire :

$$\begin{aligned}\sum_{k \geq 0} \binom{n-k}{m-k} &= \sum_{m-k \geq 0} \binom{n-(m-k)}{m-(m-k)} \\ &= \sum_{k \leq m} \binom{n-m+k}{k}.\end{aligned}$$

Nous pouvons maintenant appliquer l'identité de sommation parallèle (5.9) :

$$\sum_{k \leq m} \binom{n-m+k}{k} = \binom{(n-m)+m+1}{m} = \binom{n+1}{m}.$$

Pour finir, remettons le  $\binom{n}{m}$  que nous avons supprimé dans le dénominateur, puis appliquons (5.7) pour trouver la forme close désirée :

$$\binom{n+1}{m} / \binom{n}{m} = \frac{n+1}{n+1-m}.$$

Cette formule est valable pour tout  $n$  réel, à condition qu'il n'y ait pas de division par zéro, c'est-à-dire que  $n$  ne soit égal à aucun des entiers  $0, 1, \dots, m-1$ .

Plus un calcul est compliqué, plus il est important de vérifier que la réponse est bonne. Celui que nous venons de faire n'était pas trop compliqué, mais nous allons quand même regarder si tout va bien pour  $m=2$  et  $n=4$  :

$$\binom{2}{0} / \binom{4}{0} + \binom{2}{1} / \binom{4}{1} + \binom{2}{2} / \binom{4}{2} = 1 + \frac{1}{2} + \frac{1}{6} = \frac{5}{3}.$$

Cela correspond parfaitement à notre forme close  $(4+1)/(4+1-2)$ .

### *Problème 2 : d'après un problème de tri.*

La somme que nous allons étudier ici est apparue il y a bien longtemps (au début des années 1970), avant que les gens aient l'habitude de manipuler les coefficients binomiaux. Un article introduisant une nouvelle méthode de fusion de fichiers [196] se termine par la remarque suivante : "On peut montrer que le nombre moyen d'écritures est donné par l'expression

$$T = \sum_{r=0}^n r \frac{\binom{m-r-1}{m-1} C_{m-n-1}}{C_m},$$

où  $m$  et  $n$  sont définis comme précédemment, et  $_m C_n$  désigne le nombre de combinaisons de  $m$  objets parmi  $n$ . [...] L'auteur remercie l'arbitre qui lui a suggéré de remplacer l'équation compliquée écrite initialement par cette expression plus simple."

Nous allons voir qu'il existe une solution bien plus simple encore au problème de l'auteur. En fait, la question de trouver la solution la plus simple serait même trop facile pour un partiel.

Commençons par traduire cette somme dans notre langage habituel, car cette notation  $m-r-1 C_{m-n-1}$  est horrible à voir. Seul notre enthousiaste arbitre semble pouvoir la supporter. Ecrivons donc

$$T = \sum_{k=0}^n k \binom{m-k-1}{m-n-1} / \binom{m}{n}, \quad m > n \geq 0 \text{ entiers.}$$

Comme l'indice de sommation n'apparaît pas dans binomial du dénominateur, nous pouvons supprimer temporairement ce dernier pour travailler avec la somme

$$S = \sum_{k=0}^n k \binom{m-k-1}{m-n-1}.$$

Que faire maintenant ? Remarquons que l'indice de sommation n'apparaît dans le binomial que dans l'indice du haut. S'il n'y avait pas cet autre k en facteur, nous pourrions certainement nous arranger pour utiliser la règle de sommation de l'indice du haut (5.10). Il faudrait pouvoir utiliser une identité d'absorption pour faire avaler ce k au coefficient binomial ; alors nous pourrions sommer sur l'indice du haut. Hélas, les identités d'absorption ne marchent pas ici. Par contre, si k était remplacé par  $n - k$ , l'identité d'absorption (5.6) pourrait s'appliquer :

$$(m-k) \binom{m-k-1}{m-n-1} = (m-n) \binom{m-k}{m-n}.$$

Ça y est, nous avons trouvé. Nous allons réécrire k en  $m - (m - k)$  et séparer la somme en deux :

$$\begin{aligned} \sum_{k=0}^n k \binom{m-k-1}{m-n-1} &= \sum_{k=0}^n (m - (m - k)) \binom{m-k-1}{m-n-1} \\ &= \sum_{k=0}^n m \binom{m-k-1}{m-n-1} - \sum_{k=0}^n (m-k) \binom{m-k-1}{m-n-1} \\ &= m \sum_{k=0}^n \binom{m-k-1}{m-n-1} - \sum_{k=0}^n (m-n) \binom{m-k}{m-n} \\ &= mA - (m-n)B, \end{aligned}$$

*Pitié, ne me parlez pas du partiel.*

avec

$$A = \sum_{k=0}^n \binom{m-k-1}{m-n-1}, \quad B = \sum_{k=0}^n \binom{m-k}{m-n}.$$

Les sommes A et B ressemblent à de vieilles connaissances : l'index du haut varie tandis que l'index du bas reste constant. Occupons-nous d'abord de B, car elle a l'air plus simple. En la triturant un peu, on peut la faire ressembler au membre gauche de (5.10) :

$$\begin{aligned} \sum_{0 \leq k \leq n} \binom{m-k}{m-n} &= \sum_{0 \leq m-k \leq n} \binom{m-(m-k)}{m-n} \\ &= \sum_{m-n \leq k \leq m} \binom{k}{m-n} \\ &= \sum_{0 \leq k \leq m} \binom{k}{m-n}. \end{aligned}$$

Les termes que nous avons ajoutés à la somme dans la dernière ligne, c'est-à-dire ceux pour lesquels  $0 \leq k < m - n$ , sont tous nuls car leur indice du haut est plus petit que leur indice du bas. Il ne nous reste plus qu'à utiliser (5.10) pour sommer sur l'indice du haut, obtenant ainsi

$$B = \sum_{0 \leq k \leq m} \binom{k}{m-n} = \binom{m+1}{m-n+1}.$$

La somme A est presque identique à B : m est simplement remplacé par  $m - 1$ . Nous avons donc une forme close pour la somme S, qui peut encore être simplifiée :

$$\begin{aligned} S &= mA - (m-n)B = m \binom{m}{m-n} - (m-n) \binom{m+1}{m-n+1} \\ &= \left( m - (m-n) \frac{m+1}{m-n+1} \right) \binom{m}{m-n} \\ &= \left( \frac{n}{m-n+1} \right) \binom{m}{m-n}. \end{aligned}$$

Tout ceci nous donne une forme close de la somme de départ :

$$T = S / \binom{m}{n} = \frac{n}{m-n+1} \binom{m}{m-n} / \binom{m}{n} = \frac{n}{m-n+1}.$$

Aucun arbitre ne pourrait simplifier davantage cette formule.

Vérifions, comme de coutume, la réponse sur un petit exemple : si  $m = 4$  et  $n = 2$ , on a

$$T = 0 \cdot \binom{3}{1} / \binom{4}{2} + 1 \cdot \binom{2}{1} / \binom{4}{2} + 2 \cdot \binom{1}{1} / \binom{4}{2} = 0 + \frac{2}{6} + \frac{2}{6} = \frac{2}{3},$$

ce qui correspond bien à notre formule  $2/(4 - 2 + 1)$ .

**Problème 3 : d'après un vieux examen.**

Voyons encore une fois une somme avec un seul coefficient binomial. Contrairement à la précédente, elle est d'origine scolaire : elle est tirée d'un vieux devoir. Il s'agit de calculer la valeur de  $Q_{1000000}$ , où

*Les vieux examens  
meurent-ils un jour ?*

$$Q_n = \sum_{k \leq 2^n} \binom{2^n - k}{k} (-1)^k, \quad n \geq 0 \text{ entier.}$$

Cette somme est plus difficile que les précédentes : *aucune* des identités que nous connaissons ne peut s'appliquer. De plus, il n'est pas question d'additionner un à un  $2^{1000000} + 1$  termes. Pour corser le tout, l'indice de sommation apparaît dans l'indice du haut et l'indice du bas avec des signes opposés ; et on n'a aucun intérêt à faire un changement de signe de l'indice du haut : on fait disparaître le facteur  $(-1)^k$ , mais on introduit un  $2k$  dans l'indice du haut.

Quand aucune méthode évidente ne marche, nous savons maintenant qu'il est bon de regarder des petits exemples. Voici donc les termes non nuls et leurs sommes pour les quatre premières valeurs de  $n$ .

$n$		$Q_n$
0	$\binom{1}{0}$	= 1
1	$\binom{2}{0} - \binom{1}{1}$	= 1 - 1 = 0
2	$\binom{4}{0} - \binom{3}{1} + \binom{2}{2}$	= 1 - 3 + 1 = -1
3	$\binom{8}{0} - \binom{7}{1} + \binom{6}{2} - \binom{5}{3} + \binom{4}{4}$	= 1 - 7 + 15 - 10 + 1 = 0

Mieux vaut s'arrêter là : en calculant pour  $n = 4$ , nous risquerions de nous tromper (calculer à la main des termes du genre de  $\binom{12}{4}$  et  $\binom{11}{5}$  ne se justifie qu'en cas de nécessité absolue).

Le début de notre suite est donc 1, 0, -1, 0. Même si nous connaissons un ou deux termes de plus, la forme close ne serait pas évidente. En revanche, si nous trouvions une récurrence pour  $Q_n$ , nous pourrions probablement en déduire une forme close. Pour cela, il nous faudrait trouver une relation entre  $Q_n$  et  $Q_{n-1}$  (ou entre  $Q_n$  et  $Q_{\text{valeurs plus petites}}$ ). Il faudrait donc par exemple lier un terme comme  $\binom{128-13}{13}$ , qui apparaît pour  $n = 7$  et  $k = 13$ , à des termes comme  $\binom{64-13}{13}$ . Cela ne semble pas très prometteur,

car nous ne connaissons aucune relation évidente entre des éléments du triangle de Pascal qui sont à 64 rangées de distance. La formule d'addition, notre principal outil d'induction, ne lie que des éléments de deux lignes contiguës.

Cependant, ces réflexions nous amènent à une observation cruciale : on n'a pas besoin de manipuler des éléments distants de  $2^{n-1}$  rangées. En effet, la variable  $n$  n'apparaît jamais en dehors de  $2^n$ . Ce  $2^n$  est destiné à distraire notre attention. Si nous remplaçons  $2^n$  par  $m$ , nous n'avons plus qu'à trouver une forme close pour la somme plus générale (mais plus facile)

*Quel est le pervers qui a posé un examen pareil ?*

$$R_m = \sum_{k \leq m} \binom{m-k}{k} (-1)^k, \quad m \geq 0 \text{ entier.}$$

Nous en déduirons immédiatement une forme close de  $Q_n = R_{2^n}$ . Avec un peu de chance, le formule d'addition devrait nous donner une récurrence pour la suite  $R_m$ .

Pour trouver les premières valeurs de  $R_m$ , il suffit d'additionner et soustraire alternativement les valeurs de la table 166 qui sont sur une diagonale Sud-Ouest Nord-Est. Voici le résultat :

$m$	0	1	2	3	4	5	6	7	8	9	10
$R_m$	1	1	0	-1	-1	0	1	1	0	-1	-1

Il semblerait que beaucoup de termes se neutralisent mutuellement.

Maintenant, regardons la formule qui définit  $R_m$  et voyons si nous pouvons en faire une récurrence. Notre stratégie va consister à appliquer la formule d'addition (5.8) et essayer de trouver des  $R_k$  dans l'expression trouvée. C'est assez similaire à la méthode de perturbation que nous avons vue au chapitre 2 :

$$\begin{aligned} R_m &= \sum_{k \leq m} \binom{m-k}{k} (-1)^k \\ &= \sum_{k \leq m} \binom{m-1-k}{k} (-1)^k + \sum_{k \leq m} \binom{m-1-k}{k-1} (-1)^k \\ &= \sum_{k \leq m} \binom{m-1-k}{k} (-1)^k + \sum_{k+1 \leq m} \binom{m-2-k}{k} (-1)^{k+1} \\ &= \sum_{k \leq m-1} \binom{m-1-k}{k} (-1)^k + \binom{-1}{m} (-1)^m \\ &\quad - \sum_{k \leq m-2} \binom{m-2-k}{k} (-1)^k - \binom{-1}{m-1} (-1)^{m-1} \\ &= R_{m-1} + (-1)^{2m} - R_{m-2} - (-1)^{2(m-1)} = R_{m-1} - R_{m-2}. \end{aligned}$$

(Dans la dernière étape, nous avons utilisé la formule  $\binom{-1}{m} = (-1)^m$ , qui est vraie comme nous le savons lorsque  $m \geq 0$ ). Ce calcul est valable pour  $m \geq 2$ .

Si on calcule rapidement quelques valeurs de  $R_m$  avec cette récurrence, on s'aperçoit vite que la suite est périodique. En effet,

$$R_m = \begin{cases} 1 & \text{si } m \bmod 6 = 0 \\ 1 & \text{si } m \bmod 6 = 1 \\ 0 & \text{si } m \bmod 6 = 2 \\ -1 & \text{si } m \bmod 6 = 3 \\ -1 & \text{si } m \bmod 6 = 4 \\ 0 & \text{si } m \bmod 6 = 5 \end{cases}$$

On peut le démontrer par induction. On peut aussi le démontrer plus rapidement en développant la récurrence :

$$R_m = (R_{m-2} - R_{m-3}) - R_{m-2} = -R_{m-3}$$

pour  $m \geq 3$ . Par conséquent,  $R_m = R_{m-6}$  pour  $m \geq 6$ .

Pour finir, n'oublions pas que nous cherchons  $Q_n = R_{2^n}$ . On peut trouver  $Q_n$  en déterminant la valeur de  $2^n \bmod 6$  et en appliquant la forme close de  $R_m$ . Lorsque  $n = 0$ , on a  $2^0 \bmod 6 = 1$ ; après quoi on multiplie toujours par 2 ( $\bmod 6$ ), donc on obtient alternativement 2 ou 4. Par conséquent,

$$Q_n = R_{2^n} = \begin{cases} R_1 = 1, & \text{si } n = 0, \\ R_2 = 0, & \text{si } n \text{ est impair,} \\ R_4 = -1, & \text{si } n > 0 \text{ est pair.} \end{cases}$$

Cette formule colle parfaitement aux premières valeurs de  $Q_n$  que nous avons calculées au début du problème. Nous concluons de tout cela que  $Q_{1000000} = R_4 = -1$ .

#### **Problème 4 : Une somme avec deux binomiaux.**

Il s'agit de trouver une forme close pour

$$\sum_{k=0}^n k \binom{m-k-1}{m-n-1}, \quad m > n \geq 0 \text{ entiers.}$$

Minute ! Où est le second binomial promis dans le titre ? Et pourquoi faudrait-il refaire un travail que nous avons déjà fait ? Cette somme est exactement la même que la somme S du problème 2.

Eh bien, figurez-vous qu'il est encore plus facile de la simplifier si on voit le terme général comme un produit de deux binomiaux. On n'a plus alors qu'à utiliser une des identités de la table 181. Pour faire apparaître le

*Ceux qui ont déjà fait l'exercice 4 le savent encore mieux.*

second binomial, il suffit de réécrire  $k$  en  $\binom{k}{1}$  :

$$\sum_{k=0}^n k \binom{m-k-1}{m-n-1} = \sum_{0 \leq k \leq n} \binom{k}{1} \binom{m-k-1}{m-n-1}.$$

Comme l'indice de sommation apparaît dans les deux indices du haut et avec des signes opposés, nous devons évidemment appliquer l'identité (5.26).

Cependant, notre somme n'est pas tout à fait dans la forme désirée pour pouvoir utiliser (5.26). Il faudrait pour cela que la borne haute de la sommation soit  $m-1$ . Le problème est résolu si on remarque que les termes tels que  $n < k \leq m-1$  sont nuls. Il suffit maintenant de poser  $(l, m, n, q) \leftarrow (m-1, m-n-1, 1, 0)$  dans (5.26) pour obtenir

$$S = \binom{m}{m-n+1}.$$

Cette formule est plus simple que celle que nous avions précédemment obtenue. On peut retrouver cette dernière en appliquant (5.7) :

$$\binom{m}{m-n+1} = \frac{n}{m-n+1} \binom{m}{m-n}.$$

Nous avons résolu ce problème en prenant des valeurs particulières des paramètres de l'identité (5.26). On peut, de la même façon, obtenir d'autres résultats intéressants. Prenons par exemple  $m=n=1$  et  $q=0$  dans (5.26). Cela donne

$$\sum_{0 \leq k \leq l} (l-k)k = \binom{l+1}{3}.$$

Le membre gauche est égal à  $l((l+1)l/2) - (1^2 + 2^2 + \dots + l^2)$ . Nous avons ainsi trouvé un moyen inédit de résoudre le problème de la somme des carrés que nous avons rencontré au chapitre 2.

Voici la morale de cette histoire : lorsqu'on veut démontrer quelque chose sur une somme, il est parfois payant de manipuler une somme beaucoup plus générale. Donc, lorsque vous apprenez une formule générale, pensez aux cas particuliers qu'elle implique.

### **Problème 5 : une somme à trois facteurs.**

En voici une autre : il s'agit de simplifier

$$\sum_k \binom{n}{k} \binom{s}{k} k, \quad n \geq 0 \text{ entier.}$$

L'index de sommation  $k$  apparaît dans les deux indices du bas, et avec le même signe. Si nous nous débrouillons convenablement, nous devrions donc

arriver à quelque chose nous permettant d'appliquer l'identité (5.23) de la table 181.

La principale différence entre notre somme et (5.23), c'est le facteur  $k$  supplémentaire. Qu'à cela ne tienne, nous n'avons qu'à le faire ingurgiter par l'un des binomiaux avec une identité d'absorption :

$$\begin{aligned}\sum_k \binom{n}{k} \binom{s}{k} k &= \sum_k \binom{n}{k} \binom{s-1}{k-1} s \\ &= s \sum_k \binom{n}{k} \binom{s-1}{k-1}.\end{aligned}$$

Le  $s$  qui apparaît n'est pas gênant car c'est une constante. Nous pouvons maintenant appliquer l'identité adéquate pour obtenir une forme close :

$$s \sum_k \binom{n}{k} \binom{s-1}{k-1} = s \binom{n+s-1}{n-1}.$$

Si nous avions choisi de faire absorber  $k$  à  $\binom{n}{k}$  plutôt qu'à  $\binom{s}{k}$ , nous n'aurions pas eu le droit d'appliquer directement (5.23). En effet,  $n-1$  peut être négatif et, pour que l'identité soit valable, il faut que l'un des indices du haut au moins soit positif ou nul.

### **Problème 6 : une somme à faire frémir.**

La somme que nous allons voir maintenant est plus coriace. Nous cherchons une forme close de

$$\sum_{k \geq 0} \binom{n+k}{2k} \binom{2k}{k} \frac{(-1)^k}{k+1}, \quad n \geq 0 \text{ entier.}$$

Un bon moyen de mesurer la difficulté d'une somme est de compter le nombre de fois qu'on y trouve l'indice de sommation. D'après cette échelle, nous sommes en face d'un problème particulièrement difficile. De plus, la méthode que nous avons suivie avec succès dans le problème précédent — absorption d'un facteur par un des binomiaux — ne marche pas ici : si on absorbe  $k+1$ , on se retrouve avec un  $k$  à la place. Pire encore : l'indice  $k$  apparaît deux fois avec un facteur 2 à l'intérieur d'un binomial. Il est en général plus difficile de se débarrasser des constantes multiplicatives que des constantes additives.

Malgré tout, nous avons de la chance, car les  $2k$  se trouvent exactement là où il faut pour pouvoir appliquer l'identité (5.21) :

$$\sum_{k \geq 0} \binom{n+k}{2k} \binom{2k}{k} \frac{(-1)^k}{k+1} = \sum_{k \geq 0} \binom{n+k}{k} \binom{n}{k} \frac{(-1)^k}{k+1}.$$

*C'est donc une somme de force 6 sur l'échelle de GKP.*

Les deux 2 disparaissent, ainsi qu'une occurrence de k. Il n'en reste donc plus que cinq.

Nous pouvons maintenant faire absorber le  $k+1$  du dénominateur, qui est très gênant, par  $\binom{n}{k}$  grâce à l'identité (5.6) :

$$\begin{aligned} \sum_{k \geq 0} \binom{n+k}{k} \binom{n}{k} \frac{(-1)^k}{k+1} &= \sum_k \binom{n+k}{k} \binom{n+1}{k+1} \frac{(-1)^k}{n+1} \\ &= \frac{1}{n+1} \sum_k \binom{n+k}{k} \binom{n+1}{k+1} (-1)^k. \end{aligned}$$

(N'oublions pas que  $n \geq 0$ ). Et de deux ; plus que quatre donc.

Pour éliminer un des autres k, nous avons le choix entre deux options qui semblent également prometteuses. Nous pouvons appliquer la règle de symétrie sur  $\binom{n+k}{k}$ , ou bien faire un changement de signe de l'indice du haut  $n+k$ , cette dernière option permettant d'éliminer aussi le facteur  $(-1)^k$ . Essayons donc les deux possibilités, en commençant par la symétrie :

$$\frac{1}{n+1} \sum_k \binom{n+k}{k} \binom{n+1}{k+1} (-1)^k = \frac{1}{n+1} \sum_k \binom{n+k}{n} \binom{n+1}{k+1} (-1)^k.$$

Et de trois. Nous sommes maintenant en bonne position pour gagner gros, grâce à (5.24) : en remplaçant  $(l, m, n, s)$  par  $(n+1, 1, n, n)$  dans cette identité, on trouve

$$\frac{1}{n+1} \sum_k \binom{n+k}{n} \binom{n+1}{k+1} (-1)^k = \frac{1}{n+1} (-1)^n \binom{n-1}{-1} = 0.$$

Zéro ? Est-ce bien vrai ? Vérifions donc pour  $n = 2$  :  $\binom{2}{0} \binom{0}{0} \frac{1}{1} - \binom{3}{2} \binom{2}{1} \frac{1}{2} + \binom{4}{4} \binom{4}{2} \frac{1}{3} = 1 - \frac{6}{2} + \frac{6}{3} = 0$ . Impeccable.

Juste pour le plaisir, voyons maintenant le résultat de l'autre option, qui consiste à faire un changement de signe de l'indice du haut de  $\binom{n+k}{k}$  :

$$\frac{1}{n+1} \sum_k \binom{n+k}{k} \binom{n+1}{k+1} (-1)^k = \frac{1}{n+1} \sum_k \binom{-n-1}{k} \binom{n+1}{k+1}.$$

Appliquons maintenant (5.23), avec  $(l, m, n, s) \leftarrow (n+1, 1, 0, -n-1)$ , pour trouver

$$\frac{1}{n+1} \sum_k \binom{-n-1}{k} \binom{n+1}{k+1} = \frac{1}{n+1} \binom{0}{n}.$$

Attendez un peu ! C'est bien égal à zéro lorsque  $n > 0$ , mais cela fait 1 si  $n = 0$ . Ce n'est pas du tout ce que nous avons trouvé précédemment.

*Pendant un court instant, j'ai cru qu'on allait jouer au loto.*

Calculons donc la somme pour  $n = 0$ , en utilisant la formule de l'énoncé du problème : on trouve 1. Par conséquent, la bonne réponse est “[ $n = 0$ ]”. Nous avons donc fait une erreur dans le calcul précédent.

Pour voir où est l'erreur, relisons attentivement ce calcul en considérant que  $n = 0$ . Bon sang, mais c'est bien sûr : nous sommes tombés dans le vieux piège que nous mentionnions plus haut. Nous avons appliqué la règle de symétrie alors que l'indice du haut peut être négatif ! Nous n'avons pas le droit de remplacer  $\binom{n+k}{k}$  par  $\binom{n+k}{n}$  lorsque  $k$  parcourt tout l'ensemble des entiers, parce que dans ce cas, si  $k < -n$ , on remplace un zéro par une valeur non définie (veuillez nous en excuser).

L'autre facteur de la somme,  $\binom{n+1}{k+1}$ , se trouve être nul lorsque  $k < -n$ , sauf si  $n = 0$  et  $k = -1$ . C'est pour cela que le cas  $n = 2$  ne nous a pas permis de détecter notre erreur. L'exercice 6 explique ce que nous aurions dû faire.

### **Problème 7 : Un nouvel obstacle.**

Voici un problème encore plus difficile : trouver une forme close pour

$$\sum_{k \geq 0} \binom{n+k}{m+2k} \binom{2k}{k} \frac{(-1)^k}{k+1}, \quad m, n > 0 \text{ entiers.}$$

Si  $m$  était égal à 0, cette somme serait exactement celle du problème précédent. Hélas ce n'est pas le cas, et c'est bien embêtant : rien de ce que nous avons fait dans le problème 6 ne marche ici ; surtout pas la première étape, qui était la plus importante. Par contre, si nous pouvions nous débarrasser du  $m$ , nous pourrions utiliser le résultat précédent. Voici donc la stratégie que nous allons suivre : remplacer  $\binom{n+k}{m+2k}$  par une somme de termes du genre de  $\binom{l+k}{2k}$ , où  $l$  est un certain entier positif ou nul. Le terme général sera similaire à celui du problème 6, et nous pourrons alors modifier l'ordre de sommation. Comment modifier  $\binom{n+k}{m+2k}$  comme nous le voulons ? Un examen approfondi des identités que nous connaissons permet de voir que l'équation (5.26) de la table 181 pourrait bien nous mener au but. Appliquons la donc en remplaçant respectivement les paramètres  $l$ ,  $m$ ,  $n$ ,  $q$  et  $k$  par  $n+k-1$ ,  $2k$ ,  $m-1$ ,  $0$  et  $j$  :

$$\begin{aligned} & \sum_{k \geq 0} \binom{n+k}{m+2k} \binom{2k}{k} \frac{(-1)^k}{k+1} \\ &= \sum_{k \geq 0} \sum_{0 \leq j \leq n+k-1} \binom{n+k-1-j}{2k} \binom{j}{m-1} \binom{2k}{k} \frac{(-1)^k}{k+1} \\ &= \sum_{j \geq 0} \binom{j}{m-1} \sum_{\substack{k \geq j-n+1 \\ k \geq 0}} \binom{n+k-1-j}{2k} \binom{2k}{k} \frac{(-1)^k}{k+1}. \end{aligned}$$

*Je conseille de faire une recherche dichotomique : examiner d'abord la formule du milieu pour déterminer si l'erreur se situe avant ou après.*

Dans la dernière étape, nous avons changé l'ordre de sommation en manipulant les conditions sous les sommes selon les règles vues au chapitre 2.

Nous ne pouvons pas encore utiliser le résultat du problème 6 pour réécrire la somme interne ; la condition  $k \geq j - n + 1$  est en trop. Toutefois, remarquons que cette condition n'est nécessaire que lorsque  $j - n + 1 > 0$ , c'est-à-dire lorsque  $j \geq n$  ; et dans ce cas, le premier binomial de la somme interne est nul, car son indice du haut est compris entre 0 et  $k - 1$ , donc strictement inférieur à l'index du bas  $2k$ . Nous pouvons donc, sans affecter les termes non nuls, ajouter à la somme externe la restriction  $j < n$ . De ce fait, la restriction  $k \geq j - n + 1$  devient superflue, et nous pouvons utiliser le résultat du problème 6. Voici ce que cela donne :

$$\begin{aligned} & \sum_{j \geq 0} \binom{j}{m-1} \sum_{\substack{k \geq j-n+1 \\ k \geq 0}} \binom{n+k-1-j}{2k} \binom{2k}{k} \frac{(-1)^k}{k+1} \\ &= \sum_{0 \leq j < n} \binom{j}{m-1} \sum_{k \geq 0} \binom{n+k-1-j}{2k} \binom{2k}{k} \frac{(-1)^k}{k+1} \\ &= \sum_{0 \leq j < n} \binom{j}{m-1} [n-1-j=0] = \binom{n-1}{m-1}. \end{aligned}$$

La somme interne est nulle sauf pour  $j = n - 1$ , ce qui nous donne une forme close toute simple.

### **Problème 8 : encore un autre obstacle.**

Considérons maintenant la somme

$$S_m = \sum_{k \geq 0} \binom{n+k}{2k} \binom{2k}{k} \frac{(-1)^k}{k+1+m}, \quad m, n \geq 0 \text{ entiers},$$

qui généralise encore une fois, mais de façon différente, celle du problème 6. En effet, pour  $m = 0$ , on obtient la même somme que précédemment, mais le  $m$  n'est pas au même endroit. Ce problème est un peu plus difficile que le problème 7, mais nous sommes maintenant bien aguerris. Commençons comme dans le problème 6 :

$$S_m = \sum_{k \geq 0} \binom{n+k}{k} \binom{n}{k} \frac{(-1)^k}{k+1+m}.$$

Maintenant, comme dans le problème 7, nous allons essayer de développer la partie qui dépend de  $m$  de façon à obtenir des termes que nous savons manipuler. Dans le problème précédent,  $m$  était nul et nous faisions absorber  $k+1$  par  $\binom{n}{k}$ . Pour faire la même chose lorsque  $m > 0$ , il faut arriver à décomposer  $1/(k+1+m)$  en termes qui peuvent être absorbés. Justement,

nous avons démontré au cours du problème 1 l'identité qu'il nous faut :

$$\sum_{j=0}^m \binom{m}{j} \binom{r}{j}^{-1} = \frac{r+1}{r+1-m}. \quad m \geq 0 \text{ entier}, \quad r \notin \{0, 1, \dots, m-1\}. \quad (5.33)$$

Il suffit de remplacer  $r$  par  $-k-2$  pour obtenir

$$S_m = \sum_{k \geq 0} \binom{n+k}{k} \binom{n}{k} \frac{(-1)^k}{k+1} \sum_{j \geq 0} \binom{m}{j} \binom{-k-2}{j}^{-1}.$$

Maintenant, le  $(k+1)^{-1}$  peut être absorbé par le  $\binom{n}{k}$ . En fait, il pourrait aussi être absorbé par le  $\binom{-k-2}{j}^{-1}$ . Cette double absorption laisse penser qu'il pourrait y avoir encore d'autres simplifications. En effet, si on développe tout le nouveau terme général en factorielles pour revenir ensuite aux coefficients binomiaux, on obtient une formule que nous savons sommer sur  $k$  :

$$\begin{aligned} S_m &= \frac{m! n!}{(m+n+1)!} \sum_{j \geq 0} (-1)^j \binom{m+n+1}{n+1+j} \sum_k \binom{n+1+j}{k+j+1} \binom{-n-1}{k} \\ &= \frac{m! n!}{(m+n+1)!} \sum_{j \geq 0} (-1)^j \binom{m+n+1}{n+1+j} \binom{j}{n}. \end{aligned}$$

*Je crois que nous sommes sensés vérifier tout ça sur une feuille de brouillon.*

D'après (5.24), la même somme sur tous les entiers  $j$  (donc sans la restriction  $j \geq 0$ ) est nulle. Par conséquent,  $-S_m$  est égale à la somme pour  $j < 0$ .

Pour évaluer cette dernière somme, remplaçons  $j$  par  $-k-1$  et sommes

pour  $k \geq 0$  :

$$\begin{aligned} S_m &= \frac{m! n!}{(m+n+1)!} \sum_{k \geq 0} (-1)^k \binom{m+n+1}{n-k} \binom{-k-1}{n} \\ &= \frac{m! n!}{(m+n+1)!} \sum_{k \leq n} (-1)^{n-k} \binom{m+n+1}{k} \binom{k-n-1}{n} \\ &= \frac{m! n!}{(m+n+1)!} \sum_{k \leq n} (-1)^k \binom{m+n+1}{k} \binom{2n-k}{n} \\ &= \frac{m! n!}{(m+n+1)!} \sum_{k \leq 2n} (-1)^k \binom{m+n+1}{k} \binom{2n-k}{n}. \end{aligned}$$

Il ne reste plus qu'à appliquer (5.25) pour trouver la réponse :

$$S_m = (-1)^n \frac{m! n!}{(m+n+1)!} \binom{m}{n} = (-1)^n m^n m^{-n-1}.$$

Ouais. Une petite vérification ne fera pas de mal. Pour  $n = 2$ , on trouve

$$S_m = \frac{1}{m+1} - \frac{6}{m+2} + \frac{6}{m+3} = \frac{m(m-1)}{(m+1)(m+2)(m+3)}.$$

Dans notre démonstration,  $m$  doit être entier. Le résultat est valable toutefois pour tout réel  $m$ , car  $(m+1)^{n+1} S_m$  est un polynôme en  $m$  de degré  $\leq n$ .

### 5.3 TRUCS ET ASTUCES

Nous allons voir ici trois techniques qui renforcent encore les méthodes que nous avons vues jusqu'à présent.

*C'est plutôt le  
truc 1/2.*

#### Truc numéro 1 : prendre un demi.

Il y a souvent un réel  $r$  quelconque dans les identités que nous voyons. Lorsque  $r$  se présente sous la forme "entier moins un demi", le binomial  $\binom{r}{k}$  peut se métamorphoser en un produit de binomiaux tout à fait différents de l'original. On peut en déduire toute une famille d'identités particulièrement pratiques.

Voyons cela plus précisément en commençant par la *formule de duplication*

$$r^k (r - \frac{1}{2})^k = (2r)^{2k} / 2^{2k}, \quad k \geq 0 \text{ entier.} \quad (5.34)$$

La preuve en est évidente : développons les puissances descendantes et faisons alterner les facteurs du membre gauche ainsi :

$$\begin{aligned} r(r - \frac{1}{2})(r - 1)(r - \frac{3}{2}) \dots (r - k + 1)(r - k + \frac{1}{2}) \\ = \frac{(2r)(2r - 1) \dots (2r - 2k + 1)}{2 \cdot 2 \cdot \dots \cdot 2}. \end{aligned}$$

Si on divise les deux membres par  $k!$ , on obtient

$$\binom{r}{k} \binom{r - 1/2}{k} = \binom{2r}{2k} \binom{2k}{k} / 2^{2k}, \quad k \text{ entier.} \quad (5.35)$$

En posant  $k = r = n$  avec  $n$  entier, cela devient

$$\binom{n - 1/2}{n} = \binom{2n}{n} / 2^{2n}, \quad n \text{ entier.} \quad (5.36)$$

Avec un changement de signe de l'indice du haut, on obtient une autre formule encore :

$$\binom{-1/2}{n} = \left(\frac{-1}{4}\right)^n \binom{2n}{n}, \quad n \text{ entier.} \quad (5.37)$$

Voici ce qui se passe pour  $n = 4$  par exemple :

$$\begin{aligned}\binom{-1/2}{4} &= \frac{(-1/2)(-3/2)(-5/2)(-7/2)}{4!} \\ &= \left(\frac{-1}{2}\right)^4 \frac{1 \cdot 3 \cdot 5 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4} \\ &= \left(\frac{-1}{4}\right)^4 \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdot 2 \cdot 4 \cdot 6 \cdot 8}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 1 \cdot 2 \cdot 3 \cdot 4} = \left(\frac{-1}{4}\right)^4 \binom{8}{4}.\end{aligned}$$

Observez bien comment nous avons transformé un produit de nombres impairs en factorielle.

L'identité (5.35) donne lieu à un corollaire amusant. Prenons  $r = \frac{1}{2}n$  et sommes sur tous les entiers  $k$ . On trouve

$$\begin{aligned}\sum_k \binom{n}{2k} \binom{2k}{k} 2^{-2k} &= \sum_k \binom{n/2}{k} \binom{(n-1)/2}{k} \\ &= \binom{n-1/2}{\lfloor n/2 \rfloor}, \quad n \geq 0 \text{ entier.}\end{aligned}\tag{5.38}$$

La dernière ligne se déduit de (5.23), car soit  $n/2$ , soit  $(n-1)/2$  est égal à  $\lfloor n/2 \rfloor$  qui est un entier positif ou nul !

Autre chose encore. D'après la convolution de Vandermonde (5.27),

$$\sum_k \binom{-1/2}{k} \binom{-1/2}{n-k} = \binom{-1}{n} = (-1)^n, \quad n \geq 0 \text{ entier.}$$

D'autre part, l'équation (5.37) nous permet d'écrire

$$\begin{aligned}\binom{-1/2}{k} \binom{-1/2}{n-k} &= \left(\frac{-1}{4}\right)^k \binom{2k}{k} \left(\frac{-1}{4}\right)^{n-k} \binom{2(n-k)}{n-k} \\ &= \frac{(-1)^n}{4^n} \binom{2k}{k} \binom{2n-2k}{n-k}.\end{aligned}$$

On en déduit une remarquable propriété des éléments "centraux" du triangle de Pascal :

$$\sum_k \binom{2k}{k} \binom{2n-2k}{n-k} = 4^n, \quad n \geq 0 \text{ entier.}\tag{5.39}$$

Par exemple,  $\binom{0}{0} \binom{6}{3} + \binom{2}{1} \binom{4}{2} + \binom{4}{2} \binom{2}{1} + \binom{6}{3} \binom{0}{0} = 1 \cdot 20 + 2 \cdot 6 + 6 \cdot 2 + 20 \cdot 1 = 64 = 4^3$ .

Tout ceci montre qu'il est tout à fait raisonnable d'essayer de transformer des binomiaux de la forme  $\binom{2k}{k}$  en binomiaux de la forme  $\binom{n-1/2}{k}$ ,

où  $n$  est un entier bien choisi (en général 0, 1, ou  $k$ ). La formule qu'on obtient alors peut être bien plus simple que l'originale.

**Truc numéro 2 : Différences d'ordre élevé.**

Nous savons déjà qu'on peut calculer des sommes partielles de la série  $\binom{n}{k}(-1)^k$ , mais pas de la série  $\binom{n}{k}$ . Les binomiaux à signe alternants  $\binom{n}{k}(-1)^k$  s'avèrent très importants dans de nombreuses applications. Une des raisons en est que ces coefficients sont intimement liés à l'opérateur de différence  $\Delta$  que nous avons défini dans la section 2.6.

La différence  $\Delta f$  d'une fonction  $f$  au point  $x$  est

$$\Delta f(x) = f(x+1) - f(x).$$

Si on applique à nouveau  $\Delta$  on obtient la différence seconde

$$\begin{aligned}\Delta^2 f(x) &= \Delta f(x+1) - \Delta f(x) = (f(x+2) - f(x+1)) - (f(x+1) - f(x)) \\ &= f(x+2) - 2f(x+1) + f(x),\end{aligned}$$

analogue à la dérivée seconde. De manière similaire, on a

$$\begin{aligned}\Delta^3 f(x) &= f(x+3) - 3f(x+2) + 3f(x+1) - f(x); \\ \Delta^4 f(x) &= f(x+4) - 4f(x+3) + 6f(x+2) - 4f(x+1) + f(x);\end{aligned}$$

et ainsi de suite. Ces expressions font apparaître des coefficients binomiaux à signes alternés.

Voici la formule générale pour la  $n$ ième différence :

$$\Delta^n f(x) = \sum_k \binom{n}{k} (-1)^{n-k} f(x+k), \quad n \geq 0 \text{ entier.} \quad (5.40)$$

On peut facilement la prouver par induction. Il existe aussi une jolie façon de la démontrer en utilisant la théorie élémentaire des opérateurs. Dans la section 2.6, nous avons défini l'opérateur de décalage selon la règle

$$Ef(x) = f(x+1).$$

Ainsi, l'opérateur  $\Delta$  est égal à  $E - 1$ , où  $1$  désigne l'opérateur identité défini par  $1f(x) = f(x)$ . D'après la formule du binôme, on a donc l'équation suivante, dont les éléments sont des opérateurs :

$$\Delta^n = (E - 1)^n = \sum_k \binom{n}{k} E^k (-1)^{n-k}.$$

Elle est équivalente à (5.40) car  $E^k$  est l'opérateur qui transforme  $f(x)$  en  $f(x+k)$ .

Voici un cas intéressant qui se présente lorsqu'on considère des puissances descendantes négatives. Soit  $f(x) = (x - 1)^{-1} = 1/x$ . Alors, d'après la règle (2.45), on a  $\Delta f(x) = (-1)(x - 1)^{-2}$ ,  $\Delta^2 f(x) = (-1)(-2)(x - 1)^{-3}$ , et plus généralement

$$\Delta^n((x - 1)^{-1}) = (-1)^n (x - 1)^{-n-1} = (-1)^n \frac{n!}{x(x + 1) \dots (x + n)}.$$

Maintenant l'équation (5.40) nous dit que

$$\begin{aligned} \sum_k \binom{n}{k} \frac{(-1)^k}{x+k} &= \frac{n!}{x(x+1)\dots(x+n)} \\ &= x^{-1} \binom{x+n}{n}^{-1}, \quad x \notin \{0, -1, \dots, -n\}. \end{aligned} \quad (5.41)$$

Par exemple,

$$\begin{aligned} \frac{1}{x} - \frac{4}{x+1} + \frac{6}{x+2} - \frac{4}{x+3} + \frac{1}{x+4} \\ = \frac{4!}{x(x+1)(x+2)(x+3)(x+4)} = \frac{1}{x \binom{x+4}{4}}. \end{aligned}$$

La somme de (5.41) constitue un développement en éléments simples de  $n!/(x(x+1)\dots(x+n))$ .

On peut aussi obtenir des résultats intéressants en considérant les puissances descendantes positives. Si  $f(x)$  est un polynôme de degré  $d$ , la différence  $\Delta f(x)$  est un polynôme de degré  $d-1$ . Par conséquent,  $\Delta^d f(x)$  est une constante et  $\Delta^n f(x) = 0$  pour tout  $n > d$ . C'est un résultat extrêmement important qui permet de simplifier beaucoup de formules.

Regardons tout cela d'un peu plus près encore. Soit

$$f(x) = a_d x^d + a_{d-1} x^{d-1} + \dots + a_1 x^1 + a_0 x^0$$

un polynôme de degré  $d$ . Nous verrons au chapitre 6 que les puissances ordinaires peuvent s'exprimer comme sommes de puissances descendantes (par exemple,  $x^2 = x^2 + x^1$ ). Par conséquent, il existe des coefficients  $b_d, b_{d-1}, \dots, b_1, b_0$  tels que

$$f(x) = b_d x^d + b_{d-1} x^{d-1} + \dots + b_1 x^1 + b_0 x^0.$$

(Pour être un peu plus précis, on sait que  $b_d = a_d$  et  $b_0 = a_0$ , mais les autres coefficients sont plus compliqués). Soit  $c_k = k! b_k$  pour  $0 \leq k \leq d$ . Alors

$$f(x) = c_d \binom{x}{d} + c_{d-1} \binom{x}{d-1} + \dots + c_1 \binom{x}{1} + c_0 \binom{x}{0}.$$

Ainsi, tout polynôme peut s'exprimer en une somme de multiples de coefficients binomiaux. Ce développement est appelé la *série de Newton*, car elle a été beaucoup utilisée par Isaac Newton.

Nous avons vu un peu plus haut que la formule d'addition implique que  $\Delta(\binom{x}{k}) = \binom{x}{k-1}$ . Par induction, on en déduit que la différence de la série de Newton est très simple :

$$\Delta^n f(x) = c_d \binom{x}{d-n} + c_{d-1} \binom{x}{d-1-n} + \cdots + c_1 \binom{x}{1-n} + c_0 \binom{x}{-n}.$$

Si on pose  $x = 0$ , tous les termes  $c_k \binom{x}{k-n}$  du membre droit s'annulent sauf celui pour lequel  $k - n = 0$  ; donc

$$\Delta^n f(0) = \begin{cases} c_n, & \text{si } n \leq d; \\ 0, & \text{si } n > d. \end{cases}$$

Voici donc la série de Newton pour  $f(x)$  :

$$f(x) = \Delta^d f(0) \binom{x}{d} + \Delta^{d-1} f(0) \binom{x}{d-1} + \cdots + \Delta f(0) \binom{x}{1} + f(0) \binom{x}{0}.$$

Si  $f(x) = x^3$  par exemple, on calcule facilement

$$\begin{aligned} f(0) &= 0, & f(1) &= 1, & f(2) &= 8, & f(3) &= 27; \\ \Delta f(0) &= 1, & \Delta f(1) &= 7, & \Delta f(2) &= 19; \\ \Delta^2 f(0) &= 6, & \Delta^2 f(1) &= 12; \\ \Delta^3 f(0) &= 6. \end{aligned}$$

La série de Newton est donc égale à  $x^3 = 6 \binom{x}{3} + 6 \binom{x}{2} + 1 \binom{x}{1} + 0 \binom{x}{0}$ .

En utilisant (5.40), on peut exprimer notre formule  $\Delta^n f(0) = c_n$  de la manière suivante :

$$\sum_k \binom{n}{k} (-1)^k \left( c_0 \binom{k}{0} + c_1 \binom{k}{1} + c_2 \binom{k}{2} + \cdots \right) = (-1)^n c_n, \quad n \geq 0 \text{ entier},$$

où  $\langle c_0, c_1, c_2, \dots \rangle$  est une suite quelconque de coefficients. Il n'y a pas de problème de convergence car la somme infinie  $c_0 \binom{k}{0} + c_1 \binom{k}{1} + c_2 \binom{k}{2} + \cdots$  est en réalité finie pour tout  $k \geq 0$ . Voici un cas particulier important de ce résultat :

$$\sum_k \binom{n}{k} (-1)^k (a_0 + a_1 k + \cdots + a_n k^n) = (-1)^n n! a_n, \quad n \geq 0 \text{ entier}, \quad (5.42)$$

car le polynôme  $a_0 + a_1 k + \cdots + a_n k^n$  peut toujours s'écrire comme une série de Newton  $c_0 \binom{k}{0} + c_1 \binom{k}{1} + \cdots + c_n \binom{k}{n}$  avec  $c_n = n! a_n$ .

Beaucoup de sommes qui semblent infaisables au premier abord peuvent être résolues presque trivialement avec les différences nièmes. Considérons par exemple l'identité

$$\sum_k \binom{n}{k} \binom{r - sk}{n} (-1)^k = s^n, \quad n \geq 0 \text{ entier.} \quad (5.43)$$

Elle est très impressionnante car elle ne ressemble à rien de ce que nous avons vu jusqu'à présent. En fait, il suffit de remarquer que le fameux facteur  $\binom{n}{k}(-1)^k$  est présent dans le terme général pour qu'elle devienne tout à coup très facile à comprendre. En effet, la fonction

$$f(k) = \binom{r - sk}{n} = \frac{1}{n!} (-1)^n s^n k^n + \dots = (-1)^n s^n \binom{k}{n} + \dots$$

est un polynôme en  $k$  de degré  $n$ , de coefficient directeur  $(-1)^n s^n / n!$ . Par conséquent, (5.43) n'est rien d'autre qu'une application de (5.42).

Lorsque nous avons introduit la série de Newton, nous avons supposé que  $f(x)$  était un polynôme. Toutefois, nous avons vu aussi que la série de Newton infinie

$$f(x) = c_0 \binom{x}{0} + c_1 \binom{x}{1} + c_2 \binom{x}{2} + \dots$$

a aussi un sens car elle est en fait finie si  $x$  est un entier positif ou nul. Notre calcul de  $\Delta^n f(0) = c_n$  est donc valable aussi bien dans le cas infini que dans le cas fini. L'identité générale

$$f(x) = f(0) \binom{x}{0} + \Delta f(0) \binom{x}{1} + \Delta^2 f(0) \binom{x}{2} + \Delta^3 f(0) \binom{x}{3} + \dots, \\ x \geq 0 \text{ entier} \quad (5.44)$$

est donc valable pour toute fonction  $f(x)$  définie sur les entiers positifs ou nuls  $x$ . De plus, si le membre droit converge pour les autres valeurs de  $x$ , alors notre formule définit une fonction qui "interpole"  $f(x)$  de façon naturelle. (Comme il y a un nombre infini de façons d'interpoler une fonction, nous n'avons pas le droit de dire que (5.44) est vraie pour tout  $x$  tel que la série infinie converge. Par exemple, si  $f(x) = \sin(\pi x)$ , on a  $f(x) = 0$  pour tout entier, donc le membre droit de (5.44) est nul ; mais le membre gauche est non nul pour tout  $x$  non entier).

La série de Newton est le pendant fini de la série de Taylor du calcul infinitésimal. Une série de Taylor s'écrit

$$g(a+x) = \frac{g(a)}{0!} x^0 + \frac{g'(a)}{1!} x^1 + \frac{g''(a)}{2!} x^2 + \frac{g'''(a)}{3!} x^3 + \dots.$$

(Comme  $E = 1 + \Delta$ , De façon analogue, la série de Newton pour  $f(x) = g(a + x)$  est

$$E^x = \sum_k \binom{x}{k} \Delta^k;$$

$$\text{et } E^x g(a) = g(a + x).$$

$$g(a + x) = \frac{g(a)}{0!} x^0 + \frac{\Delta g(a)}{1!} x^1 + \frac{\Delta^2 g(a)}{2!} x^2 + \frac{\Delta^3 g(a)}{3!} x^3 + \dots \quad (5.45)$$

(Le membre droit est identique à celui de (5.44) car, lorsque  $f(x) = g(a+x)$ ),  $\Delta^n f(0) = \Delta^n g(a)$  pour tout  $n \geq 0$ ). Les séries de Taylor et de Newton sont toutes deux finies si  $g$  est un polynôme ou si  $x = 0$ . La série de Newton est finie aussi si  $x$  est un entier positif. Pour d'autres valeurs de  $x$ , les sommes convergent ou non selon le cas. Si la série de Newton converge lorsque  $x$  n'est pas un entier positif ou nul, il se peut qu'elle converge vers une valeur différente de  $g(a + x)$ , car la série de Newton (5.45) ne dépend que des valeurs espacées d'une unité  $g(a), g(a+1), g(a+2), \dots$ .

Voici un exemple de série de Newton convergente, que nous devons à la formule du binôme. Soit  $g(x) = (1+z)^x$ , où  $z$  est un nombre complexe donné tel que  $|z| < 1$ . Alors  $\Delta g(x) = (1+z)^{x+1} - (1+z)^x = z(1+z)^x$ , donc  $\Delta^n g(x) = z^n(1+z)^x$ . Dans ce cas, la série de Newton infinie

$$g(a + x) = \sum_n \Delta^n g(a) \binom{x}{n} = (1+z)^a \sum_n \binom{x}{n} z^n$$

converge vers la "bonne" valeur  $(1+z)^{a+x}$ , pour tout  $x$ .

James Stirling a tenté d'utiliser les séries de Newton pour étendre la fonction factorielle à des valeurs non entières. Pour commencer, il trouva des coefficients  $S_n$  tels que l'identité

$$x! = \sum_n S_n \binom{x}{n} = S_0 \binom{x}{0} + S_1 \binom{x}{1} + S_2 \binom{x}{2} + \dots \quad (5.46)$$

*"Forasmuch as these terms increase very fast, their differences will make a diverging progression, which hinders the ordinate of the parabola from approaching to the truth; therefore in this and the like cases, I interpolate the logarithms of the terms, whose differences constitute a series swiftly converging."*

— J. Stirling [343]

est vraie pour  $x = 0, x = 1, x = 2$ , etc. Cependant, il découvrit que la série qui en résulte ne converge pas, sauf si  $x$  est un entier positif ou nul. Il essaya donc autre chose, en écrivant cette fois

$$\ln x! = \sum_n s_n \binom{x}{n} = s_0 \binom{x}{0} + s_1 \binom{x}{1} + s_2 \binom{x}{2} + \dots \quad (5.47)$$

Dans ce cas,  $\Delta(\ln x!) = \ln(x+1)! - \ln x! = \ln(x+1)$ , donc

$$\begin{aligned} s_n &= \Delta^n (\ln x!) \Big|_{x=0} \\ &= \Delta^{n-1} (\ln(x+1)) \Big|_{x=0} \\ &= \sum_k \binom{n-1}{k} (-1)^{n-1-k} \ln(k+1) \end{aligned}$$

d'après (5.40). Les coefficients sont donc  $s_0 = s_1 = 0$ ;  $s_2 = \ln 2$ ;  $s_3 = \ln 3 - 2\ln 2 = \ln \frac{3}{4}$ ;  $s_4 = \ln 4 - 3\ln 3 + 3\ln 2 = \ln \frac{32}{27}$ ; etc. De cette

façon, Stirling obtint une série convergente (bien qu'il ne le prouvât pas). Elle converge en effet pour tout  $x > -1$ . Ainsi il était capable de donner une valeur satisfaisante à  $\frac{1}{2}!$ . Vous trouverez la suite de l'histoire dans l'exercice 88.

(Ce n'est qu'à partir du dix-neuvième siècle qu'on sut faire des preuves de convergence).

### Truc numéro 3 : inversion.

Il y a un cas particulier particulièrement intéressant de la règle (5.45). Il peut s'écrire ainsi :

$$g(n) = \sum_k \binom{n}{k} (-1)^k f(k) \iff f(n) = \sum_k \binom{n}{k} (-1)^k g(k). \quad (5.48)$$

Cette relation duale entre  $f$  et  $g$  est appelée une *formule d'inversion*. Elle a un air de famille avec les formules d'inversion de Möbius (4.56) et (4.61) que nous avons rencontrées au chapitre 4. Les formules d'inversion nous fournissent un moyen de résoudre des "réurrences implicites", dans lesquelles une suite inconnue est contenue dans une somme.

Par exemple,  $g(n)$  peut être une fonction connue et  $f(n)$  une inconnue, et si on sait démontrer que  $g(n) = \sum_k \binom{n}{k} (-1)^k f(k)$ , alors (5.48) nous permet d'exprimer  $f$  comme une somme de valeurs connues.

On peut prouver (5.48) directement en utilisant les méthodes de base que nous avons vues au début du chapitre. Si  $g(n) = \sum_k \binom{n}{k} (-1)^k f(k)$  pour tout  $n \geq 0$ , alors

$$\begin{aligned} \sum_k \binom{n}{k} (-1)^k g(k) &= \sum_k \binom{n}{k} (-1)^k \sum_j \binom{k}{j} (-1)^j f(j) \\ &= \sum_j f(j) \sum_k \binom{n}{k} (-1)^{k+j} \binom{k}{j} \\ &= \sum_j f(j) \sum_k \binom{n}{j} (-1)^{k+j} \binom{n-j}{k-j} \\ &= \sum_j f(j) \binom{n}{j} \sum_k (-1)^k \binom{n-j}{k} \\ &= \sum_j f(j) \binom{n}{j} [n-j=0] = f(n). \end{aligned}$$

Inversez ceci :  
"Esope reste ici et se repose".

L'autre moitié de la preuve se passe bien sûr exactement de la même façon, car la relation qui lie  $f$  et  $g$  est symétrique.

Illustrons (5.48) en l'appliquant au "problème de la victoire au football". Soient  $n$  supporters d'une équipe qui gagne un match de football. Chaque supporter jette alors son chapeau en l'air. Les chapeaux retombent au hasard et chaque supporter en prend un. Quel est le nombre  $h(n, k)$

de configurations telles que  $k$  chapeaux exactement aient été récupérés par leurs propriétaires respectifs ?

Par exemple, si  $n = 4$  et si les supporters s'appellent A, B, C et D, voici les  $4! = 24$  configurations possibles avec, pour chacune, le nombre de propriétaires qui ont récupéré leurs propres chapeaux :

ABCD	4	BACD	2	CABD	1	DABC	0
ABDC	2	BADC	0	CADB	0	DACB	1
ACBD	2	BCAD	1	CBAD	2	DBAC	1
ACDB	1	BCDA	0	CBDA	1	DBCA	2
ADBC	1	BDAC	0	CDAB	0	DCAB	0
ADCB	2	BDCA	1	CDBA	0	DCBA	0

Donc,  $h(4, 4) = 1$ ,  $h(4, 3) = 0$ ,  $h(4, 2) = 6$ ,  $h(4, 1) = 8$  et  $h(4, 0) = 9$ .

Remarquons que  $h(n, k)$  est égal au nombre de façons de choisir  $k$  propriétaires de chapeaux chanceux, donc  $\binom{n}{k}$ , multiplié par le nombre de façons de faire en sorte qu'aucun des  $n - k$  chapeaux restants ne soit récupéré par son propriétaire, soit  $h(n - k, 0)$ . On appelle *déarrangement* une permutation qui déplace tous ses éléments. Le nombre de dérangements de  $n$  objets est parfois noté " $n!$ ", ce qui se lit "sous-factorielle" de  $n$ . Ainsi,  $h(n - k, 0) = (n - k)!$ , ce qui nous permet d'écrire la formule

$$\begin{aligned} h(n, k) &= \binom{n}{k} h(n - k, 0) \\ &= \binom{n}{k} (n - k)!. \end{aligned}$$

(Il n'y a pas de notation standard pour la sous-factorielle. Celle que nous avons prise n'est peut être pas parfaite, mais attendons de l'avoir utilisée un peu pour voir si elle convient. Dans le cas contraire, il sera toujours temps de prendre autre chose à la place, comme  $D_n$  par exemple).

Si nous connaissons une forme close pour  $n!$ , notre problème serait résolu. Voyons voir ce que nous pouvons faire. Il y a une récurrence facile à trouver, du fait que la somme des  $h(n, k)$  pour tout  $k$  est égale au nombre total de permutations de  $n$  chapeaux :

$$\begin{aligned} n! &= \sum_k h(n, k) \\ &= \sum_k \binom{n}{k} (n - k)! \\ &= \sum_k \binom{n}{k} k!, \quad n \geq 0 \text{ entier.} \end{aligned} \tag{5.49}$$

(Lors de la dernière étape, nous avons transformé  $k$  en  $n - k$  et  $\binom{n}{n-k}$  en  $\binom{n}{k}$ ). Cette récurrence implicite nous permet de calculer tous les  $h(n, k)$

que nous voulons :

n	$h(n, 0)$	$h(n, 1)$	$h(n, 2)$	$h(n, 3)$	$h(n, 4)$	$h(n, 5)$	$h(n, 6)$
0	1						
1	0	1					
2	1	0	1				
3	2	3	0	1			
4	9	8	6	0	1		
5	44	45	20	10	0	1	
6	265	264	135	40	15	0	1

Voici par exemple ce qu'il faut faire pour calculer la ligne  $n = 4$ . Les deux nombres de droite sont évidents : il y a exactement une façon de récupérer les chapeaux comme il faut, et il n'est pas possible que trois supporters exactement récupèrent leurs propres chapeaux (car dans ce cas, qu'aurait récupéré le quatrième ?). Pour  $k = 2$  et  $k = 1$ , nous pouvons appliquer notre équation de  $h(n, k)$ , pour trouver  $h(4, 2) = \binom{4}{2}h(2, 0) = 6 \cdot 1 = 6$  et  $h(4, 1) = \binom{4}{1}h(3, 0) = 4 \cdot 2 = 8$ . Par contre, nous ne pouvons pas utiliser cette équation pour  $h(4, 0)$ . En fait nous le pouvons, mais cela donne  $h(4, 0) = \binom{4}{0}h(4, 0)$ , ce qui est vrai mais absolument inutile. Essayons donc autre chose : on peut déduire de la relation  $h(4, 0) + 8 + 6 + 0 + 1 = 4!$  que  $h(4, 0) = 9$ , qui est aussi la valeur de  $4_i$ . De la même façon,  $n_i$  dépend des valeurs de  $k_j$  pour  $k < n$ .

Maintenant, comment résoudre la récurrence (5.49) ? Facile : elle est exactement de la même forme que (5.48), avec  $g(n) = n!$  et  $f(k) = (-1)^k k_i$ . Par conséquent, voici la solution :

$$n_i = (-1)^n \sum_k \binom{n}{k} (-1)^k k_i.$$

En fait, ce n'est pas une solution vraiment satisfaisante, bien qu'elle constitue un progrès par rapport à la récurrence. N'y a-t-il pas moyen d'en trouver une forme close ? Remarquons que le  $k!$  peut être neutralisé par un  $k!$  caché dans  $\binom{n}{k}$ . Allons-y donc :

$$n_i = \sum_{0 \leq k \leq n} \frac{n!}{(n-k)!} (-1)^{n+k} = n! \sum_{0 \leq k \leq n} \frac{(-1)^k}{k!}. \quad (5.50)$$

La somme qui reste converge rapidement vers le nombre  $\sum_{k \geq 0} (-1)^k / k! = e^{-1}$ . En fait, les termes qui sont exclus de la somme sont

$$\begin{aligned} n! \sum_{k > n} \frac{(-1)^k}{k!} &= \frac{(-1)^{n+1}}{n+1} \sum_{k \geq 0} (-1)^k \frac{(n+1)!}{(k+n+1)!} \\ &= \frac{(-1)^{n+1}}{n+1} \left( 1 - \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} - \dots \right), \end{aligned}$$

*Tout l'art, en mathématiques comme dans la vie, consiste à savoir quelles vérités sont inutiles.*

et ce qui est entre parenthèses est compris entre 1 et  $1 - \frac{1}{n+2} = \frac{n+1}{n+2}$ . Par conséquent, la valeur absolue de la différence entre  $n_i$  et  $n!/e$  vaut à peu près  $1/n$ . Pour être plus précis, elle se trouve entre  $1/(n+1)$  et  $1/(n+2)$ . Or,  $n_i$  est un entier. A ce titre, s'il est strictement positif, il est forcément égal à l'arrondi de  $n!/e$  à l'entier le plus proche. Voici donc la forme close que nous cherchons :

$$n_i = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor + [n=0]. \quad (5.51)$$

C'est le nombre de configurations telles qu'aucun supporter ne récupère le bon chapeau. Lorsque  $n$  est très grand, il est plus intéressant de connaître la *probabilité* que cela arrive. Si on suppose que toutes les configurations ont la même probabilité (parce que les chapeaux ont été lancés très haut), alors la probabilité qu'aucun supporter ne récupère son propre chapeau est égale à

$$\frac{n_i}{n!} = \frac{n!/e + O(1)}{n!} \sim \frac{1}{e} = 0,367\dots$$

ce qui donne à peu près 37% (lorsque  $n$  est suffisamment grand).

Notons en passant que la récurrence (5.49) pour les sous-factorielles est exactement la même que la récurrence (5.46), celle que Stirling a d'abord considérée pour généraliser la fonction factorielle. Par conséquent,  $S_k = k_i$ . Ces coefficients sont tellement grands qu'il n'est pas étonnant que la série infinie (5.46) diverge lorsque  $x$  n'est pas un entier.

Avant de quitter ce problème, regardons rapidement deux motifs intéressants qui apparaissent dans la table des premières valeurs de  $h(n, k)$ . Tout d'abord, on dirait que les nombres 1, 3, 6, 10, 15, ... qui sont sous la diagonale des 0 sont exactement les nombres triangulaires. C'est en effet vrai, et c'est facile à prouver car ces nombres sont les  $h(n, n-2)$  et

$$h(n, n-2) = \binom{n}{n-2} 2^i = \binom{n}{2}.$$

Il semblerait aussi que les nombres des deux premières colonnes diffèrent de  $\pm 1$ . C'est encore vrai :

$$\begin{aligned} h(n, 0) - h(n, 1) &= n_i - n(n-1)_i \\ &= \left( n! \sum_{0 \leq k \leq n} \frac{(-1)^k}{k!} \right) - \left( n(n-1)! \sum_{0 \leq k \leq n-1} \frac{(-1)^k}{k!} \right) \\ &= n! \frac{(-1)^n}{n!} = (-1)^n. \end{aligned}$$

*Pour les fans de base-ball : 0,367 représente aussi la moyenne au bâton de Ty Cobb durant sa carrière. Croyez-vous que ce soit une coïncidence ?*

*(Erreur ! La moyenne de Cobb est de  $4191/11429 \approx 0,366699$ , tandis que  $1/e \approx 0,367879$ . Mais si Wade Boggs fait quelques très bonnes saisons...)*

*(La "moyenne au bâton" est un terme de baseball... québécois (N.d.T.)).*

En d'autres termes,  $n_i = n(n-1)_i + (-1)^n$ . Cette récurrence pour le nombre de dérangements est bien plus simple que celle que nous avions jusqu'ici.

Pour finir, appliquons l'inversion à quelque chose d'autre, par exemple à la formule

$$\sum_k \binom{n}{k} \frac{(-1)^k}{x+k} = \frac{1}{x} \left( \frac{x+n}{n} \right)^{-1}$$

que nous avons démontrée en (5.41). Nous trouvons

$$\frac{x}{x+n} = \sum_{k \geq 0} \binom{n}{k} (-1)^k \left( \frac{x+k}{k} \right)^{-1}.$$

C'est intéressant, mais pas franchement nouveau. En effet, si on effectue un changement de signe de l'indice du haut de  $\binom{x+k}{k}$ , on redécouvre tout bonnement l'identité (5.33).

## 5.4 FONCTIONS GÉNÉRATRICES

Nous voici arrivés à la notion la plus importante de ce livre : les *fonctions génératrices*. Toute suite infinie  $\langle a_0, a_1, a_2, \dots \rangle$  peut être représentée par une *série entière* en une variable auxiliaire  $z$  :

$$A(z) = a_0 + a_1 z + a_2 z^2 + \dots = \sum_{k \geq 0} a_k z^k. \quad (5.52)$$

Nous avons choisi d'appeler la variable auxiliaire  $z$  parce que nous la considérerons souvent comme un nombre complexe. C'est cette lettre  $z$  qui est habituellement utilisée en théorie des variables complexes, dans laquelle les séries entières (aussi appelées fonctions analytiques ou fonctions holomorphes) occupent une place prépondérante.

Au cours des chapitres suivants, nous verrons énormément de séries génératrices. Le chapitre 7 leur est même entièrement consacré. Nous allons donc nous contenter ici d'introduire les notions de base et de montrer en quoi elles sont utiles à l'étude des coefficients binomiaux.

Ce qui rend une fonction génératrice très pratique, c'est le fait qu'elle représente à elle seule un nombre infini d'éléments d'une suite. On peut résoudre beaucoup de problèmes de suites infinies en créant une ou plusieurs fonctions génératrices à partir des coefficients des suites, puis en jouant avec elles assez pour bien les connaître, enfin en regardant de nouveau les coefficients. Avec un peu de chance, on en sait assez sur les fonctions pour en déduire ce que nous cherchons sur les coefficients.

Pour toute série entière  $A(z) = \sum_{k \geq 0} a_k z^k$ , on note

$$[z^n] A(z) = a_n.$$

(On trouvera dans [223] des détails sur l'*histoire et l'utilité de cette notation*).

En d'autres termes,  $[z^n] A(z)$  désigne le coefficient de  $z^n$  dans  $A(z)$ .

Soit  $A(z)$  la fonction génératrice qui correspond à  $\langle a_0, a_1, a_2, \dots \rangle$  et  $B(z)$  celle qui correspond à une autre suite  $\langle b_0, b_1, b_2, \dots \rangle$ . Alors le produit  $A(z)B(z)$  est égal à la série entière

$$\begin{aligned} & (a_0 + a_1 z + a_2 z^2 + \dots)(b_0 + b_1 z + b_2 z^2 + \dots) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0)z + (a_0 b_2 + a_1 b_1 + a_2 b_0)z^2 + \dots. \end{aligned}$$

Le coefficient de  $z^n$  dans ce produit est

$$a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0 = \sum_{k=0}^n a_k b_{n-k}.$$

Donc, si on veut évaluer une somme de la forme

$$c_n = \sum_{k=0}^n a_k b_{n-k}, \quad (5.54)$$

et si on connaît les fonctions génératrices  $A(z)$  et  $B(z)$ , alors on peut dire que

$$c_n = [z^n] A(z)B(z).$$

La suite  $\langle c_n \rangle$  définie en (5.54) est appelée la *convolution* des suites  $\langle a_n \rangle$  et  $\langle b_n \rangle$ . Pour faire une convolution de deux suites, on écrit la somme de tous les produits à deux termes dont la somme des indices est égale à une valeur donnée. Nous venons donc de voir que faire une convolution de deux suites équivaut à effectuer le produit de leurs fonctions génératrices.

Les fonctions génératrices sont un moyen très puissant de découvrir et de prouver des identités. Par exemple, d'après la formule du binôme, la fonction génératrice de la suite  $\langle \binom{r}{0}, \binom{r}{1}, \binom{r}{2}, \dots \rangle$  est  $(1+z)^r$  :

$$(1+z)^r = \sum_{k \geq 0} \binom{r}{k} z^k.$$

Similairement,

$$(1+z)^s = \sum_{k \geq 0} \binom{s}{k} z^k.$$

En multipliant ces deux fonctions, on obtient une nouvelle fonction génératrice :

$$(1+z)^r(1+z)^s = (1+z)^{r+s}.$$

Maintenant, si on écrit l'équation correspondante pour les coefficients de  $z^n$ , on trouve

$$\sum_{k=0}^n \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}.$$

$$(5.27)! = \\ (5.27)(4.27) \\ (3.27)(2.27) \\ (1.27)(0.27)!.$$

Nous avons découvert la convolution de Vandermonde (5.27) !

Plutôt facile, non ? Essayons-en une autre. Cette fois, nous allons prendre  $(1-z)^r$ , qui est la fonction génératrice de la suite  $\langle (-1)^n \binom{r}{n} \rangle = \langle \binom{r}{0}, -\binom{r}{1}, \binom{r}{2}, \dots \rangle$ . En multipliant par  $(1+z)^r$ , on obtient une nouvelle fonction génératrice :

$$(1-z)^r(1+z)^r = (1-z^2)^r,$$

ce qui nous donne l'équation suivante pour les coefficients de  $z^n$  :

$$\sum_{k=0}^n \binom{r}{k} \binom{r}{n-k} (-1)^k = (-1)^{n/2} \binom{r}{n/2} [n \text{ pair}]. \quad (5.55)$$

Vérifions cela sur un ou deux petits exemples. Si  $n = 3$ , le résultat est

$$\binom{r}{0} \binom{r}{3} - \binom{r}{1} \binom{r}{2} + \binom{r}{2} \binom{r}{1} - \binom{r}{3} \binom{r}{0} = 0.$$

Chaque terme positif est neutralisé par un terme négatif qui lui correspond. Ce cas, qui se produit pour tout  $n$  impair, n'est pas franchement intéressant. Par contre, lorsque  $n$  est pair, par exemple  $n = 2$ , on obtient une somme non triviale différente de la convolution de Vandermonde :

$$\binom{r}{0} \binom{r}{2} - \binom{r}{1} \binom{r}{1} + \binom{r}{2} \binom{r}{0} = 2 \binom{r}{2} - r^2 = -r.$$

Pour  $n = 2$ , le résultat correspond donc bien à (5.55). Pour finir, remarquons que l'équation (5.30) est un cas particulier de notre nouvelle identité (5.55).

Les coefficients binomiaux apparaissent aussi dans d'autres fonctions génératrices. Voici deux identités particulièrement importantes, dans lesquelles l'indice du haut varie tandis que celui du bas est fixe :

$$\frac{1}{(1-z)^{n+1}} = \sum_{k \geq 0} \binom{n+k}{n} z^k, \quad n \geq 0 \text{ entier} \quad (5.56)$$

$$\frac{z^n}{(1-z)^{n+1}} = \sum_{k \geq 0} \binom{k}{n} z^k, \quad n \geq 0 \text{ entier}. \quad (5.57)$$

*Il serait bon d'encadrer ou de surligner ces deux équations.*

Pour trouver la seconde identité, on a simplement multiplié la première par  $z^n$ , pour "décaler" les coefficients de  $n$  places vers la droite. La première

identité est en fait un cas particulier de la formule du binôme : si on développe  $(1 - z)^{-n-1}$  en utilisant (5.13), on obtient  $\binom{-n-1}{k}(-1)^k$  comme coefficient de  $z^k$ . En effectuant un changement de signe de l'indice du haut, on peut le transformer en  $\binom{k+n}{k}$  ou en  $\binom{n+k}{n}$ . Ces deux cas particuliers valent la peine d'être écrits explicitement, car nous verrons qu'ils sont souvent utiles.

Voici un cas particulier d'un cas particulier : si  $n = 0$ , on a la série géométrique

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \dots = \sum_{k \geq 0} z^k.$$

C'est la fonction génératrice de la suite  $\langle 1, 1, 1, \dots \rangle$ , qui est particulièrement utile, car la convolution de n'importe quelle suite avec celle-ci donne la suite des sommes : lorsque  $b_k = 1$  pour tout  $k$ , (5.54) devient

$$c_n = \sum_{k=0}^n a_k.$$

Si  $A(z)$  est la fonction génératrice des termes généraux  $\langle a_0, a_1, a_2, \dots \rangle$ , alors  $A(z)/(1-z)$  est la fonction génératrice des sommes  $\langle c_0, c_1, c_2, \dots \rangle$ .

Le problème des dérangements (celui des supporters et des chapeaux), que nous avons résolu un peu plus haut en utilisant l'inversion, peut aussi être résolu avec les fonctions génératrices, et d'une façon assez intéressante. Si on développe  $\binom{n}{k}$  en factorielles et si on divise les deux membres de la récurrence de base

$$n! = \sum_k \binom{n}{k} (n-k)_i$$

par  $n!$ , on obtient une convolution :

$$1 = \sum_{k=0}^n \frac{1}{k!} \frac{(n-k)_i}{(n-k)!}.$$

La fonction génératrice de la suite  $\langle \frac{1}{0!}, \frac{1}{1!}, \frac{1}{2!}, \dots \rangle$  est  $e^z$  ; donc, si on pose

$$D(z) = \sum_{k \geq 0} \frac{k_i}{k!} z^k,$$

on peut écrire

$$\frac{1}{1-z} = e^z D(z).$$

En résolvant cette équation en  $D(z)$ , on trouve

$$D(z) = \frac{1}{1-z} e^{-z} = \frac{1}{1-z} \left( \frac{1}{0!} z^0 - \frac{1}{1!} z^1 + \frac{1}{2!} z^2 + \dots \right).$$

Voici l'égalité correspondante pour les coefficients de  $z^n$  :

$$\frac{n_j}{n!} = \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

C'est exactement la formule que nous avons obtenue précédemment par inversion.

Jusqu'ici, les fonctions génératrices nous ont permis d'écrire de belles preuves de résultats que nous avions démontré auparavant de manière bien moins plaisante. Cependant, nous n'avons pas encore appliqué ce nouvel outil pour trouver de nouveaux résultats, à part (5.55). Nous voici maintenant prêts à le faire. Commençons par définir deux séries qui vont engendrer une très riche famille d'identités, la *série binomiale généralisée*  $B_t(z)$  et la *série exponentielle généralisée*  $E_t(z)$  :

$$B_t(z) = \sum_{k \geq 0} (tk)^{k-1} \frac{z^k}{k!}; \quad E_t(z) = \sum_{k \geq 0} (tk+1)^{k-1} \frac{z^k}{k!}. \quad (5.58)$$

Nous démontrerons en section 7.5 que ces fonctions satisfont les identités

$$B_t(z)^{1-t} - B_t(z)^{-t} = z; \quad E_t(z)^{-t} \ln E_t(z) = z. \quad (5.59)$$

Voici pourquoi ces séries sont appelées binomiale et exponentielle “généralisées” : si  $t = 0$ , on a

$$B_0(z) = 1 + z; \quad E_0(z) = e^z;$$

Les identités suivantes sont valables pour tout réel  $r$  :

$$B_t(z)^r = \sum_{k \geq 0} \binom{tk+r}{k} \frac{r}{tk+r} z^k; \quad E_t(z)^r = \sum_{k \geq 0} r \frac{(tk+r)^{k-1}}{k!} z^k; \quad (5.60)$$

$$\frac{B_t(z)^r}{1 - t + t B_t(z)^{-1}} = \sum_{k \geq 0} \binom{tk+r}{k} z^k;$$

$$\frac{E_t(z)^r}{1 - zt E_t(z)^t} = \sum_{k \geq 0} \frac{(tk+r)^k}{k!} z^k. \quad (5.61)$$

*La série binomiale généralisée  $B_t(z)$  fut découverte vers 1750 par J. H. Lambert [236, §38]. Il remarqua quelques années plus tard [237] que ses puissances satisfaisaient la première identité de (5.60). L'exercice 84 montre comment déduire (5.61) de (5.60).*

(Lorsque  $tk + r = 0$ , il faut faire attention à la façon dont le coefficient de  $z^k$  doit être interprété. Les coefficients sont des polynômes en  $r$ . En particulier, le terme constant de  $\mathcal{E}_t(z)^r$  est  $r(0+r)^{-1}$ , qui est égal à 1 même lorsque  $r = 0$ ).

Il nous est maintenant possible d'obtenir des identités très générales en combinant les séries des équations (5.60) et (5.61), élevées à des puissances différentes. Par exemple,

$$\begin{aligned} \mathcal{B}_t(z)^r \frac{\mathcal{B}_t(z)^s}{1-t+t\mathcal{B}_t(z)^{-1}} &= \sum_{k \geq 0} \binom{tk+r}{k} \frac{r}{tk+r} z^k \sum_{j \geq 0} \binom{tj+s}{j} z^j \\ &= \sum_{n \geq 0} z^n \sum_{k \geq 0} \binom{tk+r}{k} \frac{r}{tk+r} \binom{t(n-k)+s}{n-k}. \end{aligned}$$

Cette série est aussi égale à

$$\frac{\mathcal{B}_t(z)^{r+s}}{1-t+t\mathcal{B}_t(z)^{-1}} = \sum_{n \geq 0} \binom{tn+r+s}{n} z^n;$$

ce qui nous donne l'identité

$$\sum_k \binom{tk+r}{k} \binom{t(n-k)+s}{n-k} \frac{r}{tk+r} = \binom{tn+r+s}{n}, \quad n \text{ entier},$$

valable pour tous réels  $r$ ,  $s$  et  $t$ . Lorsque  $t = 0$ , on retrouve la convolution de Vandermonde. (Si le dénominateur  $tk + r$  se trouve être nul, on évite la division par zéro en considérant qu'il est neutralisé par le facteur  $tk + r$  du numérateur du premier binomial. Les deux membres de l'identité sont des polynômes en  $r$ ,  $s$  et  $t$ ). En effectuant d'autres combinaisons, par exemple en multipliant  $\mathcal{B}_t(z)^r$  par  $\mathcal{B}_t(z)^s$ , on obtient d'autres identités. Elles sont présentées dans la table 216.

Regardons maintenant quelques cas particuliers de ces identités. Nous avons déjà remarqué qu'on peut ainsi découvrir des résultats très intéressants. Que se passe-t-il par exemple si on pose  $t = 1$ ? Le binomial généralisé  $\mathcal{B}_1(z)$  est très simple :

$$\mathcal{B}_1(z) = \sum_{k \geq 0} z^k = \frac{1}{1-z};$$

il ne nous apprend donc rien du tout. Par contre,  $\mathcal{E}_1(z)$  est plus intéressant.

$$\mathcal{E}(z) = \sum_{k \geq 0} (k+1)^{k-1} \frac{z^k}{k!} = 1 + z + \frac{3}{2}z^2 + \frac{8}{3}z^3 + \frac{125}{24}z^4 + \dots \quad (5.62)$$

**Table 216** Identités générales de convolutions, pour tout entier  $n \geq 0$ .

$$\sum_k \binom{tk+r}{k} \binom{tn-tk+s}{n-k} \frac{r}{tk+r} = \binom{tn+r+s}{n}. \quad (5.63)$$

$$\begin{aligned} \sum_k \binom{tk+r}{k} \binom{tn-tk+s}{n-k} \frac{r}{tk+r} \cdot \frac{s}{tn-tk+s} \\ = \binom{tn+r+s}{n} \frac{r+s}{tn+r+s}. \end{aligned} \quad (5.64)$$

$$\sum_k \binom{n}{k} (tk+r)^k (tn-tk+s)^{n-k} \frac{r}{tk+r} = (tn+r+s)^n. \quad (5.65)$$

$$\begin{aligned} \sum_k \binom{n}{k} (tk+r)^k (tn-tk+s)^{n-k} \frac{r}{tk+r} \cdot \frac{s}{tn-tk+s} \\ = (tn+r+s)^n \frac{r+s}{tn+r+s}. \end{aligned} \quad (5.66)$$

C'est une fonction importante que nous n'avons encore jamais rencontrée. Elle satisfait l'identité suivante :

$$\mathcal{E}(z) = e^{z\mathcal{E}(z)}. \quad (5.67)$$

Cette fonction, qui a été d'abord étudiée par Euler [117] puis par Eisenstein [91], intervient dans énormément d'applications [193, 204].

Les cas particuliers  $t = 2$  et  $t = -1$  du binomial généralisé sont particulièrement intéressants, car leurs coefficients apparaissent très souvent dans les problèmes qui portent sur des structures récursives. Ecrivons donc explicitement ces séries pour pouvoir nous y référer plus tard :

$$\begin{aligned} \mathcal{B}_2(z) &= \sum_k \binom{2k}{k} \frac{z^k}{1+k} \\ &= \sum_k \binom{2k+1}{k} \frac{z^k}{1+2k} = \frac{1-\sqrt{1-4z}}{2z}. \end{aligned} \quad (5.68)$$

$$\begin{aligned} \mathcal{B}_{-1}(z) &= \sum_k \binom{1-k}{k} \frac{z^k}{1-k} \\ &= \sum_k \binom{2k-1}{k} \frac{(-z)^k}{1-2k} = \frac{1+\sqrt{1+4z}}{2}. \end{aligned} \quad (5.69)$$

Ah ! C'est la fonction des puissances itérées  
 $\mathcal{E}(\ln z) = z^{z^{z^{\dots}}}$ .  
 Je me suis souvent posé des questions sur elle.

Zzzzzzz...

Notez aussi  
 $\mathcal{B}_{1/2}(z)^r = (\sqrt{z^2+4}+z)^{2r}/4^r$ ,  
 elle vaut le coup.

$$\mathcal{B}_2(z)^r = \sum_k \binom{2k+r}{k} \frac{r}{2k+r} z^k. \quad (5.70)$$

$$\mathcal{B}_{-1}(z)^r = \sum_k \binom{r-k}{k} \frac{r}{r-k} z^k. \quad (5.71)$$

$$\frac{\mathcal{B}_2(z)^r}{\sqrt{1-4z}} = \sum_k \binom{2k+r}{k} z^k. \quad (5.72)$$

$$\frac{\mathcal{B}_{-1}(z)^{r+1}}{\sqrt{1+4z}} = \sum_k \binom{r-k}{k} z^k. \quad (5.73)$$

Les coefficients  $\binom{2n}{n} \frac{1}{n+1}$  de la série  $\mathcal{B}_2(z)$  sont appelés les *nombres de Catalan*, du nom d'Eugène Catalan qui a rédigé vers 1830 un article très important à leur propos. Voici les premiers termes de la suite :

$n$	0	1	2	3	4	5	6	7	8	9	10
$C_n$	1	1	2	5	14	42	132	429	1430	4862	16796

Les coefficients de  $\mathcal{B}_{-1}(z)$  sont presque identiques. On ajoute juste un 1 au début, puis la suite est la même, à ceci près que les signes des éléments sont alternés :  $\langle 1, 1, -1, 2, -5, 14, \dots \rangle$ . Par conséquent,  $\mathcal{B}_{-1}(z) = 1 + z\mathcal{B}_2(-z)$ . On observe aussi que  $\mathcal{B}_{-1}(z) = \mathcal{B}_2(-z)^{-1}$ .

Pour terminer cette section, nous allons démontrer une conséquence importante des équations (5.72) et (5.73). C'est une relation qui lie d'une autre manière les fonctions  $\mathcal{B}_{-1}(z)$  et  $\mathcal{B}_2(-z)$  :

$$\frac{\mathcal{B}_{-1}(z)^{n+1} - (-z)^{n+1}\mathcal{B}_2(-z)^{n+1}}{\sqrt{1+4z}} = \sum_{k \leq n} \binom{n-k}{k} z^k.$$

C'est vrai car le coefficient de  $z^k$  dans  $(-z)^{n+1}\mathcal{B}_2(-z)^{n+1}/\sqrt{1+4z}$ , lorsque  $k > n$ , est

$$\begin{aligned} [z^k] \frac{(-z)^{n+1}\mathcal{B}_2(-z)^{n+1}}{\sqrt{1+4z}} &= (-1)^{n+1} [z^{k-n-1}] \frac{\mathcal{B}_2(-z)^{n+1}}{\sqrt{1+4z}} \\ &= (-1)^{n+1} (-1)^{k-n-1} [z^{k-n-1}] \frac{\mathcal{B}_2(z)^{n+1}}{\sqrt{1-4z}} \\ &= (-1)^k \binom{2(k-n-1)+n+1}{k-n-1} \\ &= (-1)^k \binom{2k-n-1}{k-n-1} = (-1)^k \binom{2k-n-1}{k} \\ &= \binom{n-k}{k} = [z^k] \frac{\mathcal{B}_{-1}(z)^{n+1}}{\sqrt{1+4z}}. \end{aligned}$$

Les termes ont le bon goût de se neutraliser mutuellement. Nous pouvons maintenant appliquer (5.68) et (5.69) pour obtenir une forme close :

$$\sum_{k \leq n} \binom{n-k}{k} z^k = \frac{1}{\sqrt{1+4z}} \left( \left( \frac{1+\sqrt{1+4z}}{2} \right)^{n+1} - \left( \frac{1-\sqrt{1+4z}}{2} \right)^{n+1} \right),$$

$n \geq 0$  entier. (5.74)

(Nous avons rencontré le cas particulier  $z = -1$  dans le problème numéro 3 de la section 5.2. C'est parce que les nombres  $\frac{1}{2}(1 \pm \sqrt{-3})$  sont des racines sixièmes de l'unité que les sommes  $\sum_{k \leq n} \binom{n-k}{k} (-1)^k$  ont le comportement périodique que nous avons observé dans ce problème). De manière similaire, en combinant (5.70) et (5.71), on obtient

$$\sum_{k < n} \binom{n-k}{k} \frac{n}{n-k} z^k = \left( \frac{1+\sqrt{1+4z}}{2} \right)^n + \left( \frac{1-\sqrt{1+4z}}{2} \right)^n,$$

$n > 0$  entier. (5.75)

## 5.5 FONCTIONS HYPERGÉOMÉTRIQUES

Bien que les méthodes que nous avons vues jusqu'ici dans ce chapitre soient indiscutablement efficaces, nous sommes bien obligés d'admettre qu'elles s'apparentent plutôt à des "trucs" ad hoc qu'à des techniques vérifiables. Lorsque nous travaillons sur un problème, nous nous trouvons souvent en face de plusieurs choix possibles, et nous devons tâtonner pour trouver le bon. Les coefficients binomiaux, comme les caméléons, sont maîtres dans l'art du camouflage. Il est donc naturel de se demander s'il n'existe pas une sorte de principe unificateur permettant de manipuler de façon uniforme une large classe de sommes de coefficients binomiaux. Eh bien, cela existe en effet. A la base de ce principe unificateur, on trouve la théorie des séries hypergéométriques.

L'étude des séries hypergéométriques a été lancée il y a bien longtemps par Euler, Gauss et Riemann. Elles sont encore l'objet d'énormément de recherches. Hélas, leur notation est plutôt compliquée, et il faut un certain temps pour bien s'y faire.

La série hypergéométrique générale est une série entière en  $z$  ayant  $m+n$  paramètres. Elle peut se définir en termes de puissances factorielles montantes :

$$F \left( \begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z \right) = \sum_{k \geq 0} \frac{a_1^{\bar{k}} \dots a_m^{\bar{k}}}{b_1^{\bar{k}} \dots b_n^{\bar{k}}} \frac{z^k}{k!}. \quad (5.76)$$

Pour éviter une division par zéro, il faut qu'aucun des  $b$  ne soit un entier négatif ou nul. A part cela, il n'y a aucune restriction sur les  $a$  et

*Ils font encore mieux que les caméléons : ils peuvent changer d'apparence même après qu'on les a disséqués.*

*Si quelque chose survit pendant des siècles malgré une notation si rebutante, c'est que ce quelque chose doit être vraiment utile.*

les  $b$ . On peut aussi utiliser la notation " $F(a_1, \dots, a_m; b_1, \dots, b_n; z)$ " pour la même fonction lorsque cela s'avère plus pratique, notamment pour des raisons typographiques. Les variables  $a$  sont les *paramètres du haut* ; elles apparaissent dans les numérateurs des termes de  $F$ . Les variables  $b$  sont les *paramètres du bas*, qui interviennent dans les dénominateurs. La variable  $z$  est appelée l'*argument*.

Dans les ouvrages de référence standard, on voit souvent la notation " ${}_mF_n$ ", à la place de " $F$ " pour désigner une fonction hypergéométrique ayant  $m$  paramètres en haut et  $n$  paramètres en bas. Toutefois, ces indices ont tendance à encombrer les formules et à nous faire perdre du temps lorsqu'on les manipule. De plus, ils sont redondants, car il suffit, pour connaître  $m$  et  $n$ , de compter les paramètres de la fonction.

Nous connaissons beaucoup de fonctions importantes qui sont en fait des fonctions hypergéométriques particulières. C'est d'ailleurs pour cela que ces dernières sont si puissantes. Le cas particulier le plus simple est celui où  $m = n = 0$  : il n'y a aucun paramètre, et on obtient la série bien connue

$$F\left(\begin{array}{|c} \\ \hline \end{array} \middle| z\right) = \sum_{k \geq 0} \frac{z^k}{k!} = e^z.$$

Après tout, cette notation est peut-être un peu perturbante lorsque  $m = 0$  ou  $n = 0$ . Convenons donc de corriger cela en ajoutant un 1 en haut et en bas :

$$F\left(\begin{array}{|c} 1 \\ \hline 1 \end{array} \middle| z\right) = e^z.$$

Cela a bien évidemment un sens, car la fonction n'est pas modifiée si on ajoute deux paramètres identiques en haut et en bas ou si on insère un même paramètre en haut et en bas.

Le cas le plus simple après le précédent est celui où  $m = 1$ ,  $a_1 = 1$  et  $n = 0$ . Comme précédemment, pour que  $n > 0$ , modifions un peu les paramètres de sorte que  $m = 2$ ,  $a_1 = a_2 = 1$ ,  $n = 1$  et  $b_1 = 1$ . Du fait que  $1^k = k!$ , la série obtenue nous est elle aussi familière :

$$F\left(\begin{array}{|c} 1, 1 \\ \hline 1 \end{array} \middle| z\right) = \sum_{k \geq 0} z^k = \frac{1}{1 - z}.$$

C'est notre vieille connaissance, la série géométrique. C'est justement parce qu'elle admet la série géométrique  $F(1, 1; 1; z)$  comme cas particulier que la série générale  $F(a_1, \dots, a_m; b_1, \dots, b_n; z)$  est dite hypergéométrique.

A l'aide de (5.56), on trouve facilement une forme close pour le cas

assez général où  $m = 1$  et  $n = 0$  :

$$F\left(\begin{matrix} a, 1 \\ 1 \end{matrix} \middle| z\right) = \sum_{k \geq 0} a^{\bar{k}} \frac{z^k}{k!} = \sum_k \binom{a+k-1}{k} z^k = \frac{1}{(1-z)^a} \quad (5.77)$$

En remplaçant  $a$  par  $-a$  et  $z$  par  $-z$ , on retrouve la formule du binôme :

$$F\left(\begin{matrix} -a, 1 \\ 1 \end{matrix} \middle| -z\right) = (1+z)^a.$$

En mettant un entier négatif en paramètre du haut, on est assuré que la série infinie devienne en fait finie, car  $(-a)^{\bar{k}} = 0$  pour tout entier  $a$  tel que  $k > a \geq 0$ .

Dans le cas où  $m = 0$  et  $n = 1$ , on obtient aussi une série fameuse, qu'on ne voit pourtant pas beaucoup dans la littérature concernant les mathématiques discrètes :

$$F\left(\begin{matrix} 1 \\ b, 1 \end{matrix} \middle| z\right) = \sum_{k \geq 0} \frac{(b-1)!}{(b-1+k)!} \frac{z^k}{k!} = I_{b-1}(2\sqrt{z}) \frac{(b-1)!}{z^{(b-1)/2}}. \quad (5.78)$$

La fonction  $I_{b-1}$  est une "fonction de Bessel modifiée" d'ordre  $b-1$ . Si on pose  $b = 1$ , on obtient  $F\left(\begin{matrix} 1 \\ 1, 1 \end{matrix} \middle| z\right) = I_0(2\sqrt{z})$ , qui se trouve être l'intéressante série  $\sum_{k \geq 0} z^k/k!$ .

La série obtenue pour  $m = n = 1$ , que l'on note généralement par la lettre  $M$ , est appelée une "série hypergéométrique confluente" :

$$F\left(\begin{matrix} a \\ b \end{matrix} \middle| z\right) = \sum_{k \geq 0} \frac{a^{\bar{k}}}{b^{\bar{k}}} \frac{z^k}{k!} = M(a, b, z). \quad (5.79)$$

Cette fonction, qui a été découverte par Ernst Kummer, a d'importantes applications dans les sciences de l'ingénieur.

A ce stade, certains d'entre nous se demandent certainement pourquoi nous n'avons pas encore étudié la convergence de la série infinie (5.76). En voici la raison : on peut totalement ignorer ce problème de convergence si on considère  $z$  comme un symbole formel. Il n'est pas difficile de vérifier que les sommes formelles infinies de la forme  $\sum_{k \geq n} \alpha_k z^k$  forment un corps si les coefficients  $\alpha_k$  appartiennent à un corps. On peut additionner, soustraire, multiplier, diviser, dériver, composer de telles sommes formelles sans se poser de problème de convergence ; les identités qui en résulteront n'en seront pas moins vraies. Par exemple, bien que la fonction hypergéométrique  $F\left(\begin{matrix} 1, 1, 1 \\ 1 \end{matrix} \middle| z\right) = \sum_{k \geq 0} k! z^k$  diverge pour tout  $z$  non nul, nous verrons au chapitre 7 qu'on peut néanmoins l'utiliser pour résoudre des problèmes. Cependant, si on remplace  $z$  dans une série formelle par une valeur numérique particulière, il faut absolument être sûr qu'elle est bien définie dans ce cas.

*Nous n'avons pas non plus étudié la convergence de (5.56), (5.58) ...*

"There must be many universities to-day where 95 per cent, if not 100 per cent, of the functions studied by physics, engineering, and even mathematics students, are covered by this single symbol  $F(a, b; c; x)$ ."  
—W. W. Sawyer [318]

Voyons maintenant la plus célèbre de toutes les séries hypergéométriques. C'était même la série hypergéométrique jusqu'au jour où, vers 1870, on l'a généralisée à tout  $m$  et tout  $n$ . Elle possède deux paramètres en haut et un en bas :

$$F\left(\begin{matrix} a, b \\ c \end{matrix} \middle| z\right) = \sum_{k \geq 0} \frac{a^{\bar{k}} b^{\bar{k}} z^k}{c^{\bar{k}} k!}. \quad (5.80)$$

Elle est souvent appelée la "série hypergéométrique gaussienne", car c'est Gauss qui a, en 1812, démontré le premier beaucoup de ses subtiles propriétés [143], bien que Euler [118] et Pfaff [292] eussent déjà découvert des résultats remarquables à son sujet. Voici un cas particulier important de cette série :

$$\begin{aligned} \ln(1+z) &== z F\left(\begin{matrix} 1, 1 \\ 2 \end{matrix} \middle| -z\right) \\ &= z \sum_{k \geq 0} \frac{k! k!}{(k+1)!} \frac{(-z)^k}{k!} \\ &= z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots. \end{aligned}$$

Remarquez que  $z^{-1} \ln(1+z)$  est une fonction hypergéométrique, mais que  $\ln(1+z)$  ne peut pas l'être car une série hypergéométrique doit toujours prendre la valeur 1 lorsque  $z = 0$ .

Jusqu'à présent, les fonctions hypergéométriques ne nous ont pas servi à grand chose, à part à faire une liste de formules sans grand rapport les unes avec les autres. C'est justement cela qui va nous servir à présent : nous savons que beaucoup de fonctions très différentes les unes des autres peuvent être vues comme des séries hypergéométriques. Nous allons voir qu'il existe une large classe de sommes que l'on peut écrire de façon "canonique" sous forme de séries hypergéométriques. Cela nous permettra de classer en quelque sorte les résultats que nous connaissons sur les coefficients binomiaux.

Quelles sont les séries qui sont hypergéométriques ? Pour répondre à cette question, regardons le rapport de deux termes consécutifs d'une série hypergéométrique :

$$F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) = \sum_{k \geq 0} t_k, \quad t_k = \frac{a_1^{\bar{k}} \dots a_m^{\bar{k}} z^k}{b_1^{\bar{k}} \dots b_n^{\bar{k}} k!}.$$

Le premier terme est  $t_0 = 1$ , et les rapports des termes suivants sont donnés

par la formule

$$\begin{aligned}\frac{t_{k+1}}{t_k} &= \frac{a_1^{\overline{k+1}} \dots a_m^{\overline{k+1}}}{a_1^{\overline{k}} \dots a_m^{\overline{k}}} \frac{b_1^{\overline{k}} \dots b_n^{\overline{k}}}{b_1^{\overline{k+1}} \dots b_n^{\overline{k+1}}} \frac{k!}{(k+1)!} \frac{z^{k+1}}{z^k} \\ &= \frac{(k+a_1) \dots (k+a_m) z}{(k+b_1) \dots (k+b_n)(k+1)}. \end{aligned} \quad (5.81)$$

On obtient une *fonction rationnelle* de  $k$ , c'est-à-dire un quotient de polynômes en  $k$ . D'après le Théorème Fondamental de l'Algèbre, toute fonction rationnelle de  $k$  peut être factorisée sur le corps des nombres complexes pour être écrite sous la forme ci-dessus. Les variables  $a$  sont les opposés des racines du polynôme du numérateur, et les  $b$  sont les opposés des racines du dénominateur. Si le dénominateur ne contient pas le facteur particulier  $(k+1)$ , il est toujours possible de l'ajouter à la fois dans le numérateur et le dénominateur. Il reste un facteur constant qu'on peut appeler  $z$ . Par conséquent, les séries hypergéométriques sont exactement les séries dont le premier terme est égal à 1 et dont le rapport de deux termes consécutifs  $t_{k+1}/t_k$  est une fonction rationnelle de  $k$ .

Supposons par exemple qu'on nous donne une série infinie dont le rapport de deux termes consécutifs est la fonction rationnelle de  $k$  suivante :

$$\frac{t_{k+1}}{t_k} = \frac{k^2 + 7k + 10}{4k^2 + 1}.$$

Le polynôme du numérateur se factorise gentiment en deux facteurs  $(k+2)(k+5)$ , et le dénominateur en  $4(k+i/2)(k-i/2)$ . Ajoutons le facteur  $(k+1)$  qui manque, pour arriver à

$$\frac{t_{k+1}}{t_k} = \frac{(k+2)(k+5)(k+1)(1/4)}{(k+i/2)(k-i/2)(k+1)}.$$

Nous pouvons alors écrire explicitement la série recherchée :

$$\sum_{k \geq 0} t_k = t_0 F\left(\begin{matrix} 2, 5, 1 \\ i/2, -i/2 \end{matrix} \middle| 1/4\right).$$

Nous disposons donc d'une méthode générale pour trouver la représentation hypergéométrique d'une quantité donnée  $S$ , lorsque cette représentation existe. On commence par écrire  $S$  comme une série infinie dont le premier terme est non nul, par exemple  $\sum_{k \geq 0} t_k$  avec  $t_0 \neq 0$ . Puis on calcule  $t_{k+1}/t_k$ . Si ce rapport n'est pas une fonction rationnelle de  $k$ , c'est perdu. Sinon, on l'exprime sous la forme (5.81) ; on obtient ainsi des paramètres  $a_1, \dots, a_m, b_1, \dots, b_n$  et un argument  $z$  tels que  $S = t_0 F(a_1, \dots, a_m; b_1, \dots, b_n; z)$ .

*(C'est le moment de faire l'exercice d'échauffement 11.)*

Voici une façon d'écrire la série hypergéométrique de Gauss

$$F\left(\begin{matrix} a, b \\ c \end{matrix} \middle| z\right) = 1 + \frac{a}{1} \frac{b}{c} z \left( 1 + \frac{a+1}{2} \frac{b+1}{c+1} z \left( 1 + \frac{a+2}{3} \frac{b+2}{c+2} z (1 + \dots) \right) \right)$$

qui permet de mettre en évidence l'importance des rapports de termes successifs.

Essayons maintenant de reformuler les identités sur le coefficients binomiaux que nous avons vues dans ce chapitre en termes de séries hypergéométriques. Voyons par exemple ce que cela donne pour la règle de sommation parallèle

$$\sum_{k \leq n} \binom{r+k}{k} = \binom{r+n+1}{n}, \quad n \text{ entier.}$$

Commençons par remplacer  $k$  par  $n - k$  pour exprimer la somme comme une série infinie qui démarre à  $k = 0$  :

$$\sum_{k \geq 0} \binom{r+n-k}{n-k} = \sum_{k \geq 0} \frac{(r+n-k)!}{r! (n-k)!} = \sum_{k \geq 0} t_k.$$

Cette série, qui est infinie formellement, est en réalité finie, car, en raison du  $(n-k)!$  au dénominateur,  $t_k$  sera nul pour tout  $k > n$ . (Nous verrons plus tard que  $1/x!$  est défini pour tout  $x$  et que  $1/x! = 0$  lorsque  $x$  est strictement négatif. Evitons pour l'instant ces considérations techniques ; nous y reviendrons lorsque nous connaîtrons mieux les séries hypergéométriques). Le rapport des termes est égal à

$$\begin{aligned} \frac{t_{k+1}}{t_k} &= \frac{(r+n-k-1)! r! (n-k)!}{r! (n-k-1)! (r+n-k)!} = \frac{n-k}{r+n-k} \\ &= \frac{(k+1)(k-n)(1)}{(k-n-r)(k+1)}. \end{aligned}$$

Nous savons de plus que  $t_0 = \binom{r+n}{n}$ . Par conséquent, la règle de sommation parallèle est équivalente à l'identité hypergéométrique

$$\binom{r+n}{n} F\left(\begin{matrix} 1, -n \\ -n-r \end{matrix} \middle| 1\right) = \binom{r+n+1}{n}.$$

En divisant le tout par  $\binom{r+n}{n}$ , on obtient une version plus simple :

$$F\left(\begin{matrix} 1, -n \\ -n-r \end{matrix} \middle| 1\right) = \frac{r+n+1}{r+1}, \quad \text{si } \binom{r+n}{n} \neq 0. \quad (5.82)$$

Passons à une autre identité. Le rapport des termes de (5.16),

$$\sum_{k \leq m} \binom{r}{k} (-1)^k = (-1)^m \binom{r-1}{m}, \quad m \text{ entier,}$$

est égal à  $(k-m)/(r-m+k+1) = (k+1)(k-m)(1)/(k-m+r+1)(k+1)$ , une fois qu'on a remplacé  $k$  par  $m-k$ . On en déduit que (5.16) donne une forme close de

$$F\left(\begin{matrix} 1, -m \\ -m+r+1 \end{matrix} \middle| 1\right).$$

On reconnaît la fonction hypergéométrique du membre gauche de (5.82), dans laquelle on a remplacé  $n$  par  $m$  et  $-r$  par  $r+1$ . Ainsi, l'identité (5.16) est une conséquence immédiate de (5.82), la version hypergéométrique de (5.9). (Pas étonnant qu'il ait été facile de démontrer (5.16) à partir de (5.9)).

Avant d'aller plus loin, n'oublions pas de penser aux cas dégénérés. Les séries hypergéométriques ne sont pas définies lorsque l'un des paramètres du bas est un entier négatif ou nul. Si  $r$  et  $n$  sont des entiers strictement positifs, on peut appliquer la règle de sommation parallèle ; mais, dans ce cas,  $-n-r$  est un entier strictement négatif et la série hypergéométrique (5.76) n'est pas définie. Comment alors peut-on considérer (5.82) comme légitime ? Tout simplement en prenant la limite de  $F\left(\begin{matrix} 1, -n \\ -n-r+\epsilon \end{matrix} \middle| 1\right)$  lorsque  $\epsilon \rightarrow 0$ .

Nous verrons ce genre de considérations plus en détail un peu plus loin. Retenons simplement pour l'instant qu'il y a des dénominateurs à manier avec d'extrêmes précautions. Il est toutefois intéressant que la toute première somme que nous avons essayé d'exprimer en termes de fonctions hypergéométrique s'avère dégénérée.

Il y a un autre point sensible dans notre calcul de (5.82) : c'est lorsque nous avons réécrit  $\binom{r+n-k}{n-k}$  en  $(r+n-k)!/r!(n-k)!$ . Ce développement n'est pas autorisé si  $r$  est un entier strictement négatif, car il faut que  $(-m)!$  soit  $\infty$  pour que la règle

$$0! = 0 \cdot (-1) \cdot (-2) \cdot \dots \cdot (-m+1) \cdot (-m)!$$

puisse s'appliquer. Encore une fois, il nous faut recourir à la limite de  $r+\epsilon$  lorsque  $\epsilon \rightarrow 0$  pour obtenir des résultats cohérents pour les entiers

Il y a cependant encore un problème : la factorielle, et a fortiori la représentation factorielle  $\binom{r}{k} = r!/k!(r-k)!$  du binomial, ne sont définies que si  $r$  est un entier ! Pour pouvoir réellement travailler avec les séries hypergéométriques, il nous faut une factorielle généralisée aux nombres complexes. Justement cette fonction existe. Elle peut être définie de diverses manières. Voici donc une définition de  $z!$  (en fait une définition de  $1/z!$ ), qui est l'une des plus pratiques :

$$\frac{1}{z!} = \lim_{n \rightarrow \infty} \binom{n+z}{n} n^{-z}. \quad (5.83)$$

*Tout à l'heure les dérangements, maintenant les dégénérés. Ça ne s'arrange pas, dites donc.*

*(A l'origine, nous avons prouvé les identités pour  $r$  entier, puis nous avons invoqué l'argument polynomial pour les étendre à tout réel ; et maintenant, nous les prouvons d'abord pour tout  $r$  irrationnel, puis nous invoquons la limite pour montrer qu'elles s'appliquent aussi aux entiers !)*

(Voir l'exercice 21. Euler [99, 100, 72] découvrit cette fonction à l'âge de 22 ans). On peut montrer que la limite est définie pour tout nombre complexe  $z$  et qu'elle est nulle seulement si  $z$  est un entier négatif. Voici une autre définition possible :

$$z! = \int_0^\infty t^z e^{-t} dt, \quad \text{si } \Re z > -1. \quad (5.84)$$

Cette intégrale ne peut exister que si la partie réelle de  $z$  est strictement supérieure à  $-1$ . Toutefois, il suffit d'appliquer la formule

$$z! = z(z-1)! \quad (5.85)$$

pour étendre la définition à tous les complexes (sauf les entiers strictement négatifs). Une autre définition possible fait appel à l'interpolation de Stirling pour  $\ln z!$  que nous avons vue en (5.47). Toutes ces approches sont en fait équivalentes, et conduisent donc à la même fonction factorielle généralisée.

Il existe une fonction très similaire, appelée la *fonction Gamma*. Elle est à la factorielle ordinaire ce que les puissances montantes sont aux puissances descendantes. Généralement, les ouvrages de référence standard utilisent simultanément la factorielle et la fonction Gamma. On peut passer de l'une à l'autre avec les formules de conversion suivantes :

$$\Gamma(z+1) = z! ; \quad (5.86)$$

$$(-z)! \Gamma(z) = \frac{\pi}{\sin \pi z}. \quad (5.87)$$

Comment écrivez-vous  $z$  puissance  $\bar{w}$ , lorsque  $\bar{w}$  est le conjugué de  $w$  ?

$$z^{(\bar{w})}.$$

Je vois. L'indice du bas atteint sa limite le premier. C'est pour cela que  $\binom{z}{w}$  est nul lorsque  $w$  est un entier strictement négatif. En revanche, la valeur est infinie lorsque  $z$  est un entier strictement négatif et  $w$  n'est pas entier.

Avec les factorielles généralisées, on peut définir des puissances factorielles généralisées à tous nombres complexes  $z$  et  $w$  :

$$z^w = \frac{z!}{(z-w)!}; \quad (5.88)$$

$$z^{\bar{w}} = \frac{\Gamma(z+w)}{\Gamma(z)}. \quad (5.89)$$

La seule condition est qu'il convient de prendre des valeurs limites appropriées lorsque ces formules donnent  $\infty/\infty$ . (Elles ne peuvent pas donner  $0/0$  car la factorielle et la fonction Gamma ne sont jamais nulles). Tout coefficient binomial peut s'écrire

$$\binom{z}{w} = \lim_{\zeta \rightarrow z} \lim_{\omega \rightarrow w} \frac{\zeta!}{\omega! (\zeta - \omega)!} \quad (5.90)$$

quels que soient les nombres complexes  $z$  et  $w$ .

Maintenant que nous sommes équipés de notre factorielle généralisée, revenons à notre propos : mettre en évidence l'essence hypergéométrique des identités que nous avons démontrées. Nous savons déjà que la formule du binôme (5.13) n'est rien d'autre que (5.77), à peu de chose près. La prochaine identité à examiner est donc la convolution de Vandermonde (5.27) :

$$\sum_k \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}, \quad n \text{ entier.}$$

Le  $k$ ième terme est égal à

$$t_k = \frac{r!}{(r-k)! k!} \frac{s!}{(s-n+k)! (n-k)!}.$$

Nous pouvons maintenant nous permettre d'utiliser les factorielles généralisées. Pour chaque facteur de type  $(\alpha+k)!$  (avec un signe plus devant le  $k$ ) contenu dans  $t_k$ , on obtient, d'après (5.85),  $(\alpha+k+1)!/(\alpha+k)! = k+\alpha+1$  dans le rapport des termes  $t_{k+1}/t_k$ . Ceci entraîne qu'il y a un paramètre " $\alpha+1$ " dans la série hypergéométrique correspondante, paramètre du haut si  $(\alpha+k)!$  est dans le numérateur de  $t_k$ , paramètre du bas sinon. De façon similaire, un facteur de type  $(\alpha-k)!$  dans  $t_k$  entraîne une expression  $(\alpha-k-1)!/(\alpha-k)! = (-1)/(k-\alpha)$  dans le rapport des termes. Ceci induit un paramètre " $-\alpha$ " dans la série hypergéométrique (en bas si  $(\alpha-k)!$  est au numérateur, en haut sinon) et entraîne un changement de signe de l'argument. Les facteurs comme  $r!$ , qui sont indépendants de  $k$ , sont présents dans  $t_0$  mais disparaissent dans le rapport des termes. Ces considérations nous permettent d'affirmer sans plus de calculs que le rapport des termes de (5.27) est

$$\frac{t_{k+1}}{t_k} = \frac{k-r}{k+1} \frac{k-n}{k+s-n+1}$$

multiplié par  $(-1)^2 = 1$ . Par conséquent, la convolution de Vandermonde devient

$$\binom{s}{n} F\left(\begin{matrix} -r, -n \\ s-n+1 \end{matrix} \middle| 1\right) = \binom{r+s}{n}. \quad (5.91)$$

Cette équation peut être utilisée pour calculer  $F(a, b; c; z)$  lorsque  $z = 1$  et  $b$  est un entier négatif.

Réécrivons (5.91) de façon un peu plus lisible :

$$F\left(\begin{matrix} a, b \\ c \end{matrix} \middle| 1\right) = \frac{\Gamma(c-a-b)\Gamma(c)}{\Gamma(c-a)\Gamma(c-b)}; \quad \begin{aligned} &b \leq 0 \text{ entier} \\ &\text{ou } \Re c > \Re a + \Re b. \end{aligned} \quad (5.92)$$

*Il y a quelque semaines seulement, nous avons étudié ce que Gauss avait découvert à la maternelle. En ce moment, nous voyons ce qu'il a fait après sa thèse. C'est un peu rapide, non ?*

La convolution de Vandermonde (5.27) ne couvre que le cas où l'un des paramètres du haut, par exemple  $b$ , est un entier négatif ou nul. Gauss [143] a toutefois démontré que (5.92) est valide aussi lorsque  $a$ ,  $b$  et  $c$  sont des nombres complexes dont les parties réelles satisfont  $\Re c > \Re a + \Re b$ . Dans tous les autres cas, la série infinie  $F\left(\frac{a,b}{c} \mid 1\right)$  diverge. Si  $b = -n$ , l'identité peut s'écrire plus simplement, avec des puissances factorielles plutôt que des Gamma :

$$F\left(\begin{matrix} a, -n \\ c \end{matrix} \mid 1\right) = \frac{(c-a)^{\bar{n}}}{c^{\bar{n}}} = \frac{(a-c)^{\bar{n}}}{(-c)^{\bar{n}}}, \quad n \geq 0 \text{ entier.} \quad (5.93)$$

Nous pouvons maintenant affirmer que les cinq identités de la table 181 sont des cas particuliers de la convolution de Vandermonde, car elles sont toutes couvertes par la formule (5.93). Il convient pour s'en assurer de bien considérer les cas dégénérés.

Remarquez que (5.82) représente juste le cas particulier  $a = 1$  de (5.93). Nous n'avons donc pas besoin de retenir (5.82) ; ni d'ailleurs l'identité (5.9) qui nous a conduits à (5.82), même si la table 186 affirmait qu'il fallait la mémoriser. On peut calculer automatiquement  $\sum_{k \leq n} \binom{r+k}{k}$ , en convertissant la somme en fonction hypergéométrique puis en lui appliquant l'identité générale de la convolution de Vandermonde. Ce travail pourrait être fait sans problème par un programme informatique.

Dans le problème numéro 1 de la section 5.2, nous cherchions la valeur de

$$\sum_{k \geq 0} \binom{m}{k} / \binom{n}{k}.$$

Ce problème est naturellement d'essence hypergéométrique : avec un peu de pratique, tout hypergéomètre peut voir que cette expression est égale à  $F(1, -m; -n; 1)$ . C'est encore un cas particulier de la convolution de Vandermonde !

La somme des problèmes 2 et 4 s'écrit  $F(2, 1-n; 2-m; 1)$  (en remplaçant d'abord  $k$  par  $k+1$ ). La somme du problème 6, celle qui nous faisait froid dans le dos, n'est autre que  $F(n+1, -n; 2; 1)$ . On est en droit de se poser la question suivante : existe-t-il des sommes qui ne sont pas des versions camouflées de la convolution de Vandermonde ?

Eh bien oui. Celle du problème 3 est différente, justement. Il s'agit d'un cas particulier de la somme  $\sum_k \binom{n-k}{k} z^k$  que nous avons vue en (5.74), qui nous conduit à une forme close de

$$F\left(\begin{matrix} 1+2[n/2], -n \\ 1/2 \end{matrix} \mid -z/4\right).$$

Nous avons aussi démontré quelque chose de nouveau en (5.55), lorsque nous regardions les coefficients de  $(1-z)^r(1+z)^r$  :

$$F\left(\begin{matrix} 1-c-2n, -2n \\ c \end{matrix} \middle| -1\right) = (-1)^n \frac{(2n)!}{n!} \frac{(c-1)!}{(c+n-1)!}, \quad n \geq 0 \text{ entier.}$$

Lorsqu'on généralise cette équation aux nombres complexes, on trouve ce qu'on appelle la *formule de Kummer* :

*Kummer le Sommeur.*

$$F\left(\begin{matrix} a, b \\ 1+b-a \end{matrix} \middle| -1\right) = \frac{(b/2)!}{b!} (b-a)^{b/2}. \quad (5.94)$$

(Ernst Kummer [229] l'a démontrée en 1836).

Il est intéressant de comparer ces deux formules. Si on remplace  $c$  par  $1-2n-a$ , on trouve que les résultats sont cohérents si et seulement si

$$(-1)^n \frac{(2n)!}{n!} = \lim_{b \rightarrow -2n} \frac{(b/2)!}{b!} = \lim_{x \rightarrow -n} \frac{x!}{(2x)!} \quad (5.95)$$

lorsque  $n$  est un entier strictement positif. Supposons par exemple que  $n=3$ . Alors il faut que  $-6!/3! = \lim_{x \rightarrow -3} x!/(2x)!$ . Nous savons que les factorielles  $(-3)!$  et  $(-6)!$  sont toutes deux infinies. On pourrait être tenté d'ignorer cette difficulté en posant par exemple  $(-3)! = (-3)(-4)(-5)(-6)!$ , de façon que les deux occurrences de  $(-6)!$  se neutralisent mutuellement. Il faut absolument résister aux tentations de ce genre, car elles mènent à un résultat faux ! D'après (5.95), La limite de  $x!/(2x)!$  lorsque  $x \rightarrow -3$  n'est pas  $(-3)(-4)(-5)$ , mais  $-6!/3! = (-4)(-5)(-6)$ .

La bonne façon d'évaluer la limite de (5.95) consiste à utiliser l'équation (5.87), qui lie la factorielle d'un nombre négatif à la valeur de la fonction Gamma en l'opposé (positif) de ce nombre. Si on remplace  $x$  par  $-n-\epsilon$  et si on fait tendre  $\epsilon$  vers zéro, on obtient, après deux applications de (5.87),

$$\frac{(-n-\epsilon)!}{(-2n-2\epsilon)!} \frac{\Gamma(n+\epsilon)}{\Gamma(2n+2\epsilon)} = \frac{\sin(2n+2\epsilon)\pi}{\sin(n+\epsilon)\pi}.$$

Comme  $\sin(x+y) = \sin x \cos y + \cos x \sin y$ , ce rapport de sinus devient

$$\frac{\cos 2n\pi \sin 2\epsilon\pi}{\cos n\pi \sin \epsilon\pi} = (-1)^n (2+O(\epsilon)),$$

si on utilise les méthodes du chapitre 9. Par conséquent, d'après (5.86), on obtient

$$\lim_{\epsilon \rightarrow 0} \frac{(-n-\epsilon)!}{(-2n-2\epsilon)!} = 2(-1)^n \frac{\Gamma(2n)}{\Gamma(n)} = 2(-1)^n \frac{(2n-1)!}{(n-1)!} = (-1)^n \frac{(2n)!}{n!}$$

comme désiré.

Complétons ce panorama en exprimant en termes hypergéométriques les autres identités que nous avons vues dans ce chapitre. La somme à trois binomiaux de (5.29) peut s'écrire

$$\begin{aligned} & {}_F\left(\begin{matrix} 1-a-2n, 1-b-2n, -2n \\ a, b \end{matrix} \middle| 1\right) \\ &= (-1)^n \frac{(2n)! (a+b+2n-1)^{\bar{n}}}{n! a^{\bar{n}} b^{\bar{n}}}, \quad n \geq 0 \text{ entier.} \end{aligned}$$

Sa généralisation aux nombres complexes est appelée la *formule de Dixon*.

$${{}_F\left(\begin{matrix} a, b, c \\ 1+c-a, 1+c-b \end{matrix} \middle| 1\right)} = \frac{(c/2)!}{c!} \frac{(c-a)^{c/2} (c-b)^{c/2}}{(c-a-b)^{c/2}}, \quad (5.96)$$

$\Re a + \Re b < 1 + \Re c/2.$

La somme de trois binomiaux (5.28) est parmi les formules les plus générales que nous avons rencontrées. Elle nous donne l'*identité de Saalschütz* :

(Note historique :  
Saalschütz [315]  
a découvert indépendamment cette  
formule presque 100  
ans après qu'elle  
eût été publiée  
par Pfaff [292]. En  
prenant la limite  
lorsque  $n \rightarrow \infty$ , on  
trouve (5.92)).

Cette formule donne la valeur en  $z = 1$  de la série hypergéométrique générale ayant trois paramètres en haut et deux en bas, pourvu que l'un des paramètres du haut soit un entier négatif ou nul et que  $b_1 + b_2 = a_1 + a_2 + a_3 + 1$ . (Si la somme des paramètres du bas dépasse la somme des paramètres du haut de 2 unités au lieu de 1, on peut appliquer la formule de l'exercice 25 pour exprimer  $F(a_1, a_2, a_3; b_1, b_2; 1)$  en fonction de deux séries hypergéométriques qui satisfont l'identité de Saalschütz).

L'identité si durement gagnée du problème 8 de la section 5.2 équivaut à

$$\frac{1}{1+x} {}_F\left(\begin{matrix} x+1, n+1, -n \\ 1, x+2 \end{matrix} \middle| 1\right) = (-1)^n x^n x^{-n-1}.$$

Ce n'est en fait que le cas  $c = 1$  de l'identité de Saalschütz (5.97). Nous aurions vraiment économisé notre peine si nous avions directement utilisé les séries hypergéométriques !

La somme du problème 7, quant à elle, nous donne la formule

$${{}_F\left(\begin{matrix} n+1, m-n, 1, \frac{1}{2} \\ \frac{1}{2}m+1, \frac{1}{2}m+\frac{1}{2}, 2 \end{matrix} \middle| 1\right)} = \frac{m}{n}. \quad n \geq m > 0 \text{ entier.}$$

C'est la première fois que nous voyons trois paramètres en bas ; cette formule a donc l'air inédite. Ce n'est pourtant pas le cas : l'exercice 26 nous dit qu'on peut remplacer le membre gauche par un multiple de

$$F\left(\begin{matrix} n, m-n-1, -\frac{1}{2} \\ \frac{1}{2}m, \frac{1}{2}m-\frac{1}{2} \end{matrix} \middle| 1\right) = 1,$$

et c'est encore l'identité de Saalschütz qui remporte la mise. Voici une raison de plus d'apprécier la puissance des séries hypergéométriques.

Les identités de convolution de la table 216 n'ont pas d'équivalents hypergéométriques, car les rapports de leurs termes ne sont des fonctions rationnelles de  $k$  que si  $t$  est un nombre entier. Pire encore, les équations (5.65) et (5.66) ne sont même pas hypergéométriques lorsque  $t = 1$ . Remarquons cependant que, lorsque  $t$  est un petit entier, l'identité (5.63) nous dit que

*(Note historique : c'est George Andrews qui a, le premier, souligné l'importance des séries hypergéométriques pour les identités binomiales. C'était en 1974 [9, section 5].)*

$$\begin{aligned} F\left(\begin{matrix} \frac{1}{2}r, \frac{1}{2}r+\frac{1}{2}, -n, -n-s \\ r+1, -n-\frac{1}{2}s, -n-\frac{1}{2}s+\frac{1}{2} \end{matrix} \middle| 1\right) &= \binom{r+s+2n}{n} / \binom{s+2n}{n}; \\ F\left(\begin{matrix} \frac{1}{3}r, \frac{1}{3}r+\frac{1}{3}, \frac{1}{3}r+\frac{2}{3}, -n, -n-\frac{1}{2}s, -n-\frac{1}{2}s+\frac{1}{2} \\ \frac{1}{2}r+\frac{1}{2}, \frac{1}{2}r+1, -n-\frac{1}{3}s, -n-\frac{1}{3}s+\frac{1}{3}, -n-\frac{1}{3}s+\frac{2}{3} \end{matrix} \middle| 1\right) \\ &= \binom{r+s+3n}{n} / \binom{s+3n}{n}. \end{aligned}$$

On retrouve dans la première de ces formules le résultat du problème 7, en remplaçant respectivement  $(r, s, n)$  par  $(1, m-2n-1, n-m)$ .

Pour finir, la somme "inattendue" (5.20) donne lieu à une identité hypergéométrique inattendue aussi et très instructive. Voyons cela au ralenti. Commençons par convertir notre somme finie en une somme infinie :

$$\sum_{k \leq m} \binom{m+k}{k} 2^{-k} = 2^m \iff \sum_{k \geq 0} \binom{2m-k}{m-k} 2^k = 2^{2m}.$$

Le rapport des termes  $(2m-k)! 2^k / m! (m-k)!$  est égal à  $2(k-m)/(k-2m)$ . Nous en déduisons une identité hypergéométrique avec  $z = 2$  :

$$\binom{2m}{m} F\left(\begin{matrix} 1, -m \\ -2m \end{matrix} \middle| 2\right) = 2^{2m}, \quad m \geq 0 \text{ entier.} \quad (5.98)$$

Le paramètre du bas est " $-2m$ ". Comme les entiers négatifs sont absolument interdits à cet endroit, cette identité n'est pas définie !

Il est grand temps maintenant de nous pencher sérieusement, comme nous l'avions promis, sur ce genre de problèmes de limites. Pour évaluer une série hypergéométrique dégénérée, la bonne solution consiste souvent à prendre sa limite au voisinage du point qui pose problème. Ce faisant, il

faut toutefois prendre des précautions, car, selon la façon dont on calcule une limite, on peut trouver des résultats différents. Voici par exemple deux limites tout à fait différentes d'une même fonction en un même point :

$$\begin{aligned}\lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} -1+\epsilon, -3 \\ -2+\epsilon \end{matrix} \middle| 1\right) &= \lim_{\epsilon \rightarrow 0} \left(1 + \frac{(-1+\epsilon)(-3)}{(-2+\epsilon) 1!} + \frac{(-1+\epsilon)(\epsilon)(-3)(-2)}{(-2+\epsilon)(-1+\epsilon) 2!} \right. \\ &\quad \left. + \frac{(-1+\epsilon)(\epsilon)(1+\epsilon)(-3)(-2)(-1)}{(-2+\epsilon)(-1+\epsilon)(\epsilon) 3!}\right) \\ &= 1 - \frac{3}{2} + 0 + \frac{1}{2} = 0; \\ \lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} -1, -3 \\ -2+\epsilon \end{matrix} \middle| 1\right) &= \lim_{\epsilon \rightarrow 0} \left(1 + \frac{(-1)(-3)}{(-2+\epsilon) 1!} + 0 + 0\right) \\ &= 1 - \frac{3}{2} + 0 + 0 = -\frac{1}{2}.\end{aligned}$$

Pour trouver ces résultats différents, il a suffi d'ajouter  $\epsilon$  à l'un des paramètres du haut. De la même façon, nous avons défini  $\binom{-1}{-1} = 0 = \lim_{\epsilon \rightarrow 0} \binom{-1+\epsilon}{-1}$ , ce qui est tout à fait différent de  $\lim_{\epsilon \rightarrow 0} \binom{-1+\epsilon}{-1+\epsilon} = 1$ . Pour prendre la bonne limite de (5.98), il faut bien voir que le paramètre du haut  $-m$  est là pour que tous les termes de la série  $\sum_{k \geq 0} \binom{2m-k}{m-k} 2^k$  soient nuls si  $k > m$ . Par conséquent, c'est l'expression suivante qu'il nous faut :

$$\binom{2m}{m} \lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} 1, -m \\ -2m+\epsilon \end{matrix} \middle| 2\right) = 2^{2m}, \quad m \geq 0 \text{ entier.} \quad (5.99)$$

Tous les termes de cette limite sont bien définis, car le facteur  $(-2m)^{\overline{k}}$  du dénominateur n'est nul que si  $k > 2m$ . Cette série nous donne donc exactement la somme (5.20) du départ.

## 5.6 IDENTITÉS HYPERGÉOMÉTRIQUES

Maintenant, il est indubitable que si on dispose d'une base de données contenant les identités hypergéométriques connues, on est particulièrement bien outillé pour calculer des sommes de coefficients binomiaux. Pour évaluer une somme, on la met sous sa forme hypergéométrique canonique et on la compare au contenu de la base. Si elle y est, nous avons notre réponse ; sinon, on peut l'ajouter à la base si on sait l'exprimer en forme close. La base peut aussi contenir des informations du genre "cette somme n'a pas de forme close simple en général". Par exemple, la somme  $\sum_{k \leq m} \binom{n}{k}$  correspond à la série hypergéométrique

$$\binom{n}{m} F\left(\begin{matrix} 1, -m \\ n-m+1 \end{matrix} \middle| -1\right), \quad n \geq m \geq 0 \text{ entiers,} \quad (5.100)$$

qui n'admet une forme close que si  $m$  est proche de 0,  $\frac{1}{2}n$  ou  $n$ .

Cela ne s'arrête pas là : les fonctions hypergéométriques ont aussi leurs identités propres. Cela signifie que toute forme close d'une série hypergéométrique peut conduire à d'autres formes closes, donc à d'autres entrées de la base de données. Par exemple, les identités des exercices 25 et 26 nous montrent comment transformer une série hypergéométrique en deux autres séries ayant des paramètres similaires mais différents, qui peuvent elles-même être encore transformées.

En 1797, J. F. Pfaff [292] découvrit une surprenante *règle de réflexion*,

$$\frac{1}{(1-z)^a} F\left(\begin{matrix} a, b \\ c \end{matrix} \middle| \frac{-z}{1-z}\right) = F\left(\begin{matrix} a, c-b \\ c \end{matrix} \middle| z\right), \quad (5.101)$$

qui est une transformation d'un autre type. Si on remplace  $(-z)^k/(1-z)^{k+a}$  par la série infinie  $(-z)^k (1 + \binom{k+a}{1} z + \binom{k+a+1}{2} z^2 + \dots)$  après avoir développé le membre gauche, on obtient une identité formelle sur des séries (voir l'exercice 50). On peut appliquer cette loi pour trouver d'autres formules à partir de celles que nous connaissons déjà, si  $z \neq 1$ .

Par exemple, on peut combiner la formule de Kummer (5.94) avec la règle de réflexion (5.101) si on prend la précaution de choisir des paramètres tels que les deux puissent s'appliquer :

$$\begin{aligned} 2^{-a} F\left(\begin{matrix} a, 1-a \\ 1+b-a \end{matrix} \middle| \frac{1}{2}\right) &= F\left(\begin{matrix} a, b \\ 1+b-a \end{matrix} \middle| -1\right) \\ &= \frac{(b/2)!}{b!} (b-a)^{b/2}. \end{aligned} \quad (5.102)$$

Nous pouvons maintenant poser  $a = -n$  pour trouver une identité de coefficients binomiaux qui pourrait bien nous servir un jour :

$$\begin{aligned} \sum_{k \geq 0} \frac{(-n)^{\bar{k}} (1+n)^{\bar{k}}}{(1+b+n)^{\bar{k}}} \frac{2^{-k}}{k!} &= \sum_k \binom{n}{k} \left(\frac{-1}{2}\right)^k \binom{n+k}{k} / \binom{n+b+k}{k} \\ &= 2^{-n} \frac{(b/2)! (b+n)!}{b! (b/2+n)!}, \quad n \geq 0 \text{ entier.} \end{aligned} \quad (5.103)$$

Par exemple, lorsque  $n = 3$  cette identité donne

$$\begin{aligned} 1 - 3 \frac{4}{2(4+b)} + 3 \frac{4 \cdot 5}{4(4+b)(5+b)} - \frac{4 \cdot 5 \cdot 6}{8(4+b)(5+b)(6+b)} \\ = \frac{(b+3)(b+2)(b+1)}{(b+6)(b+4)(b+2)}. \end{aligned}$$

C'est presque incroyable, mais vrai pour tout  $b$  (sauf bien évidemment si un facteur du dénominateur est nul).

*La base de données hypergéométrique serait donc véritablement une "base de connaissances".*

Amusant, non ? Essayons encore une fois. Peut-être allons-nous trouver quelque chose qui pourra réellement impressionner nos amis. Que donne la loi de réflexion de Pfaff si on l'applique à l'étrange formule (5.99), dans laquelle  $z = 2$  ? Dans ce cas, il nous faut poser  $a = -m$ ,  $b = 1$  et  $c = -2m + \epsilon$ , obtenant ainsi

$$\begin{aligned} (-1)^m \lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} -m, 1 \\ -2m + \epsilon \end{matrix} \middle| 2\right) &= \lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} -m, -2m-1+\epsilon \\ -2m + \epsilon \end{matrix} \middle| 2\right) \\ &= \lim_{\epsilon \rightarrow 0} \sum_{k \geq 0} \frac{(-m)^{\bar{k}} (-2m-1+\epsilon)^{\bar{k}}}{(-2m+\epsilon)^{\bar{k}}} \frac{2^k}{k!} \\ &= \sum_{k \leq m} \binom{m}{k} \frac{(2m+1)^{\underline{k}}}{(2m)^{\underline{k}}} (-2)^k, \end{aligned}$$

car aucun des termes de la limite n'est proche de zéro. Ceci nous mène à une autre formule miraculeuse :

(Note hystérique : si vous trouvez un résultat différent, allez voir l'exercice 51).

$$\begin{aligned} \sum_{k \leq m} \binom{m}{k} \frac{2m+1}{2m+1-k} (-2)^k &= (-1)^m 2^{2m} / \binom{2m}{m} \\ &= 1 / \binom{-1/2}{m}, \quad m \geq 0 \text{ entier.} \quad (5.104) \end{aligned}$$

Pour  $m = 3$  par exemple, la somme vaut

$$1 - 7 + \frac{84}{5} - 14 = -\frac{16}{5},$$

et  $\binom{-1/2}{3}$  est en effet égal à  $-\frac{5}{16}$ .

Lorsque nous avons passé en revue nos identités binomiales pour les transformer en identités hypergéométriques, nous avons snobé l'équation (5.19) sous prétexte que ce n'était pas une forme close, mais une relation entre deux sommes. Maintenant, nous pouvons considérer (5.19) comme une identité entre des séries hypergéométriques. En la dérivant  $n$  fois par rapport à  $y$  puis en remplaçant  $k$  par  $m - n - k$ , on obtient

$$\begin{aligned} \sum_{k \geq 0} \binom{m+r}{m-n-k} \binom{n+k}{n} x^{m-n-k} y^k \\ = \sum_{k \geq 0} \binom{-r}{m-n-k} \binom{n+k}{n} (-x)^{m-n-k} (x+y)^k. \end{aligned}$$

Ceci donne lieu à l'identité hypergéométrique suivante :

$$F\left(\begin{matrix} a, -n \\ c \end{matrix} \middle| z\right) = \frac{(a-c)_n}{(-c)_n} F\left(\begin{matrix} a, -n \\ 1-n+a-c \end{matrix} \middle| 1-z\right), \quad n \geq 0 \text{ entier.} \quad (5.105)$$

Notez que si  $z = 1$ , on se ramène à la convolution de Vandermonde (5.93).

Cet exemple semble indiquer que la dérivation est quelque chose d'utile. Elle nous a d'ailleurs déjà aidés au chapitre 2, lorsqu'il nous fallait sommer  $x + 2x^2 + \dots + nx^n$ . Voyons donc ce qui se passe si on dérive une série hypergéométrique générale par rapport à  $z$  :

$$\begin{aligned} \frac{d}{dz} F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) &= \sum_{k \geq 1} \frac{a_1^{\bar{k}} \dots a_m^{\bar{k}} z^{k-1}}{b_1^{\bar{k}} \dots b_n^{\bar{k}} (k-1)!} \\ &= \sum_{k+1 \geq 1} \frac{a_1^{\bar{k+1}} \dots a_m^{\bar{k+1}} z^k}{b_1^{\bar{k+1}} \dots b_n^{\bar{k+1}} k!} \\ &= \sum_{k \geq 0} \frac{a_1(a_1+1)^{\bar{k}} \dots a_m(a_m+1)^{\bar{k}} z^k}{b_1(b_1+1)^{\bar{k}} \dots b_n(b_n+1)^{\bar{k}} k!} \\ &= \frac{a_1 \dots a_m}{b_1 \dots b_n} F\left(\begin{matrix} a_1+1, \dots, a_m+1 \\ b_1+1, \dots, b_n+1 \end{matrix} \middle| z\right). \quad (5.106) \end{aligned}$$

Les paramètres se déplacent et se décalent.

Il est aussi possible d'utiliser la dérivation pour agir juste sur un paramètre en laissant les autres inchangés. Pour cela, introduisons l'opérateur

$$\vartheta = z \frac{d}{dz},$$

qui agit sur une fonction en la dérivant puis en la multipliant par  $z$ . Cela donne

$$\vartheta F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) = z \sum_{k \geq 1} \frac{a_1^{\bar{k}} \dots a_m^{\bar{k}} z^{k-1}}{b_1^{\bar{k}} \dots b_n^{\bar{k}} (k-1)!} = \sum_{k \geq 0} \frac{k a_1^{\bar{k}} \dots a_m^{\bar{k}} z^k}{b_1^{\bar{k}} \dots b_n^{\bar{k}} k!},$$

ce qui n'est pas franchement utile en soi. Cependant, si on multiplie  $F$  par l'un de ses paramètres du haut, par exemple  $a_1$ , et si on ajoute le tout à  $\vartheta F$ , on obtient

$$\begin{aligned} (\vartheta + a_1) F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) &= \sum_{k \geq 0} \frac{(k+a_1) a_1^{\bar{k}} \dots a_m^{\bar{k}} z^k}{b_1^{\bar{k}} \dots b_n^{\bar{k}} k!}, \\ &= \sum_{k \geq 0} \frac{a_1(a_1+1)^{\bar{k}} a_2^{\bar{k}} \dots a_m^{\bar{k}} z^k}{b_1^{\bar{k}} \dots b_n^{\bar{k}} k!} \\ &= a_1 F\left(\begin{matrix} a_1+1, a_2, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right). \end{aligned}$$

Un seul paramètre a été déplacé et décalé.

Comment prononcez-vous  $\vartheta$ ?  
(Sais pas, mais TeX l'appelle "vartheta").

Il existe une astuce similaire pour les paramètres du bas. Dans ce cas, le décalage se fait dans l'autre sens :

$$\begin{aligned}
 (\vartheta + b_1 - 1) F \left( \begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z \right) &= \sum_{k \geq 0} \frac{(k + b_1 - 1) a_1^{\bar{k}} \dots a_m^{\bar{k}} z^k}{b_1^{\bar{k}} \dots b_n^{\bar{k}} k!}, \\
 &= \sum_{k \geq 0} \frac{(b_1 - 1) a_1^{\bar{k}} \dots a_m^{\bar{k}} z^k}{(b_1 - 1)^{\bar{k}} b_2^{\bar{k}} \dots b_n^{\bar{k}} k!} \\
 &= (b_1 - 1) F \left( \begin{matrix} a_1, \dots, a_m \\ b_1 - 1, b_2, \dots, b_n \end{matrix} \middle| z \right).
 \end{aligned}$$

Nous pouvons maintenant combiner toutes ces opérations pour faire un "jeu de mots" mathématique en exprimant la même chose de deux façons différentes. Nous avons donc

$$(\vartheta + a_1) \dots (\vartheta + a_m) F = a_1 \dots a_m F \left( \begin{matrix} a_1 + 1, \dots, a_m + 1 \\ b_1, \dots, b_n \end{matrix} \middle| z \right),$$

et

$$\begin{aligned}
 (\vartheta + b_1 - 1) \dots (\vartheta + b_n - 1) F \\
 &= (b_1 - 1) \dots (b_n - 1) F \left( \begin{matrix} a_1, \dots, a_m \\ b_1 - 1, \dots, b_n - 1 \end{matrix} \middle| z \right),
 \end{aligned}$$

avec  $F = F(a_1, \dots, a_m; b_1, \dots, b_n; z)$ . Or, d'après (5.106), la formule du haut est la dérivée de celle du bas. Par conséquent, la fonction hypergéométrique générale  $F$  satisfait l'équation différentielle

$$D(\vartheta + b_1 - 1) \dots (\vartheta + b_n - 1) F = (\vartheta + a_1) \dots (\vartheta + a_m) F, \quad (5.107)$$

où  $D$  désigne l'opérateur  $\frac{d}{dz}$ .

Un exemple ne nous fera pas de mal. Ecrivons donc l'équation différentielle satisfait par le série hypergéométrique 2-sur-1 standard,  $F(z) = F(a, b; c; z)$ . D'après (5.107), nous avons

$$D(\vartheta + c - 1) F = (\vartheta + a)(\vartheta + b) F.$$

En notation "ordinaire", qu'est-ce que ça donne ? On réécrit  $(\vartheta + c - 1) F$  en  $zF'(z) + (c - 1)F(z)$ , et on dérive pour obtenir le membre gauche :

$$F'(z) + zF''(z) + (c - 1)F'(z).$$

Dans le membre droit, on a

$$\begin{aligned} (\vartheta + a)(zF'(z) + bF(z)) &= z \frac{d}{dz}(zF'(z) + bF(z)) + a(zF'(z) + bF(z)) \\ &= zF'(z) + z^2F''(z) + bzF'(z) + azF'(z) + abF(z). \end{aligned}$$

Après regroupement des termes, on obtient l'équation suivante,

$$z(1 - z)F''(z) + (c - z(a + b + 1))F'(z) - abF(z) = 0. \quad (5.108)$$

qui est équivalente à la formule (5.107).

Réciproquement, il est possible de revenir à la série depuis l'équation différentielle. Supposons que  $F(z) = \sum_{k \geq 0} t_k z^k$  est une série satisfaisant (5.107). Il est alors facile de montrer que

$$\frac{t_{k+1}}{t_k} = \frac{(k + a_1) \dots (k + a_m)}{(k + b_1) \dots (k + b_n)(k + 1)};$$

donc  $F(z)$  est forcément égal à  $t_0 F(a_1, \dots, a_m; b_1, \dots, b_n; z)$ . Nous venons de démontrer que la série hypergéométrique (5.76) est l'unique série formelle qui satisfait l'équation différentielle (5.107) et dont le terme constant est égal à 1.

L'idéal serait que les séries hypergéométriques puissent résoudre toutes les équations différentielles du monde. Ce n'est hélas pas le cas. Le membre droit de (5.107) se développe toujours en une somme de termes de la forme  $\alpha_k z^k F^{(k)}(z)$ , où  $F^{(k)}(z)$  est la dérivée kième  $D^k F(k)$ . Le membre gauche, lui, se développe en une somme de termes de la forme  $\beta_k z^{k-1} F^{(k)}(z)$  avec  $k > 0$ . Ainsi, l'équation différentielle (5.107) est toujours de la forme

$$z^{n-1}(\beta_n - z\alpha_n)F^{(n)}(z) + \dots + (\beta_1 - z\alpha_1)F'(z) - \alpha_0 F(z) = 0.$$

L'équation (5.108) en est une illustration pour le cas  $n = 2$ . Réciproquement, nous montrerons dans l'exercice 6.13 que toutes les équations différentielles de ce type peuvent se factoriser en fonction de l'opérateur  $\vartheta$  pour donner des équations de la même forme que (5.107). Ce sont donc exactement les équations différentielles dont les solutions sont des séries dont le rapport des termes est rationnel.

Si on multiplie les deux membres de (5.107) par  $z$ , l'opérateur  $D$  s'en va et on obtient une équation très instructive :

$$\vartheta(\vartheta + b_1 - 1) \dots (\vartheta + b_n - 1)F = z(\vartheta + a_1) \dots (\vartheta + a_m)F. \quad (5.109)$$

Le premier facteur du membre gauche,  $\vartheta = (\vartheta + 1 - 1)$ , correspond au  $(k+1)$  du rapport des termes (5.81), qui correspond lui-même au  $k!$  du dénominateur du kième terme de la série hypergéométrique générale. Les autres

*La fonction  $F(z) = (1 - z)^r$  satisfait  $\vartheta F = z(\vartheta - r)F$ . Cela donne une autre preuve de la formule du binôme.*

facteurs  $(\vartheta + b_j - 1)$  correspondent au facteur  $(k + b_j)$  du dénominateur, qui correspond lui-même au  $b_j^{\bar{k}}$  de (5.76). Du côté droit, le  $z$  correspond à  $z^k$  et  $(\vartheta + a_j)$  correspond à  $a_j^{\bar{k}}$ .

On peut trouver et démontrer de nouvelles identités à l'aide de ces équations différentielles. Par exemple, on vérifie facilement que les deux séries hypergéométriques

$$F\left(\begin{matrix} 2a, 2b \\ a+b+\frac{1}{2} \end{matrix} \middle| z\right) \quad \text{et} \quad F\left(\begin{matrix} a, b \\ a+b+\frac{1}{2} \end{matrix} \middle| 4z(1-z)\right)$$

satisfont l'équation différentielle

$$z(1-z)F''(z) + (a+b+\frac{1}{2})(1-2z)F'(z) - 4abF(z) = 0.$$

Nous venons ainsi de démontrer l'*identité de Gauss* [143, équation 102]

$$F\left(\begin{matrix} 2a, 2b \\ a+b+\frac{1}{2} \end{matrix} \middle| z\right) = F\left(\begin{matrix} a, b \\ a+b+\frac{1}{2} \end{matrix} \middle| 4z(1-z)\right). \quad (5.110)$$

En particulier,

$$F\left(\begin{matrix} 2a, 2b \\ a+b+\frac{1}{2} \end{matrix} \middle| \frac{1}{2}\right) = F\left(\begin{matrix} a, b \\ a+b+\frac{1}{2} \end{matrix} \middle| 1\right), \quad (5.111)$$

à condition que ces deux sommes infinies convergent. En fait elles convergent toujours, sauf dans le cas dégénéré où  $a+b+\frac{1}{2}$  est un entier négatif ou nul.

Toute nouvelle identité sur les séries hypergéométriques a des conséquences sur les coefficients binomiaux, et celle-ci ne fait pas exception. Considérons la somme

$$\sum_{k \leq m} \binom{m-k}{n} \binom{m+n+1}{k} \left(\frac{-1}{2}\right)^k, \quad m \geq n \geq 0 \text{ entiers.}$$

Si  $0 \leq k \leq m-n$ , les termes sont non nuls et, en passant à la limite avec les précautions d'usage, on peut exprimer la somme comme la série hypergéométrique

$$\lim_{\epsilon \rightarrow 0} \binom{m}{n} F\left(\begin{matrix} n-m, -n-m-1+\alpha\epsilon \\ -m+\epsilon \end{matrix} \middle| \frac{1}{2}\right).$$

Comme le paramètre du haut  $n-m$  arrête la somme assez vite, la limite ne peut pas être affectée par la valeur de  $\alpha$ . Nous pouvons donc poser  $\alpha=2$  pour pouvoir appliquer (5.111). La limite peut maintenant être

*(Attention : on ne peut pas utiliser (5.110) en toute sécurité si  $|z| > 1/2$ , sauf si les deux membres sont des polynômes ; voir l'exercice 53.)*

calculée car le membre droit est un cas particulier de (5.92). On montre dans l'exercice 54 que le résultat peut s'exprimer sous la forme simplifiée

$$\sum_{k \leq m} \binom{m-k}{n} \binom{m+n+1}{k} \left(\frac{-1}{2}\right)^k = \binom{(m+n)/2}{n} 2^{n-m} [m+n \text{ est pair}], \quad \begin{matrix} m \geq n \geq 0 \\ \text{entiers.} \end{matrix} \quad (5.112)$$

Par exemple, si  $m = 5$  et  $n = 2$ , on trouve  $\binom{5}{2} \binom{8}{0} - \binom{4}{2} \binom{8}{1}/2 + \binom{3}{2} \binom{8}{2}/4 - \binom{2}{2} \binom{8}{3}/8 = 10 - 24 + 21 - 7 = 0$ ; si  $m = 4$  et  $n = 2$ , on trouve  $\frac{3}{4}$  des deux côtés.

On peut aussi trouver des cas où (5.110) donne des sommes de binomiaux lorsque  $z = -1$ , mais ces sommes s'avèrent très bizarres. Si on pose  $a = \frac{1}{6} - \frac{n}{3}$  et  $b = -n$ , on obtient la formule monstrueuse

$$F\left(\begin{matrix} \frac{1}{3} - \frac{2}{3}n, -2n \\ \frac{2}{3} - \frac{4}{3}n \end{matrix} \middle| -1\right) = F\left(\begin{matrix} \frac{1}{6} - \frac{1}{3}n, -n \\ \frac{2}{3} - \frac{4}{3}n \end{matrix} \middle| -8\right).$$

Si  $n \not\equiv 2 \pmod{3}$ , ces séries sont des polynômes non dégénérés. Les paramètres ont été particulièrement bien choisis pour que le membre gauche puisse être calculé avec (5.94). Nous arrivons ainsi à un résultat proprement sidérant,

$$\sum_k \binom{n}{k} \binom{\frac{1}{3}n - \frac{1}{6}}{k} 8^k / \binom{\frac{4}{3}n - \frac{2}{3}}{k} = \binom{2n}{n} / \binom{\frac{4}{3}n - \frac{2}{3}}{n}, \quad n \geq 0 \text{ entier, } n \not\equiv 2 \pmod{3}. \quad (5.113)$$

C'est l'identité binomiale la plus surprenante qu'il nous ait été donné de voir. Il est même difficile de la vérifier à la main sur des petites valeurs (pour  $n = 3$ , on trouve  $\frac{81}{7}$  dans les deux membres). Bien évidemment, elle est absolument inutile. Elle n'apparaîtra certainement jamais dans un problème pratique.

Nous savons maintenant que les séries hypergéométriques nous fournissent un outil de haut niveau pour nous aider à comprendre les sommes de binomiaux. On trouvera beaucoup d'informations complémentaires dans l'ouvrage de Bailey [18] et dans sa suite, par Gasper et Rahman [141].

*L'identité (5.113) est utile à une seule chose : elle démontre l'existence d'identités inutiles à un point inimaginable.*

## 5.7 SOMMES HYPERGÉOMÉTRIQUES FINIES

La plupart des sommes que nous avons calculées dans ce chapitre portent sur tous les indices  $k \geq 0$ . Dans certains cas toutefois, nous avons pu trouver une forme close pour un indice  $a \leq k < b$ . Par exemple, nous avons

vu en (5.16) que

$$\sum_{k < m} \binom{n}{k} (-1)^k = (-1)^{m-1} \binom{n-1}{m-1}, \quad m \text{ entier.} \quad (5.114)$$

La théorie développée au chapitre 2 nous fournit un bon moyen de comprendre les formules de ce genre : si  $f(k) = \Delta g(k) = g(k+1) - g(k)$ , alors on peut écrire  $\sum f(k) \delta k = g(k) + C$ , et

$$\sum_a^b f(k) \delta k = g(k) \Big|_a^b = g(b) - g(a).$$

Si  $a$  et  $b$  sont deux entiers tels que  $a \leq b$ , on peut aussi écrire

$$\sum_a^b f(k) \delta k = \sum_{a \leq k < b} f(k) = g(b) - g(a).$$

Ainsi, à l'identité (5.114) correspondent la sommation infinie

$$\sum \binom{n}{k} (-1)^k \delta k = (-1)^{k-1} \binom{n-1}{k-1} + C$$

et la différence

$$\Delta \left( (-1)^k \binom{n}{k} \right) = (-1)^{k+1} \binom{n+1}{k+1}.$$

Il est facile, en partant d'une fonction  $g(k)$ , de calculer la fonction  $\Delta g(k) = f(k)$ , dont la somme vaut  $g(k) + C$ . Il est bien plus difficile en revanche de partir de  $f(k)$  pour trouver sa somme indéfinie  $\sum f(k) \delta k = g(k) + C$ . Cette dernière peut très bien ne pas avoir de forme simple. Par exemple, la somme  $\sum \binom{n}{k} \delta k$  n'a apparemment pas de forme simple ; sinon, nous saurions évaluer des sommes du genre de  $\sum_{k \leq n/3} \binom{n}{k}$ , sur lesquelles nous sommes tout à fait impuissants ; ou alors, peut-être existe-t-il une forme simple de  $\sum \binom{n}{k} \delta k$  que nous n'avons pas encore trouvée. Comment savoir ?

En 1977, R. W. Gosper [154] a découvert une façon élégante de calculer des sommes indéfinies  $\sum f(k) \delta k = g(k) + C$  lorsque  $f$  et  $g$  appartiennent à une classe de fonctions que l'on appelle termes hypergéométriques. Convensions d'écrire

$$F \left( \begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z \right)_k = \frac{a_1^{\bar{k}} \dots a_m^{\bar{k}}}{b_1^{\bar{k}} \dots b_n^{\bar{k}}} \frac{z^k}{k!} \quad (5.115)$$

le  $k$ ième terme de la série hypergéométrique  $F(a_1, \dots, a_m; b_1, \dots, b_n; z)$ . Nous considérerons  $F(a_1, \dots, a_m; b_1, \dots, b_n; z)_k$  non comme une fonction

de  $z$ , mais comme une fonction de  $k$ . Il se trouve que, dans beaucoup de cas, il existe des paramètres  $c, A_1, \dots, A_M, B_1, \dots, B_N$  et  $Z$  tels que

$$\sum F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right)_k \delta k = c F\left(\begin{matrix} A_1, \dots, A_M \\ B_1, \dots, B_N \end{matrix} \middle| Z\right)_k + C, \quad (5.116)$$

pour  $a_1, \dots, a_m, b_1, \dots, b_n$  et  $z$  donnés. Nous dirons qu'une fonction  $F(a_1, \dots, a_m; b_1, \dots, b_n; z)_k$  est *sommable en termes hypergéométriques* si de telles constantes  $c, A_1, \dots, A_M, B_1, \dots, B_N, Z$  existent. L'algorithme de Gosper sait trouver ces constantes si elles existent ; dans le cas contraire, il prouve qu'elles n'existent pas.

On dit que  $t(k)$  est un *terme hypergéométrique* si  $t(k+1)/t(k)$  est une fonction rationnelle de  $k$  non identiquement nulle. Cela signifie essentiellement que  $t(k)$  est le produit d'une constante et d'un terme du même type que (5.115). (En fait, il faut ajouter à cela un petit détail technique pour régler le problème des zéros. Nous voulons en effet que  $t(k)$  ait un sens aussi lorsque  $k$  est négatif ou lorsque un ou plusieurs des  $b$  de (5.115) sont négatifs ou nuls. Formellement, un terme hypergéométrique s'obtient en multipliant (5.115) par un nombre constant non nul de puissances de 0, puis en simplifiant les zéros du numérateur et du dénominateur. Les exemples de l'exercice 12 peuvent aider à comprendre cette règle).

Supposons que nous cherchions  $\sum t(k) \delta k$ , où  $t(k)$  est un terme hypergéométrique. L'algorithme de Gosper procède en deux étapes très simples. La première étape consiste à exprimer le rapport des termes sous la forme particulière suivante :

$$\frac{t(k+1)}{t(k)} = \frac{p(k+1)}{p(k)} \frac{q(k)}{r(k+1)}, \quad (5.117)$$

où  $p, q$  et  $r$  sont des polynômes qui satisfont la condition

$$(k+\alpha) \nmid q(k) \quad \text{et} \quad (k+\beta) \nmid r(k) \\ \implies \alpha - \beta \text{ n'est pas un entier strictement positif.} \quad (5.118)$$

Cette condition est facile à obtenir : on commence par poser provisoirement  $p(k) = 1$ , et considérer que  $q(k)$  et  $r(k+1)$  sont respectivement le numérateur et le dénominateur, factorisés en facteurs linéaires, du rapport des termes. Par exemple, si  $t(k)$  est de la forme donnée dans (5.115), on démarre avec  $q(k) = (k+a_1) \dots (k+a_m)z$  et  $r(k) = (k+b_1-1) \dots (k+b_n-1)k$ . Puis on regarde si la condition (5.118) est respectée ou non. Si  $q$  et  $r$  contiennent respectivement des facteurs  $(k+\alpha)$  et  $(k+\beta)$  tels que  $\alpha - \beta = N > 0$ , on retire ces facteurs de  $q$  et  $r$  et on remplace  $p(k)$  par

$$p(k)(k+\alpha-1)^{\frac{N-1}{2}} = p(k)(k+\alpha-1)(k+\alpha-2) \dots (k+\beta+1). \quad (5.119)$$

(La divisibilité des polynômes est analogue à la divisibilité des entiers. Par exemple, l'expression  $(k+\alpha) \nmid q(k)$  signifie que le quotient  $q(k)/(k+\alpha)$  est un polynôme. Il est facile de voir que  $(k+\alpha) \nmid q(k)$  si et seulement si  $q(-\alpha) = 0$ .)

Les nouveaux polynômes  $p$ ,  $q$  et  $r$  satisfont (5.117), et le processus peut être répété jusqu'à ce que la condition (5.118) soit respectée. Nous verrons bientôt en quoi cette condition est importante.

La deuxième étape de l'algorithme de Gosper consiste à finir le travail, c'est-à-dire trouver, si possible, un terme hypergéométrique  $T(k)$  tel que

$$t(k) = T(k+1) - T(k). \quad (5.120)$$

Pour bien comprendre cette partie de l'algorithme, il nous faut d'abord aborder quelques notions théoriques. Gosper remarqua, après avoir étudié un certain nombre de cas particuliers, qu'il vaut mieux écrire la fonction  $T(k)$  sous la forme

$$T(k) = \frac{r(k)s(k)t(k)}{p(k)}, \quad (5.121)$$

(L'exercice 55 aide à comprendre pourquoi cette transformation est souhaitable).

où  $s(k)$  est une fonction inconnue qu'il faut trouver. En injectant (5.121) dans (5.120) et en appliquant (5.117), on obtient

$$\begin{aligned} t(k) &= \frac{r(k+1)s(k+1)t(k+1)}{p(k+1)} - \frac{r(k)s(k)t(k)}{p(k)} \\ &= \frac{q(k)s(k+1)t(k)}{p(k)} - \frac{r(k)s(k)t(k)}{p(k)}. \end{aligned}$$

Par conséquent, il faut que

$$p(k) = q(k)s(k+1) - r(k)s(k). \quad (5.122)$$

S'il existe un  $s(k)$  qui satisfait cette récurrence, alors on a trouvé  $\sum t(k) \delta k$ . Sinon,  $T$  n'existe pas.

Comme  $T(k)$  est supposé être un terme hypergéométrique, le rapport  $T(k+1)/T(k)$  est une fonction rationnelle de  $k$ . Par conséquent, d'après (5.121) et (5.122),  $r(k)s(k)/p(k) = T(k)/(T(k+1) - T(k))$  est aussi une fonction rationnelle de  $k$ , et  $s(k)$  lui-même est forcément un quotient de polynômes :

$$s(k) = f(k)/g(k).$$

En fait, nous pouvons même prouver que  $s(k)$  est tout simplement un polynôme. Voici comment. Si  $g(k)$  n'est pas constant et si  $f(k)$  et  $g(k)$  n'ont pas de facteur commun, soit  $N$  le plus grand entier pour lequel il existe un nombre complexe  $\beta$  tel que  $(k+\beta)$  et  $(k+\beta+N-1)$  soient tous deux des facteurs de  $g(k)$ . L'entier  $N$  est forcément positif, puisque  $N=1$  satisfait la condition précédente. L'équation (5.122) peut se réécrire

$$p(k)g(k+1)g(k) = q(k)f(k+1)g(k) - r(k)g(k+1)f(k),$$

et en posant  $k = -\beta$  et  $k = -\beta - N$  on obtient

$$r(-\beta)g(1-\beta)f(-\beta) = 0 = q(-\beta-N)f(1-\beta-N)g(-\beta-N).$$

Nous savons que  $f(-\beta) \neq 0$  et  $f(1-\beta-N) \neq 0$ , car  $f$  et  $g$  n'ont pas de racine commune. De même,  $g(1-\beta) \neq 0$  et  $g(-\beta-N) \neq 0$ , sinon  $g(k)$  contiendrait un des facteurs  $(k+\beta-1)$  ou  $(k+\beta+N)$ , ce qui contredirait le fait que  $N$  est maximal. Par conséquent,

$$r(-\beta) = q(-\beta-N) = 0.$$

C'est en contradiction avec la condition (5.118) ; donc  $s(k)$  est forcément un polynôme.

Notre tâche se réduit donc à trouver un polynôme  $s(k)$  qui satisfait (5.122), où  $p(k)$ ,  $q(k)$  et  $r(k)$  sont des polynômes donnés, ou bien à prouver que ce polynôme n'existe pas. C'est facile si on connaît le degré  $d$  de  $s(k)$ , car dans ce cas on peut poser

$$s(k) = \alpha_d k^d + \alpha_{d-1} k^{d-1} + \cdots + \alpha_0, \quad \alpha_d \neq 0 \quad (5.123)$$

où les coefficients  $(\alpha_d, \dots, \alpha_0)$  sont des inconnues, et injecter cette expression dans la récurrence (5.122). Le polynôme  $s(k)$  satisfara cette récurrence si et seulement si les  $\alpha$  satisfont chacune des équations linéaires correspondant à chaque puissance de  $k$  dans (5.122).

Il s'agit maintenant de trouver un moyen de déterminer le degré de  $s$ . En fait, nous allons voir que ce degré ne peut prendre que deux valeurs possibles. L'équation (5.122) peut se réécrire sous la forme

$$2p(k) = Q(k)(s(k+1) + s(k)) + R(k)(s(k+1) - s(k)), \quad (5.124)$$

$$\text{avec } Q(k) = q(k) - r(k) \text{ et } R(k) = q(k) + r(k).$$

Si  $s(k)$  est de degré  $d$ , alors la somme  $s(k+1) + s(k) = 2\alpha_d k^d + \cdots$  est aussi de degré  $d$ , tandis que la différence  $s(k+1) - s(k) = \Delta s(k) = d\alpha_d k^{d-1} + \cdots$  est de degré  $d-1$  (on peut considérer que le polynôme nul est de degré  $-1$ ). Notons  $\deg(P)$  le degré d'un polynôme  $P$ . Si  $\deg(Q) \geq \deg(R)$ , alors le degré du membre droit de (5.124) est égal à  $\deg(Q) + d$ , et par conséquent  $d = \deg(p) - \deg(Q)$ . En revanche, si  $\deg(Q) < \deg(R) = d'$ , on peut écrire  $Q(k) = \lambda' k^{d'-1} + \cdots$  et  $R(k) = \lambda k^{d'} + \cdots$  où  $\lambda \neq 0$ ; et le membre droit de (5.124) s'écrit sous la forme

$$(2\lambda' \alpha_d + \lambda d \alpha_d)k^{d+d'-1} + \cdots.$$

Il y a donc deux possibilités : soit  $2\lambda' + \lambda d \neq 0$ , et dans ce cas  $d = \deg(p) - \deg(R) + 1$ ; soit  $2\lambda' + \lambda d = 0$ , alors  $d > \deg(p) - \deg(R) + 1$ .

*J'ai compris : c'est pour que cette preuve marche que Gosper a introduit la condition (5.118).*

Le second cas ne mérite d'être considéré que si  $-2\lambda'/\lambda$  est un entier  $d$  plus grand que  $\deg(p) - \deg(R) + 1$ .

Parfait. Nous en savons assez pour lancer la deuxième étape de l'algorithme de Gosper : il suffit de tester deux valeurs de  $d$  pour trouver  $s(k)$  si l'équation (5.122) admet un polynôme comme solution. Si  $s(k)$  existe, il ne reste qu'à l'injecter dans (5.121) pour obtenir  $T$ . Si  $s(k)$  n'existe pas, on a démontré que  $t(k)$  n'est pas sommable en termes hypergéométriques.

Il est temps de voir cela sur un exemple : essayons donc la somme partielle (5.114). Si tout va bien, la méthode de Gosper devrait venir à bout de

$$\sum \binom{n}{k} (-1)^k \delta k$$

pour  $n$  donné. Nous cherchons donc la somme de

$$t(k) = \binom{n}{k} (-1)^k = \frac{n! (-1)^k}{k! (n-k)!}.$$

La première étape consiste à écrire le rapport des termes sous la forme adéquate (5.117) :

$$\frac{t(k+1)}{t(k)} = \frac{k-n}{k+1} = \frac{p(k+1)q(k)}{p(k)r(k+1)}.$$

*Pourquoi pas  
 $r(k) = k+1$  ?  
Oh, je vois.*

Nous pouvons donc simplement prendre  $p(k) = 1$ ,  $q(k) = k-n$  et  $r(k) = k$ . Ainsi la condition (5.118) est satisfaite, sauf si  $n$  est un entier strictement négatif. Supposons que ce n'est pas le cas.

Passons à la deuxième étape. D'après (5.124), nous devons considérer les polynômes  $Q(k) = -n$  et  $R(k) = 2k-n$ . Comme  $R$  est de degré plus élevé que  $Q$ , nous avons deux cas à examiner. Soit  $d = \deg(p) - \deg(R) + 1$ , ce qui donne 0 ; soit  $d = -2\lambda'/\lambda$ , où  $\lambda' = -n$  et  $\lambda = 2$ , donc  $d = n$ . Le premier cas est le plus sympathique car il n'impose pas que  $n$  soit un entier strictement positif. Commençons donc par lui, et nous n'essaierons l'autre que si celui-ci n'aboutit pas. Si on suppose que  $d = 0$ , alors  $s(k)$  est simplement égal à  $\alpha_0$ , et l'équation (5.122) se réduit à  $1 = (k-n)\alpha_0 - k\alpha_0$ . Il nous faut donc prendre  $\alpha_0 = -1/n$ , ce qui donne

$$\begin{aligned} T(k) &= \frac{r(k)s(k)t(k)}{p(k)} = k \cdot \left(\frac{-1}{n}\right) \cdot \binom{n}{k} (-1)^k \\ &= \binom{n-1}{k-1} (-1)^{k-1}, \quad \text{si } n \neq 0. \end{aligned}$$

C'est exactement ce que nous voulions confirmer.

Si on applique la même méthode pour évaluer la somme indéfinie  $\sum \binom{n}{k} \delta k$ , sans le  $(-1)^k$ , tout se passera quasiment de la même façon, sauf

que  $q(k)$  sera égal à  $n - k$  ; alors  $Q(k) = n - 2k$  aura un degré plus élevé que  $R(k) = n$ , et on en concluera que  $d$  est égal à  $\deg(p) - \deg(Q) = -1$ , ce qui est impossible. En effet, le polynôme  $s(k)$  ne peut pas avoir un degré négatif car il ne peut pas être nul. Par conséquent, la fonction  $\binom{n}{k}$  n'est pas sommable en termes hypergéométriques.

Toutefois, si on en croit S. Holmes [83], une fois qu'on a écarté l'impossible, ce qui reste, même improbable, ne peut être que vrai. Lorsque nous avons défini  $p$ ,  $q$  et  $r$  lors de la première étape, nous avons décidé d'ignorer la possibilité que  $n$  soit un entier strictement négatif. Que se passerait-il s'il l'était ? Posons  $n = -N$ , où  $N$  est un entier strictement positif. Dans ce cas, le rapport des termes de  $\sum \binom{n}{k} \delta_k$  est

$$\frac{t(k+1)}{t(k)} = \frac{-(k+N)}{(k+1)} = \frac{p(k+1)}{p(k)} \frac{q(k)}{r(k+1)}$$

et on prend, conformément à (5.119),  $p(k) = (k+1)^{\overline{N-1}}$ ,  $q(k) = -1$  et  $r(k) = 1$ . À présent, selon la deuxième étape de l'algorithme de Gosper, il nous faut chercher un polynôme  $s(k)$  de degré  $d = N - 1$ . Essayons, nous verrons bien. Prenons par exemple  $N = 2$ . D'après la récurrence (5.122), il nous faut résoudre

$$k+1 = -((k+1)\alpha_1 + \alpha_0) - (k\alpha_1 + \alpha_0).$$

En séparant les coefficients de  $k$  et de 1, cela revient à résoudre

$$1 = -\alpha_1 - \alpha_0; \quad 1 = -\alpha_1 - \alpha_0 - \alpha_0;$$

donc  $s(k) = -\frac{1}{2}k - \frac{1}{4}$  est solution, et

$$T(k) = \frac{1 \cdot (-\frac{1}{2}k - \frac{1}{4}) \cdot \binom{-2}{k}}{k+1} = (-1)^{k-1} \frac{2k+1}{4}.$$

Cela pourrait-il être la bonne somme ? En effet :

$$(-1)^k \frac{2k+3}{4} - (-1)^{k-1} \frac{2k+1}{4} = (-1)^k (k+1) = \binom{-2}{k}.$$

*"Excellent, Holmes !"  
"Elementaire, mon  
cher Watson."*

Notons qu'en ajoutant une borne à cette somme, on peut l'écrire différemment :

$$\begin{aligned} \sum_{k < m} \binom{-2}{k} &= (-1)^{k-1} \frac{2k+1}{4} \Big|_0^m = \frac{(-1)^{m-1}}{2} \left( m + \frac{1 - (-1)^m}{2} \right) \\ &= (-1)^{m-1} \left[ \frac{m}{2} \right], \quad m \geq 0 \text{ entier.} \end{aligned}$$

Le défaut de cette représentation est qu'elle cache le fait que  $\binom{-2}{k}$  est sommable en termes hypergéométriques, car  $\lceil m/2 \rceil$  n'est pas un terme hypergéométrique (voir l'exercice 12).

Un problème pourrait survenir dans le dénominateur de (5.121) s'il existait un  $k$  tel que  $p(k) = 0$ . L'exercice 97 donne un aperçu de ce qu'on peut faire dans ce genre de situation.

Nous plaidions, un peu plus haut, pour la conception d'une base de données des sommes hypergéométriques définies, qui nous aiderait à résoudre bien des problèmes. Remarquez que, par contre, la même démarche serait tout à fait inutile pour les fonctions indéfiniment sommables en termes hypergéométriques, car l'algorithme de Gosper fournit une méthode rapide et standard pour résoudre tous les cas sommables.

Marko Petkovsek [291] a généralisé l'algorithme de Gosper pour résoudre des problèmes d'inversion plus compliqués. Il donne une méthode pour trouver tous les termes hypergéométriques  $T(k)$  qui satisfont la récurrence d'ordre  $l$

$$t(k) = p_l(k)T(k+l) + \cdots + p_1(k)T(k+1) + p_0(k)T(k), \quad (5.125)$$

quelques soient le terme hypergéométrique  $t(k)$  et les polynômes  $p_l(k), \dots, p_1(k), p_0(k)$  donnés.

## 5.8 SOMMATION AUTOMATIQUE

L'algorithme de Gosper, aussi élégant soit-il, ne peut trouver des formes closes que pour un petit nombre de sommes parmi toutes celles que nous sommes amenés à rencontrer. Cependant, nous n'avons pas encore tout dit. Doron Zeilberger [383] a montré comment étendre l'algorithme de Gosper à des cas bien plus généraux. Ainsi, grâce aux travaux de Zeilberger, on peut manipuler non plus seulement des sommes partielles, mais aussi des sommes sur tout  $k$ . Cela nous fournit donc une alternative aux méthodes que nous avons vues dans les sections 5.5 et 5.6. De plus, comme pour la méthode originale de Gosper, les calculs peuvent être faits par ordinateur, sans aucun appel à l'intuition ou à la chance.

L'idée de départ est de considérer le terme à sommer comme une fonction  $t(n, k)$  de deux variables  $n$  et  $k$  (dans l'algorithme de Gosper, on écrivait simplement  $t(k)$ ). Lorsque  $t(n, k)$  n'est pas indéfiniment sommable en termes hypergéométriques en fonction de  $k$  — ne nous voilons pas la face, cela est très fréquent — Zeilberger fait remarquer qu'on peut souvent modifier  $t(n, k)$  pour obtenir quelque chose qui *est* indéfiniment sommable. Par exemple, en pratique, l'expression  $\beta_0(n)t(n, k) + \beta_1(n)t(n+1, k)$  est souvent indéfiniment sommable sur  $k$ , pour des polynômes  $\beta_0(n)$  and  $\beta_1(n)$  adéquats. Quand on effectue la somme sur  $k$ , on obtient une récurrence en  $n$

qui résout le problème.

Commençons par un cas simple pour nous familiariser avec cette approche générale. Supposons que nous ayons oublié la formule du binôme et que nous voulions calculer  $\sum_k \binom{n}{k} z^k$ . Comment trouver la réponse sans faire appel à l'intuition et sans conjecturer la formule ? Nous avons vu dans ce chapitre par exemple dans le problème 3 de la section 5.2, qu'on peut remplacer  $\binom{n}{k}$  par  $\binom{n-1}{k} + \binom{n-1}{k-1}$  et modéliser le résultat de façon à arriver à nos fins. Il existe cependant un moyen plus systématique de procéder.

*Et sans regarder la page 186.*

Soit  $t(n, k) = \binom{n}{k} z^k$  la quantité que nous voulons sommer. L'algorithme de Gosper nous dit qu'on ne peut pas évaluer les sommes partielles  $\sum_{k \leq m} t(n, k)$  pour tout  $n$  en termes hypergéométriques, sauf si  $z = -1$ . Considérons donc un terme plus général

$$\hat{t}(n, k) = \beta_0(n)t(n, k) + \beta_1(n)t(n+1, k). \quad (5.126)$$

Nous allons chercher des valeurs de  $\beta_0(n)$  et  $\beta_1(n)$  pour lesquelles l'algorithme de Gosper marche. D'abord, simplifions (5.126) en utilisant la relation entre  $t(n+1, k)$  et  $t(n, k)$  pour éliminer le  $t(n+1, k)$ . Comme

$$\begin{aligned} \frac{t(n+1, k)}{t(n, k)} &= \frac{(n+1)! z^k}{(n+1-k)! k!} \frac{(n-k)! k!}{n! z^k} \\ &= \frac{n+1}{n+1-k}, \end{aligned}$$

on a

$$\hat{t}(n, k) = p(n, k) \frac{t(n, k)}{n+1-k},$$

avec

$$p(n, k) = (n+1-k)\beta_0(n) + (n+1)\beta_1(n).$$

Maintenant, appliquons l'algorithme de Gosper à  $\hat{t}(n, k)$ , avec  $n$  fixé. Nous écrivons donc

$$\frac{\hat{t}(n, k+1)}{\hat{t}(n, k)} = \frac{\hat{p}(n, k+1)}{\hat{p}(n, k)} \frac{q(n, k)}{r(n, k+1)} \quad (5.127)$$

comme en (5.117). Selon la méthode originale de Gosper, on commencerait en prenant  $\hat{p}(n, k) = 1$ . Avec l'extension de Zeilberger, il est préférable de partir avec  $\hat{p}(n, k) = p(n, k)$ . Remarquons que si on pose  $\bar{t}(n, k) = \hat{t}(n, k)/p(n, k)$  et  $\bar{p}(n, k) = \hat{p}(n, k)/p(n, k)$ , l'équation (5.127) est équivalente à

$$\frac{\bar{t}(n, k+1)}{\bar{t}(n, k)} = \frac{\bar{p}(n, k+1)}{\bar{p}(n, k)} \frac{q(n, k)}{r(n, k+1)}. \quad (5.128)$$

Ainsi, si on trouve  $\bar{p}$ ,  $q$  et  $r$  satisfaisant (5.128) en partant de  $\bar{p}(n, k) = 1$ , on trouve aussi  $\hat{p}$ ,  $q$  et  $r$  satisfaisant (5.127). Ceci nous rend la vie plus facile, car  $\bar{t}(n, k)$  ne dépend pas des inconnues  $\beta_0(n)$  et  $\beta_1(n)$  qui apparaissent dans  $\hat{t}(n, k)$ . Dans notre cas,  $\bar{t}(n, k) = t(n, k)/(n+1-k) = n! z^k/(n+1-k)! k!$ , donc

$$\frac{\bar{t}(n, k+1)}{\bar{t}(n, k)} = \frac{(n+1-k)z}{k+1}.$$

*Cette fois, je sais pourquoi  $r(n, k)$  ne vaut pas  $k+1$ .*

Nous pouvons prendre  $q(n, k) = (n+1-k)z$  et  $r(n, k) = k$ . Ces polynômes en  $k$  sont supposés satisfaire la condition (5.118). S'ils ne la satisfont pas, nous devons supprimer des facteurs de  $q$  et  $r$  et ajouter les facteurs correspondants (5.119) à  $\bar{p}(n, k)$ . Cependant, ceci ne doit être fait que lorsque la quantité  $\alpha - \beta$  de (5.118) est un entier strictement positif indépendant de  $n$ , car nous voulons que nos calculs soient valides pour tout  $n$ . (En fait, avec la factorielle généralisée (5.83), les formules que nous allons obtenir seront valides même lorsque  $n$  et  $k$  ne seront pas des entiers).

Notre choix initial de  $q$  et  $r$  satisfait bien (5.118) dans le sens du paragraphe précédent. Nous pouvons donc passer à la deuxième étape de l'algorithme de Gosper : il nous faut résoudre l'analogue de (5.122), en utilisant (5.127) à la place de (5.117). Il s'agit donc de résoudre

$$\hat{p}(n, k) = q(n, k)s(n, k+1) - r(n, k)s(n, k) \quad (5.129)$$

où l'inconnue est le polynôme

$$s(n, k) = \alpha_d(n)k^d + \alpha_{d-1}(n)k^{d-1} + \cdots + \alpha_0(n). \quad (5.130)$$

(Les coefficients de  $s$  ne sont pas des constantes, ce sont des fonctions de  $n$ ). Dans notre cas, l'équation (5.129) s'écrit

$$\begin{aligned} & (n+1-k)\beta_0(n) + (n+1)\beta_1(n) \\ &= (n+1-k)zs(n, k+1) - ks(n, k) \end{aligned}$$

Il faut la voir comme une équation polynomiale en  $k$  dont les coefficients sont des fonctions de  $n$ . Le degré  $d$  de  $s$  se détermine comme nous l'avons fait auparavant : on considère  $Q(n, k) = q(n, k) - r(n, k)$  et  $R(n, k) = q(n, k) + r(n, k)$ . Comme  $\deg(Q) = \deg(R) = 1$  (on suppose que  $z \neq \pm 1$ ), on a  $d = \deg(\hat{p}) - \deg(Q) = 0$ , et  $s(n, k) = \alpha_0(n)$  ne dépend pas de  $k$ . Notre équation devient alors

$$(n+1-k)\beta_0(n) + (n+1)\beta_1(n) = (n+1-k)z\alpha_0(n) - k\alpha_0(n),$$

et en séparant les puissances de  $k$ , on obtient le système équivalent

$$\begin{aligned} (n+1)\beta_0(n) + (n+1)\beta_1(n) - (n+1)z\alpha_0(n) &= 0, \\ -\beta_0(n) &+ (z+1)\alpha_0(n) = 0. \end{aligned}$$

*La fonction  $\deg(Q)$  représente le degré en  $k$ , le nombre  $n$  étant considéré comme une constante.*

Nous avons donc une solution de (5.129), avec

$$\beta_0(n) = z + 1, \quad \beta_1(n) = -1, \quad \alpha_0(n) = s(n, k) = 1.$$

(Par chance, le  $n$  a complètement disparu).

Nous avons donc démontré, de façon totalement automatique, que le terme  $\hat{t}(n, k) = (z + 1)t(n, k) - t(n + 1, k)$  est sommable en termes hypergéométriques. Autrement dit,

Ou "en d'autres termes" ?

$$\hat{t}(n, k) = T(n, k + 1) - T(n, k), \quad (5.131)$$

où  $T(n, k)$  est un terme hypergéométrique en  $k$ . Voyons ce que vaut exactement ce  $T(n, k)$ . D'après (5.121) et (5.128), on a

$$T(n, k) = \frac{r(n, k)s(n, k)\hat{t}(n, k)}{\hat{p}(n, k)} = r(n, k)s(n, k)\bar{t}(n, k), \quad (5.132)$$

car  $\bar{p}(n, k) = 1$  (en pratique,  $\bar{p}(n, k)$  est presque toujours égal à 1). Par conséquent,

$$T(n, k) = \frac{k}{n + 1 - k} t(n, k) = \frac{k}{n + 1 - k} \binom{n}{k} z^k = \binom{n}{k-1} z^k.$$

Vérifions, juste pour la forme, que l'équation (5.131) est bien respectée :

$$(z + 1)\binom{n}{k} z^k - \binom{n+1}{k} z^k = \binom{n}{k} z^{k+1} - \binom{n}{k-1} z^k.$$

En fait, nous n'avons même pas besoin de connaître précisément  $T(n, k)$  car nous allons sommer sur tous les entiers  $k$ . Tout ce que nous avons besoin de savoir, c'est que  $T(n, k)$  n'est non nul que pour un nombre fini de valeurs de  $k$ , pour tout entier  $n$  positif ou nul donné. Alors la somme télescopique des  $T(n, k + 1) - T(n, k)$  sur tout  $k$  vaut forcément 0.

Soit  $S_n = \sum_k t(n, k) = \sum_k \binom{n}{k} z^k$  la somme dont nous sommes partis. Nous en savons maintenant largement assez sur  $t(n, k)$  pour pouvoir la calculer. L'algorithme de Gosper-Zeilberger a établi que

$$\sum_k ((z + 1)t(n, k) - t(n + 1, k)) = 0.$$

Or, cette somme n'est autre que  $(z + 1) \sum_k t(n, k) - \sum_k t(n + 1, k) = (z + 1)S_n - S_{n+1}$ . Par conséquent,

$$S_{n+1} = (z + 1)S_n. \quad (5.133)$$

C'est une récurrence que nous savons résoudre, à condition de connaître  $S_0$  ; et bien évidemment,  $S_0 = 1$ . Nous en déduisons donc que  $S_n = (z + 1)^n$ , pour tout entier  $n \geq 0$ . CQFD.

En fait,  
 $\lim_{k \rightarrow \infty} T(n, k) = 0$   
si  $|z| < 1$   
et  $n$  est un nombre complexe quelconque. Donc (5.133) est vraie pour tout  $n$ , et en particulier  $S_n = (z + 1)^n$  lorsque  $n$  est un entier strictement négatif.

Revenons sur le calcul que nous venons de faire et faisons en un petit résumé qui pourra s'appliquer en d'autres occasions. Pour un terme général  $t(n, k)$  donné, l'algorithme de Gosper-Zeilberger peut se formuler ainsi :

- 0 Poser  $l := 0$  (nous allons chercher des récurrences en  $n$  d'ordre  $l$ ).
- 1 Soit  $\hat{t}(n, k) = \beta_0(n)t(n, k) + \dots + \beta_l(n)t(n+l, k)$ , où  $\beta_0(n), \dots, \beta_l(n)$  sont des fonctions inconnues. Utiliser les propriétés de  $t(n, k)$  pour trouver une combinaison linéaire  $p(n, k)$  de  $\beta_0(n), \dots, \beta_l(n)$  dont les coefficients sont des polynômes en  $n$  et  $k$ , de sorte que  $\hat{t}(n, k)$  puisse s'écrire sous la forme  $p(n, k)\bar{t}(n, k)$ , où  $\bar{t}(n, k)$  est un terme hypergéométrique en  $k$ . Trouver des polynômes  $\bar{p}(n, k)$ ,  $q(n, k)$ ,  $r(n, k)$  tels que le rapport des termes de  $\bar{t}(n, k)$  s'exprime sous la forme (5.128), où  $q(n, k)$  et  $r(n, k)$  satisfont la condition de Gosper (5.118). Poser  $\hat{p}(n, k) = p(n, k)\bar{p}(n, k)$ .
- 2a Poser  $d_Q := \deg(q - r)$ ,  $d_R := \deg(q + r)$  et

$$d := \begin{cases} \deg(\hat{p}) - d_Q, & \text{si } d_Q \geq d_R; \\ \deg(\hat{p}) - d_R + 1, & \text{si } d_Q < d_R. \end{cases}$$

- 2b Si  $d \geq 0$ , définir  $s(n, k)$  par (5.130), et considérer les équations linéaires en  $\alpha_0, \dots, \alpha_d, \beta_0, \dots, \beta_l$  obtenues en mettant en équations les coefficients des puissances de  $k$  dans l'équation fondamentale (5.129). Si ces équations admettent une solution telle que l'un des  $\beta_0, \dots, \beta_l$  au moins ne soit pas nul, aller en 4. Sinon, si  $d_Q < d_R$  et si  $-2\lambda'/\lambda$  est un entier strictement supérieur à  $d$ , où  $\lambda$  est le coefficient de  $k^{d_R}$  dans  $q + r$  et  $\lambda'$  est le coefficient de  $k^{d_R-1}$  dans  $q - r$ , poser  $d := -2\lambda'/\lambda$  et revenir en 2.
- 3 (Le terme  $\hat{t}(n, k)$  n'est pas hypergéométriquement sommable). Ajouter 1 à  $l$  et revenir en 1.
- 4 (C'est gagné). Poser  $T(n, k) := r(n, k)s(n, k)\bar{t}(n, k)/\bar{p}(n, k)$ . L'algorithme a établi que  $\hat{t}(n, k) = T(n, k+1) - T(n, k)$ .

Nous démontrerons un peu plus loin que cet algorithme se termine avec succès dès lors que  $t(n, k)$  appartient à une classe importante de termes, que l'on appelle termes propres.

Il existe bien des façons de démontrer la formule du binôme. Notre premier exemple d'application de l'approche de Gosper-Zeilberger était donc plus instructif que spectaculaire. Attaquons-nous maintenant à la convolution de Vandermonde. Gosper et Zeilberger peuvent-ils trouver automatiquement la forme simple de  $\sum_k \binom{a}{k} \binom{b}{n-k}$ ? L'algorithme part de  $l = 0$ , tout comme dans la méthode originale de Gosper, et essaie de voir si  $\binom{a}{k} \binom{b}{n-k}$  est sommable en termes hypergéométriques. Surprise : il se trouve que c'est sommable si  $a + b$  est un entier positif ou nul particulier (voir l'exercice 94). Cependant, c'est le cas général qui nous intéresse, et l'algorithme découvre rapidement qu'alors la somme indéfinie n'est pas

un terme hypergéométrique. La valeur de  $l$  passe donc de 0 à 1, et l'algorithme considère  $\hat{t}(n, k) = \beta_0(n)t(n, k) + \beta_1(n)t(n+1, k)$ . L'étape suivante, comme pour la formule du binôme, consiste à écrire  $\hat{t}(n, k) = p(n, k)\bar{t}(n, k)$ , où  $p(n, k)$  est obtenu en simplifiant  $t(n+1, k)/t(n, k)$ . Dans notre cas — le lecteur ferait bien de vérifier ces calculs sur une feuille de brouillon, bien qu'ils ne soient pas aussi difficiles qu'ils paraissent — tout se passe comme précédemment, et on obtient les valeurs

$$\begin{aligned} p(n, k) &= (n+1-k)\beta_0(n) + (b-n+k)\beta_1(n) = \hat{p}(n, k), \\ \bar{t}(n, k) &= t(n, k)/(n+1-k) = a!b!/(a-k)!k!(b-n+k)!(n+1-k)!, \\ q(n, k) &= (n+1-k)(a-k), \\ r(n, k) &= (b-n+k)k. \end{aligned}$$

L'étape 2a trouve que  $\deg(q - r) < \deg(q + r)$  et  $d = \deg(\hat{p}) - \deg(q + r) + 1 = 0$ , donc  $s(n, k)$  est encore une fois indépendant de  $k$ . L'équation fondamentale de Gosper (5.129) est équivalente à deux équations à trois inconnues,

$$\begin{aligned} (n+1)\beta_0(n) + (b-n)\beta_1(n) - (n+1)a\alpha_0(n) &= 0, \\ -\beta_0(n) + \beta_1(n) + (a+b+1)\alpha_0(n) &= 0, \end{aligned}$$

dont la solution est

$$\beta_0(n) = a + b - n, \quad \beta_1(n) = -n - 1, \quad \alpha_0(n) = 1.$$

Nous en concluons que  $(a+b-n)t(n, k) - (n+1)t(n+1, k)$  est sommable sur  $k$ . Donc, si on pose

$$S_n = \sum_k \binom{a}{k} \binom{b}{n-k},$$

alors  $S_n$  satisfait la récurrence

$$S_{n+1} = \frac{a+b-n}{n+1} S_n.$$

Par conséquent,  $S_n = \binom{a+b}{n}$  puisque  $S_0 = 1$ . C'est du gâteau.

Et si nous essayions l'identité Saalschützienne à trois binomiaux de (5.28) ? La preuve qui en est donnée dans l'exercice 43 est intéressante, mais elle requiert de l'inspiration. Quand un art devient science, l'inspiration devient transpiration. Voyons donc si la méthode de Gosper-Zeilberger est capable de découvrir et démontrer (5.28) de façon totalement automatique. Permettons-nous tout d'abord, sans changer le fond du problème, d'effectuer quelques changements de variables pour que (5.28) soit un peu

*Voici le point essentiel : la méthode de Gosper-Zeilberger conduit toujours à des équations linéaires en les  $\alpha$  et en les  $\beta$ , car le membre gauche de (5.129) est linéaire en les  $\beta$  et le membre droit est linéaire en les  $\alpha$ .*

plus symétrique : posons  $m = b + d$ ,  $n = a$ ,  $r = a + b + c + d$ ,  $s = a + b + c$ .

Ainsi l'identité s'écrit

$$\begin{aligned} \sum_k & \frac{(a+b+c+d+k)!}{(a-k)!(b-k)!(c+k)!(d+k)!k!} \\ &= \frac{(a+b+c+d)!(a+b+c)!(a+b+d)!}{a!b!(a+c)!(a+d)!(b+c)!(b+d)!}. \end{aligned} \quad (5.134)$$

Pour que la somme soit finie, nous supposons que soit  $a$  soit  $b$  est un entier positif ou nul.

Soit  $t(n, k) = (n + b + c + d + k)!/(n - k)!(b - k)!(c + k)!(d + k)!k!$  et  $\tilde{t}(n, k) = \beta_0(n)t(n, k) + \beta_1(n)t(n + 1, k)$ . Maintenant, nous n'avons plus qu'à suivre un chemin balisé. Posons

$$\begin{aligned} p(n, k) &= (n + 1 - k)\beta_0(n) + (n + 1 + b + c + d + k)\beta_1(n) \\ &= \hat{p}(n, k), \\ \tilde{t}(n, k) &= \frac{t(n, k)}{n + 1 - k} = \frac{(n + b + c + d + k)!}{(n + 1 - k)!(b - k)!(c + k)!(d + k)!k!}, \\ q(n, k) &= (n + b + c + d + k + 1)(n + 1 - k)(b - k), \\ r(n, k) &= (c + k)(d + k)k, \end{aligned}$$

et essayons de résoudre (5.129) pour  $s(n, k)$ . On a de nouveau  $\deg(q - r) < \deg(q + r)$ , mais cette fois  $\deg(\hat{p}) - \deg(q + r) + 1 = -1$ . On dirait bien que nous sommes coincés. Cependant, n'oublions pas que l'étape 2b nous propose un autre choix,  $d = -2\lambda'/\lambda$ , pour le degré de  $s$ . Il serait bon de l'essayer avant d'abandonner la partie. Ici,  $R(n, k) = q(n, k) + r(n, k) = 2k^3 + \dots$ , donc  $\lambda = 2$ , tandis que le polynôme  $Q(n, k) = q(n, k) - r(n, k)$  se trouve, miraculeusement ou presque, être de degré 1 en  $k$  (par chance, le coefficient de  $k^2$  disparaît). Ainsi,  $\lambda' = 0$ , et Gosper nous invite à prendre  $d = 0$  et  $s(n, k) = \alpha_0(n)$ .

Voici maintenant les équations à résoudre :

$$\begin{aligned} (n + 1)\beta_0(n) + (n + 1 + b + c + d)\beta_1(n) \\ - (n + 1)(n + 1 + b + c + d)b\alpha_0(n) &= 0, \\ -\beta_0(n) + \beta_1(n) \\ - ((n + 1)b - (n + 1 + b)(n + 1 + b + c + d) - cd)\alpha_0(n) &= 0; \end{aligned}$$

Avec un petit effort, on trouve

$$\begin{aligned} \beta_0(n) &= (n + 1 + b + c)(n + 1 + b + d)(n + 1 + b + c + d), \\ \beta_1(n) &= -(n + 1)(n + 1 + c)(n + 1 + d), \\ \alpha_0(n) &= 2n + 2 + b + c + d. \end{aligned}$$

Décider quel paramètre doit s'appeler  $n$ , voilà la seule partie non automatique de la méthode.

Remarquez que  $\lambda'$  n'est pas le coefficient directeur de  $Q$ , bien que  $\lambda$  soit celui de  $R$ . Le nombre  $\lambda'$  est le coefficient de  $k^{\deg(R)-1}$  dans  $Q$ .

Le cerveau cuit,  
l'identité suit.

L'identité (5.134) s'ensuit immédiatement.

## 252 COEFFICIENTS BINOMIAUX

Si on travaille en prenant  $n = d$  au lieu de  $n = a$ , on trouve une autre preuve, similaire à celle-ci, de (5.134) (voir l'exercice 99).

La méthode de Gosper-Zeilberger peut aussi bien nous aider à évaluer des sommes définies sur un intervalle limité que des sommes sur tout  $k$ . Considérons par exemple

$$S_n(z) = \sum_{k=0}^n \binom{n+k}{k} z^k. \quad (5.135)$$

Pour  $z = \frac{1}{2}$ , nous avons obtenu un résultat "inattendu" en (5.20). Gosper et Zeilberger s'y seraient-ils attendus ? Posons  $t(n, k) = \binom{n+k}{k} z^k$ , ce qui nous mène à

$$\begin{aligned} p(n, k) &= (n+1)\beta_0(n) + (n+1+k)\beta_1(n) = \hat{p}(n, k), \\ \bar{t}(n, k) &= t(n, k)/(n+1) = (n+k)! z^k / k! (n+1)!, \\ q(n, k) &= (n+1+k)z, \\ r(n, k) &= k, \end{aligned}$$

et  $\deg(s) = \deg(\hat{p}) - \deg(q-r) = 0$ . On résout l'équation (5.129) en prenant  $\beta_0(n) = 1$ ,  $\beta_1(n) = z-1$  et  $s(n, k) = 1$ . Nous trouvons donc

$$t(n, k) + (z-1)t(n+1, k) = T(n, k+1) - T(n, k), \quad (5.136)$$

où  $T(n, k) = r(n, k)s(n, k)\bar{t}(n, k)/\hat{p}(n, k) = \binom{n+k}{k-1} z^k$ . Nous pouvons maintenant sommer (5.136) pour  $0 \leq k \leq n+1$ , obtenant ainsi

$$\begin{aligned} S_n(z) + t(n, n+1) + (z-1)S_{n+1}(z) &= T(n, n+2) - T(n, 0) \\ &= \binom{2n+2}{n+1} z^{n+2} \\ &= 2 \binom{2n+1}{n} z^{n+2}. \end{aligned}$$

Or,  $t(n, n+1) = \binom{2n+1}{n+1} z^{n+1} = \binom{2n+1}{n} z^{n+1}$ , donc

$$S_{n+1}(z) = \frac{1}{1-z} \left( S_n(z) + (1-2z) \binom{2n+1}{n} z^{n+1} \right). \quad (5.137)$$

On voit tout de suite que le cas  $z = \frac{1}{2}$  est bien particulier et que  $S_{n+1}(\frac{1}{2}) = 2S_n(\frac{1}{2})$ . De plus, la récurrence (5.137) peut être simplifiée en appliquant le facteur de sommation  $(1-z)^{n+1}$  aux deux membres. On aboutit à une identité générale,

$$(1-z)^n \sum_{k=0}^n \binom{n+k}{k} z^k = 1 + \frac{1-2z}{2-2z} \sum_{k=1}^n \binom{2k}{k} (z(1-z))^k, \quad (5.138)$$

dont relativement peu de gens auraient pu soupçonner l'existence avant que Gosper et Zeilberger ne s'en mêlent. Grâce à eux, c'est maintenant pure routine que de produire de telles identités.

Que dire de la somme similaire

$$S_n(z) = \sum_{k=0}^n \binom{n-k}{k} z^k, \quad (5.139)$$

que nous avons rencontrée en (5.74)? En toute confiance, nous posons  $t(n, k) = \binom{n-k}{k} z^k$  et calculons

$$\begin{aligned} p(n, k) &= (n+1-2k)\beta_0(n) + (n+1-k)\beta_1(n) = \hat{p}(n, k), \\ \bar{t}(n, k) &= t(n, k)/(n+1-2k) = (n-k)! z^k / k! (n+1-2k)!, \\ q(n, k) &= (n+1-2k)(n-2k)z, \\ r(n, k) &= (n+1-k)k. \end{aligned}$$

$S_n(-\frac{1}{4})$  égale  
 $(n+1)/2^n$ .

Aïe ! Impossible de résoudre (5.129) si  $z \neq -\frac{1}{4}$ , car le degré de  $s$  serait égal à  $\deg(\hat{p}) - \deg(q - r) = -1$ .

Qu'à cela ne tienne. Ajoutons simplement un autre paramètre  $\beta_2(n)$  et essayons  $\hat{t}(n, k) = \beta_0(n)t(n, k) + \beta_1(n)t(n+1, k) + \beta_2(n)t(n+2, k)$  à la place :

$$\begin{aligned} p(n, k) &= (n+1-2k)(n+2-2k)\beta_0(n) \\ &\quad + (n+1-k)(n+2-2k)\beta_1(n) \\ &\quad + (n+1-k)(n+2-k)\beta_2(n) = \hat{p}(n, k), \\ \bar{t}(n, k) &= t(n, k)/(n+1-2k)(n+2-2k) = (n-k)! z^k / k! (n+2-2k)!, \\ q(n, k) &= (n+2-2k)(n+1-2k)z, \\ r(n, k) &= (n+1-k)k. \end{aligned}$$

Maintenant, nous pouvons prendre  $s(n, k) = \alpha_0(n)$ , et (5.129) admet une solution :

$$\beta_0(n) = z, \quad \beta_1(n) = 1, \quad \beta_2(n) = -1, \quad \alpha_0(n) = 1.$$

Nous venons de découvrir que

$$zt(n, k) + t(n+1, k) - t(n+2, k) = T(n, k+1) - T(n, k),$$

où  $T(n, k)$  est égal à  $r(n, k)s(n, k)\hat{t}(n, k)/\hat{p}(n, k) = (n+1-k)k\bar{t}(n, k) = \binom{n+1-k}{k-1}z^k$ . En sommant de  $k=0$  à  $k=n$ , on trouve

$$\begin{aligned} zS_n(z) + (S_{n+1}(z) - \binom{0}{n+1}z^{n+1}) - (S_{n+2}(z) - \binom{0}{n+2}z^{n+2} - \binom{1}{n+1}z^{n+1}) \\ = T(n, n+1) - T(n, 0). \end{aligned}$$

Comme  $\binom{1}{n+1}z^{n+1} = \binom{0}{n}z^{n+1} = T(n, n+1)$  pour tout  $n \geq 0$ , on aboutit à

$$S_{n+2}(z) = S_{n+1}(z) + zS_n(z), \quad n \geq 0. \quad (5.140)$$

Nous étudierons les récurrences de ce type dans les chapitres 6 et 7. Les méthodes que nous y verrons permettent de déduire directement de (5.140) la forme close (5.74), lorsque  $S_0(z) = S_1(z) = 1$ .

Voici, pour compléter le tableau, un dernier exemple très célèbre. Le mathématicien français Roger Apéry a résolu, en 1978, un problème ouvert depuis bien longtemps. Il prouva que le nombre  $\zeta(3) = 1 + 2^{-3} + 3^{-3} + 4^{-3} + \dots$  est irrationnel [14]. Une des principaux ingrédients de sa preuve faisait intervenir les sommes de binomiaux

$$A_n = \sum_k \binom{n}{k}^2 \binom{n+k}{k}^2, \quad (5.141)$$

pour lesquelles il conjecturait une récurrence que personne n'était incapable de vérifier à l'époque (les  $A_n$  sont maintenant appelés nombres d'Apéry :  $A_0 = 1, A_1 = 5, A_2 = 73, A_3 = 1445, A_4 = 33001$ , etc). Finalement, Don Zagier et Henri Cohen [356] trouvèrent une preuve de la conjecture d'Apéry. D'ailleurs, cette preuve, pour ce cas particulier mais particulièrement difficile, constitue l'un des principaux indices qui ont mis Zeilberger sur la voie de la méthode générale que nous sommes en train d'étudier.

Nous sommes maintenant tellement entraînés que la somme de (5.141) nous paraît presque triviale. Posons  $t(n, k) = \binom{n}{k}^2 \binom{n+k}{k}^2$  et  $\tilde{t}(n, k) = \beta_0(n)t(n, k) + \beta_1(n)t(n+1, k) + \beta_2(n)t(n+2, k)$ , puis essayons de résoudre (5.129) avec

$$\begin{aligned} p(n, k) &= (n+1-k)^2(n+2-k)^2\beta_0(n) \\ &\quad + (n+1+k)^2(n+2-k)^2\beta_1(n) \\ &\quad + (n+1+k)^2(n+2+k)^2\beta_2(n) = \hat{p}(n, k), \\ \tilde{t}(n, k) &= t(n, k)/(n+1-k)^2(n+2-k)^2 = (n+k)!^2/k!^4(n+2-k)!^2, \\ q(n, k) &= (n+1+k)^2(n+2-k)^2, \\ r(n, k) &= k^4. \end{aligned}$$

(Nous avons d'abord essayé sans  $\beta_2$ , sans succès).

(Inutile de s'inquiéter sous prétexte que  $q$  contient le facteur  $(k+n+1)$  tandis que  $r$  contient le facteur  $k$  ; ceci ne viole pas (5.118), car nous considérons  $n$  non comme un entier fixé, mais comme un paramètre variable). Comme  $q(n, k) - r(n, k) = -2k^3 + \dots$ , nous pouvons poser  $\deg(s) = -2\lambda'/\lambda = 2$ , et donc prendre

$$s(n, k) = \alpha_2(n)k^2 + \alpha_1(n)k + \alpha_0(n).$$

Avec  $s$  ainsi choisi, la récurrence (5.129) se ramène à cinq équations en les six inconnues  $\beta_0(n)$ ,  $\beta_1(n)$ ,  $\beta_2(n)$ ,  $\alpha_0(n)$ ,  $\alpha_1(n)$ ,  $\alpha_2(n)$ . Par exemple, l'équation provenant des coefficients de  $k^0$  se simplifie en

$$\beta_0 + \beta_1 + \beta_2 - \alpha_0 - \alpha_1 - \alpha_2 = 0;$$

l'équation correspondant aux coefficients de  $k^4$  est

$$\beta_0 + \beta_1 + \beta_2 + \alpha_1 + (6 + 6n + 2n^2)\alpha_2 = 0.$$

Les trois autres équations sont plus compliquées. Toutefois, le principal est que ces équations linéaires, comme toutes les équations qui apparaissent à ce stade de l'algorithme de Gosper-Zeilberger, sont *homogènes* (leurs membres droits sont nuls). Par conséquent, elles admettent toujours une solution non nulle s'il y a plus d'inconnues que d'équations. Voici une solution pour le cas qui nous intéresse :

$$\begin{aligned}\beta_0(n) &= (n+1)^3, \\ \beta_1(n) &= -(2n+3)(17n^2+51n+39), \\ \beta_2(n) &= (n+2)^3, \\ \alpha_0(n) &= -16(n+1)(n+2)(2n+3), \\ \alpha_1(n) &= -12(2n+3), \\ \alpha_2(n) &= 8(2n+3).\end{aligned}$$

Par conséquent,

$$(n+1)^3 t(n, k) - (2n+3)(17n^2+51n+39)t(n+1, k) + (n+2)^3 t(n+2, k) = T(n, k+1) - T(n, k),$$

où  $T(n, k) = k^4 s(n, k) \bar{t}(n, k) = (2n+3)(8k^2 - 12k - 16(n+1)(n+2)) \times (n+k)!^2 / (k-1)!^4 (n+2-k)!^2$ . En sommant sur  $k$ , on trouve l'incroyable mais vraie récurrence d'Apéry,

$$(n+1)^3 A_n + (n+2)^3 A_{n+2} = (2n+3)(17n^2+51n+39)A_{n+1} \quad (5.142)$$

La méthode de Gosper-Zeilberger est-elle capable de résoudre toutes les sommes que nous avons rencontrées dans ce chapitre ? Non. Elle ne marche pas lorsque  $t(n, k)$  est le terme général  $\binom{n}{k} (k+1)^{k-1} (n-k+1)^{n-k-1}$  de (5.66), car le rapport des termes  $t(n, k+1)/t(n, k)$  n'est pas une fonction rationnelle de  $k$ . Elle échoue aussi dans des cas comme  $t(n, k) = \binom{n}{k} n^k$ , car l'autre rapport des termes,  $t(n+1, k)/t(n, k)$ , n'est pas une fonction rationnelle de  $k$  (on peut quand même s'en tirer en sommant  $\binom{n}{k} z^k$  puis en posant  $z = n$ ). Elle échoue encore pour des termes généraux relativement simples comme  $t(n, k) = 1/(nk+1)$ , alors même que

*"Professor Littlewood, when he makes use of an algebraic identity, always saves himself the trouble of proving it; he maintains that an identity, if true, can be verified in a few lines by anybody obtuse enough to feel the need of verification. My object in the following pages is to confute this assertion."*

—F. J. Dyson [89]

$t(n, k+1)/t(n, k)$  et  $t(n+1, k)/t(n, k)$  sont des fonctions rationnelles de  $n$  et  $k$  (voir l'exercice 107).

En revanche, le succès de l'algorithme de Gosper-Zeilberger est garanti pour un nombre extrêmement grand de cas, à savoir lorsque le terme général  $t(n, k)$  est ce qu'on appelle un *terme propre*, c'est-à-dire un terme qui peut s'écrire sous la forme

$$t(n, k) = f(n, k) \frac{(a_1 n + a'_1 k + a''_1)! \dots (a_p n + a'_p k + a''_p)!}{(b_1 n + b'_1 k + b''_1)! \dots (b_q n + b'_q k + b''_q)!} w^n z^k. \quad (5.143)$$

Ici,  $f(n, k)$  est un polynôme en  $n$  et  $k$ ; les coefficients  $a_1, a'_1, \dots, a_p, a'_p, b_1, b'_1, \dots, b_q, b'_q$  sont des constantes entières spécifiques; les paramètres  $w$  et  $z$  sont non nuls; les  $a''_1, \dots, a''_p, b''_1, \dots, b''_q$  sont des nombres complexes quelconques. Nous allons prouver que, si  $t(n, k)$  est un terme propre, alors il existe des polynômes  $\beta_0(n), \dots, \beta_l(n)$  non tous nuls et un terme propre  $T(n, k)$  tel que

$$\beta_0(n)t(n, k) + \dots + \beta_l(n)t(n+l, k) = T(n, k+1) - T(n, k). \quad (5.144)$$

Que se passe-t-il si  
 $t(n, k)$  est indépen-  
dant de  $n$  ?

La preuve qui suit est due à Wilf et Zeilberger [374].

Soit  $N$  l'opérateur qui incrémente  $n$  de 1, et  $K$  l'opérateur qui incrémente  $k$  de 1, de façon que, par exemple,  $N^2 K^3 t(n, k) = t(n+2, k+3)$ . Nous allons étudier les opérateurs de différence linéaire en  $N$ ,  $K$ , et  $n$ , c'est-à-dire les polynômes d'opérateurs de la forme

$$H(N, K, n) = \sum_{i=0}^I \sum_{j=0}^J \alpha_{i,j}(n) N^i K^j, \quad (5.145)$$

où chaque  $\alpha_{i,j}(n)$  est un polynôme en  $n$ . Montrons pour commencer que si  $t(n, k)$  est un terme propre et  $H(N, K, n)$  un opérateur de différence linéaire, alors  $H(N, K, n)t(n, k)$  est un terme propre. Supposons que  $t$  et  $H$  sont donnés respectivement par (5.143) et (5.145), et définissons un "terme de base"

$$\bar{t}(n, k)_{I,J} = \frac{\prod_{i=1}^p (a_i n + a'_i k + a_i I[a_i < 0] + a'_i J[a'_i < 0] + a''_i)!}{\prod_{i=1}^q (b_i n + b'_i k + b_i I[b_i > 0] + b'_i J[b'_i > 0] + b''_i)!} w^n z^k.$$

Par exemple, si  $t(n, k)$  est égal à  $\binom{n-2k}{k} = (n-2k)!/k! (n-3k)!$ , le terme de base correspondant à un opérateur de différence linéaire de degrés  $I$  et  $J$  est  $\bar{t}(n, k)_{I,J} = (n-2k-2J)!/(k+J)! (n-3k+I)!$ . L'important est que, pour tous  $0 \leq i \leq I$  et  $0 \leq j \leq J$ ,  $\alpha_{i,j}(n)N^i K^j t(n, k)$  est égal à  $\bar{t}(n, k)_{I,J}$  multiplié par un polynôme en  $n$  et  $k$ . Comme une somme finie de polynômes est un polynôme,  $H(N, K, n)t(n, k)$  est bien de la forme requise dans (5.143).

L'étape suivante consiste à montrer que, pour tout terme propre  $t(n, k)$ , il existe un opérateur de différence linéaire non nul  $H(N, K, n)$  tel que

$$H(N, K, n)t(n, k) = 0.$$

Si  $0 \leq i \leq I$  et  $0 \leq j \leq J$ , le terme décalé  $N^i K^j t(n, k)$  est égal à  $\bar{t}(n, k)_{I,j}$  multiplié par un polynôme en  $n$  et  $k$  de degré au plus

$$\begin{aligned} D_{I,J} &= \deg(f) + |a_1|I + |a'_1|J + \cdots + |a_p|I + |a'_p|J \\ &\quad + |b_1|I + |b'_1|J + \cdots + |b_q|I + |b'_q|J \end{aligned}$$

en  $k$ . Donc, nous trouverons le  $H$  que nous cherchons si nous pouvons résoudre un système de  $D_{I,J} + 1$  équations linéaires homogènes en les  $(I + 1)(J + 1)$  variables  $\alpha_{i,j}(n)$ , dont les coefficients sont des polynômes en  $n$ . Il nous suffit pour cela de choisir  $I$  et  $J$  suffisamment grands pour que  $(I + 1)(J + 1) > D_{I,J} + 1$ . Nous pouvons par exemple prendre  $I = 2A' + 1$  et  $J = 2A + \deg(f)$ , où

$$\begin{aligned} A &= |a_1| + \cdots + |a_p| + |b_1| + \cdots + |b_q|; \\ A' &= |a'_1| + \cdots + |a'_p| + |b'_1| + \cdots + |b'_q|. \end{aligned}$$

Pour finir la preuve, nous allons déduire du fait que  $H(N, K, n)t(n, k) = 0$  une solution de (5.144). Choisissons  $H$  de façon à minimiser  $J$ , c'est-à-dire de sorte que le degré de  $H$  en  $K$  soit le plus petit possible. Il existe un opérateur de différence linéaire  $G(N, K, n)$  tel que

$$H(N, K, n) = H(N, 1, n) - (K - 1)G(N, K, n).$$

Soient maintenant  $H(N, 1, n) = \beta_0(n) + \beta_1(n)N + \cdots + \beta_l(n)N^l$  et  $T(n, k) = G(N, K, n)t(n, k)$ . Alors  $T(n, k)$  est un terme propre et (5.144) est vérifiée.

Pour que la preuve soit complète, il nous reste à vérifier que  $H(N, 1, n)$  n'est pas simplement l'opérateur nul. Supposons qu'il l'est. Dans ce cas,  $T(n, k)$  ne dépend pas de  $k$ ; alors il existe des polynômes  $\beta_0(n)$  and  $\beta_1(n)$  tels que  $(\beta_0(n) + \beta_1(n)N)T(n, k) = 0$ . Donc  $(\beta_0(n) + \beta_1(n)N)G(N, K, n)$  est un opérateur de différence linéaire non nul de degré  $J - 1$  qui annule  $t(n, k)$ . Ceci contredit le fait que  $J$  est minimal. Nous en avons terminé avec notre preuve.

Une fois qu'on sait qu'il existe un terme propre  $T$  qui vérifie (5.144), on est absolument certain que l'algorithme de Gosper va trouver  $T$  (éventuellement à une constante près). Nous n'avons prouvé l'algorithme de Gosper que pour des termes hypergéométriques  $t(k)$  en une seule variable  $k$ . Toutefois, on peut étendre la démonstration au cas de deux variables. Voici comment : il existe une infinité de nombres complexes  $n$  pour lesquels la condition (5.118) est valide lorsque  $q(n, k)$  et  $r(n, k)$  sont complètement

Ce truc consiste à considérer  $H$  comme un polynôme en  $K$ , puis à remplacer  $K$  par  $\Delta + 1$ .

factorisés en tant que polynômes en  $k$ , et pour lesquels le calcul de  $d$  lors de l'étape 2 s'accorde avec le calcul de l'algorithme de Gosper à une variable. Nous avons montré précédemment que, pour tous ces  $n$ , il existe un polynôme  $s(n, k)$  en  $k$  adéquat ; donc il existe un polynôme  $s(n, k)$  en  $n$  et  $k$  adéquat. CQFD.

Nous avons démontré que l'algorithme de Gosper-Zeilberger trouve une solution de (5.144), pour un certain  $l$  aussi petit que possible. Cette solution nous donne une récurrence en  $n$  pour calculer la somme sur  $k$  de tout terme propre  $t(n, k)$ , dès lors que  $t(n, k)$  n'est non nul que pour un nombre fini de  $k$ . Les rôles de  $n$  et  $k$  peuvent évidemment être inversés, car la définition d'un terme propre en (5.143) est symétrique en  $n$  et  $k$ .

On trouvera dans les exercices 98 à 108 d'autres exemples d'applications de l'algorithme Gosper-Zeilberger, qui donnent une idée de ses multiples talents. Wilf et Zeilberger [374] ont considérablement étendu ces résultats, pour concevoir des méthodes qui permettent de manipuler des coefficients binomiaux généralisés et des indices de sommation multiples.

## Exercices

### Echauffements

- 1 Combien vaut  $11^4$  ? Pourquoi ce nombre est-il facile à calculer pour qui connaît les coefficients binomiaux ?
- 2 Pour quelle(s) valeur(s) de  $k$  le binomial  $\binom{n}{k}$  est-il maximum, si  $n$  est un entier strictement positif donné ? Prouvez votre réponse.
- 3 Démontrez la propriété de l'hexagone,

$$\binom{n-1}{k-1} \binom{n}{k+1} \binom{n+1}{k} = \binom{n-1}{k} \binom{n+1}{k+1} \binom{n}{k-1}.$$

- 4 Calculez  $\binom{-1}{k}$  en effectuant un changement de signe de l'indice du haut .
- 5 Soit  $p$  un nombre premier. Montrez que  $\binom{p}{k} \bmod p = 0$  pour tout  $0 < k < p$ . Qu'est-ce que cela implique pour les coefficients binomiaux  $\binom{p-1}{k}$  ?
- 6 Réparez l'erreur de calcul dans le problème 6, section 5.2, en appliquant comme il faut la symétrie. *Un cas de confusion d'identité.*
- 7 La formule (5.34) est-elle vraie aussi lorsque  $k < 0$  ?
- 8 Calculez

$$\sum_k \binom{n}{k} (-1)^k (1 - k/n)^n.$$

Quelle est la valeur approchée de cette somme lorsque  $n$  est très grand ?

*Suggestion :* la somme est égale à  $\Delta^n f(0)$ , pour une certaine fonction  $f$ .

- 9 Montrez que les exponentielles généralisées de (5.58) satisfont

$$\mathcal{E}_t(z) = \mathcal{E}(tz)^{1/t}, \quad \text{si } t \neq 0,$$

où  $\mathcal{E}(z)$  est une abréviation de  $\mathcal{E}_1(z)$ .

- 10 Montrez que  $-2(\ln(1-z) + z)/z^2$  est une fonction hypergéométrique.

- 11 Exprimez les deux fonctions

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots$$

$$\arcsin z = z + \frac{1 \cdot z^3}{2 \cdot 3} + \frac{1 \cdot 3 \cdot z^5}{2 \cdot 4 \cdot 5} + \frac{1 \cdot 3 \cdot 5 \cdot z^7}{2 \cdot 4 \cdot 6 \cdot 7} + \dots$$

sous forme de séries hypergéométriques.

- 12 Parmi les fonctions de  $k$  suivantes, quelles sont celles qui sont des termes hypergéométriques, selon la définition de la section 5.7 ? Argumentez vos réponses.

a  $n^k$ .

b  $k^n$ .

c  $(k! + (k+1)!)/2$ .

d  $H_k$ , c'est-à-dire  $\frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{k}$ .

e  $1/\binom{n}{k}$ .

f  $t(k)T(k)$ , où  $t$  et  $T$  sont des termes hypergéométriques.

g  $t(k) + T(k)$ , où  $t$  et  $T$  sont des termes hypergéométriques.

h  $t(n-k)$ , où  $t$  est un terme hypergéométrique.

i  $a t(k) + b t(k+1) + c t(k+2)$ , où  $t$  est un terme hypergéométrique.

j  $\lceil k/2 \rceil$ .

k  $k [k > 0]$ .

(Ici,  $t$  et  $T$  ne sont pas nécessairement liés comme en (5.120)).

### Exercices de base

- 13 Trouvez des relations entre la fonction superfactorielle  $P_n = \prod_{k=1}^n k!$  de l'exercice 4.55, la fonction hyperfactorielle  $Q_n = \prod_{k=1}^n k^k$  et le produit  $R_n = \prod_{k=0}^n \binom{n}{k}$ .

- 14 Prouvez l'identité (5.25) en faisant un changement de signe de l'indice du haut dans la convolution de Vandermonde (5.22) ; puis montrez qu'avec un autre changement de signe on obtient (5.26).

- 15 Que vaut  $\sum_k \binom{n}{k}^3 (-1)^k$  ? *Suggestion :* voyez (5.29).

- 16 Calculez la somme  $\sum_k \binom{2a}{a+k} \binom{2b}{b+k} \binom{2c}{c+k} (-1)^k$ , où  $a, b, c$  sont des entiers positifs ou nuls.

17 Trouvez une relation simple entre  $\binom{2n-1/2}{n}$  et  $\binom{2n-1/2}{2n}$ .

18 Trouvez une autre expression, analogue à (5.35), pour le produit

$$\binom{r}{k} \binom{r-1/3}{k} \binom{r-2/3}{k}.$$

19 Montrez que les binomiaux généralisés de (5.58) obéissent à la règle

$$\mathcal{B}_t(z) = \mathcal{B}_{1-t}(-z)^{-1}.$$

20 Soit la "série bloopergéométrique généralisée", définie par la formule

$$G\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) = \sum_{k \geq 0} \frac{a_1^k \dots a_m^k}{b_1^k \dots b_n^k} \frac{z^k}{k!},$$

où les puissances montantes de (5.76) sont remplacées par des puissances descendantes. Exprimez G en fonction de F.

21 Prouvez que la définition des factorielles donnée par Euler est en accord avec la définition ordinaire, en montrant que la limite définie en (5.83) est égale à  $1/m!$  lorsque  $z = m$  est un entier strictement positif.

22 Utilisez (5.83) pour démontrer la *formule de duplication factorielle* :

$$x! (x - \frac{1}{2})! = (2x)! (-\frac{1}{2})! / 2^{2x}. \quad \text{A propos, } (-\frac{1}{2})! = \sqrt{\pi}.$$

23 Que vaut  $F(-n, 1; ; 1)$  ?

24 Calculez  $\sum_k \binom{n}{m+k} \binom{m+k}{2k} 4^k$  en utilisant les séries hypergéométriques.

25 Montrez que

$$\begin{aligned} (a_1 - b_1) F\left(\begin{matrix} a_1, a_2, \dots, a_m \\ b_1+1, b_2, \dots, b_n \end{matrix} \middle| z\right) \\ = a_1 F\left(\begin{matrix} a_1+1, a_2, \dots, a_m \\ b_1+1, b_2, \dots, b_n \end{matrix} \middle| z\right) - b_1 F\left(\begin{matrix} a_1, a_2, \dots, a_m \\ b_1, b_2, \dots, b_n \end{matrix} \middle| z\right). \end{aligned}$$

Trouvez une relation similaire entre les séries hypergéométriques

$$F\left(\begin{matrix} a_1, a_2, a_3, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right),$$

$$F\left(\begin{matrix} a_1+1, a_2, a_3, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right), \quad \text{et}$$

$$F\left(\begin{matrix} a_1, a_2+1, a_3, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right).$$

**26** Exprimez le  $G(z)$  de la formule

$$F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) = 1 + G(z)$$

comme un multiple d'une série hypergéométrique.

**27** Prouvez que

$$\begin{aligned} F\left(\begin{matrix} a_1, a_1 + \frac{1}{2}, \dots, a_m, a_m + \frac{1}{2} \\ b_1, b_1 + \frac{1}{2}, \dots, b_n, b_n + \frac{1}{2}, \frac{1}{2} \end{matrix} \middle| (2^{m-n-1}z)^2\right) \\ = \frac{1}{2} \left( F\left(\begin{matrix} 2a_1, \dots, 2a_m \\ 2b_1, \dots, 2b_n \end{matrix} \middle| z\right) + F\left(\begin{matrix} 2a_1, \dots, 2a_m \\ 2b_1, \dots, 2b_n \end{matrix} \middle| -z\right) \right). \end{aligned}$$

**28** Démontrez l'*identité d'Euler*

$$F\left(\begin{matrix} a, b \\ c \end{matrix} \middle| z\right) = (1-z)^{c-a-b} F\left(\begin{matrix} c-a, c-b \\ c \end{matrix} \middle| z\right)$$

en appliquant deux fois la règle de réflexion de Pfaff (5.101).

**29** Montrez que les séries hypergéométriques confluentes satisfont

$$e^z F\left(\begin{matrix} a \\ b \end{matrix} \middle| -z\right) = F\left(\begin{matrix} b-a \\ b \end{matrix} \middle| z\right).$$

**30** Quelle série hypergéométrique  $F$  satisfait l'équation  $zF'(z) + F(z) = 1/(1-z)$  ?

**31** Montrez que si  $f(k)$  est une fonction sommable en termes hypergéométriques, alors  $f$  elle-même est un terme hypergéométrique. Par exemple, si  $\sum f(k) \delta k = cF(A_1, \dots, A_M; B_1, \dots, B_N; Z)_k + C$ , alors il existe des constantes  $a_1, \dots, a_m, b_1, \dots, b_n$  et  $z$  telles que  $f(k)$  est un multiple de (5.115).

**32** Calculez  $\sum k^2 \delta k$  par la méthode de Gosper.

**33** Appliquez la méthode de Gosper pour calculer  $\sum \delta k/(k^2 - 1)$ .

**34** Montrez qu'une somme hypergéométrique partielle peut toujours être représentée par une limite d'une série hypergéométrique ordinaire :

$$\sum_{k \leq c} F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right)_k = \lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} -c, a_1, \dots, a_m \\ \epsilon - c, b_1, \dots, b_n \end{matrix} \middle| z\right),$$

si  $c$  est un entier positif ou nul (voir (5.115)). Utilisez ce fait pour évaluer  $\sum_{k \leq m} \binom{n}{k} (-1)^k$ .

**Devoirs à la maison**

- 35 La notation  $\sum_{k \leq n} \binom{n}{k} 2^{k-n}$  est ambiguë en dehors de tout contexte. Evaluez la
- en la considérant comme une somme sur  $k$  ;
  - en la considérant comme une somme sur  $n$ .
- 36 Soit  $p^k$  la plus grande puissance du nombre premier  $p$  qui divise  $\binom{m+n}{m}$ , où  $m$  et  $n$  sont des entiers positifs ou nuls. Démontrez que  $k$  est le nombre de retenues effectuées lorsqu'on additionne  $m$  et  $n$  en base  $p$ . *Suggestion* : l'exercice 4.24 peut vous aider.
- 37 Montrez qu'il existe un analogue de la formule du binôme pour les puissances factorielles. En d'autres termes, prouvez les identités

$$(x+y)^n = \sum_k \binom{n}{k} x^k y^{n-k},$$

$$(x+y)^{\bar{n}} = \sum_k \binom{n}{k} x^{\bar{k}} y^{\bar{n-k}},$$

pour tout entier positif ou nul  $n$ .

- 38 Montrez que tout entier positif ou nul  $n$  peut s'écrire de façon unique sous la forme  $n = \binom{a}{1} + \binom{b}{2} + \binom{c}{3}$ , où  $a$ ,  $b$  et  $c$  sont des entiers tels que  $0 \leq a < b < c$  (*c'est ce qu'on appelle le système de numération binomial*).
- 39 Montrez que si  $xy = ax + by$ , alors

$$x^n y^n = \sum_{k=1}^n \binom{2n-1-k}{n-1} (a^n b^{n-k} x^k + a^{n-k} b^n y^k)$$

pour tout  $n > 0$ . Trouvez une formule similaire pour le produit plus général  $x^m y^n$ . (On obtient ainsi des développements en éléments simples parfois bien utiles, par exemple pour  $x = 1/(z-c)$  et  $y = 1/(z-d)$ ).

- 40 Trouvez une forme close pour

$$\sum_{j=1}^m (-1)^{j+1} \binom{r}{j} \sum_{k=1}^n \binom{-j + rk + s}{m-j}, \quad m, n \geq 0 \text{ entiers.}$$

- 41 Calculez  $\sum_k \binom{n}{k} k! / (n+1+k)!$ , où  $n$  est un entier positif ou nul.
- 42 Calculez la somme indéfinie  $\sum ((-1)^x / \binom{n}{x}) \delta x$ , puis utilisez le résultat pour trouver une forme close de  $\sum_{k=0}^n (-1)^k / \binom{n}{k}$ .
- 43 Démontrez l'identité à trois binomiaux (5.28). *Suggestion* : commencez par remplacer  $\binom{r+k}{m+n}$  par  $\sum_j \binom{r}{m+n-j} \binom{k}{j}$ .

- 44** Utilisez l'identité (5.32) pour trouver des formes closes pour les sommes doubles

$$\sum_{j,k} (-1)^{j+k} \binom{j+k}{j} \binom{a}{j} \binom{b}{k} \binom{m+n-j-k}{m-j} \quad \text{et}$$

$$\sum_{j,k \geq 0} (-1)^{j+k} \binom{a}{j} \binom{m}{j} \binom{b}{k} \binom{n}{k} / \binom{m+n}{j+k},$$

où  $m \geq a \geq 0$  et  $n \geq b \geq 0$  sont donnés.

- 45** Trouvez une forme close pour  $\sum_{k \leq n} \binom{2k}{k} 4^{-k}$ .

- 46** Trouvez une forme close de la somme suivante, où  $n$  est un entier strictement positif :

$$\sum_k \binom{2k-1}{k} \binom{4n-2k-1}{2n-k} \frac{(-1)^{k-1}}{(2k-1)(4n-2k-1)}.$$

*Suggestion* : pensez aux fonctions génératrices.

- 47** La somme

$$\sum_k \binom{rk+s}{k} \binom{rn-rk-s}{n-k}$$

est un polynôme en  $r$  et  $s$ . Montrez qu'en fait elle ne dépend pas de  $s$ .

- 48** L'identité  $\sum_{k \leq n} \binom{n+k}{n} 2^{-k} = 2^n$  peut être combinée avec la formule  $\sum_{k \geq 0} \binom{n+k}{n} z^k = 1/(1-z)^{n+1}$  pour donner

$$\sum_{k>n} \binom{n+k}{n} 2^{-k} = 2^n.$$

Quel est l'équivalent hypergéométrique de cette dernière identité ?

- 49** Utilisez la méthode hypergéométrique pour calculer

$$\sum_k (-1)^k \binom{x}{k} \binom{x+n-k}{n-k} \frac{y}{y+n-k}.$$

- 50** Démontrez la règle de réflexion de Pfaff (5.101) en comparant les coefficients de  $z^n$  des deux membres de l'équation.

- 51** On montre dans le calcul de (5.104) que

$$\lim_{\epsilon \rightarrow 0} F(-m, -2m-1+\epsilon; -2m+\epsilon; 2) = 1/\left(\frac{-1/2}{m}\right).$$

Dans cet exercice, nous allons voir qu'il suffit d'une légère différence dans le mode de passage à la limite pour que la série hypergéométrique

dégénérée  $F(-m, -2m-1; -2m; 2)$  donne des réponses totalement différentes.

a Montrez que  $\lim_{\epsilon \rightarrow 0} F(-m + \epsilon, -2m - 1; -2m + 2\epsilon; 2) = 0$ , en utilisant la règle de réflexion de Pfaff pour prouver que  $F(a, -2m-1; 2a; 2) = 0$  pour tout entier  $m \geq 0$ .

b Que vaut  $\lim_{\epsilon \rightarrow 0} F(-m + \epsilon, -2m - 1; -2m + \epsilon; 2)$  ?

52 Montrez que si  $N$  est un entier positif ou nul, alors

$$\begin{aligned} & b_1^N \dots b_n^N F \left( \begin{matrix} a_1, \dots, a_m, -N \\ b_1, \dots, b_n \end{matrix} \middle| z \right) \\ &= a_1^N \dots a_m^N (-z)^N F \left( \begin{matrix} 1-b_1-N, \dots, 1-b_n-N, -N \\ 1-a_1-N, \dots, 1-a_m-N \end{matrix} \middle| \frac{(-1)^{m+n}}{z} \right). \end{aligned}$$

53 Si on pose  $b = -\frac{1}{2}$  et  $z = 1$  dans l'identité de Gauss (5.110), le membre gauche se réduit à  $-1$  tandis que celui de droite donne  $+1$ . Pourquoi n'est-ce pas une preuve que  $-1 = +1$ ?

54 Expliquez comment obtenir le membre droit de (5.112).

55 Si les termes hypergéométriques  $t(k) = F(a_1, \dots, a_m; b_1, \dots, b_n; z)_k$  et  $T(k) = F(A_1, \dots, A_M; B_1, \dots, B_N; Z)_k$  satisfont  $t(k) = c(T(k+1) - T(k))$  pour tout  $k \geq 0$ , montrez que  $z = Z$  et  $m - n = M - N$ .

56 Trouvez une formule générale pour  $\sum \binom{-3}{k} \delta k$  en appliquant la méthode de Gosper. Montrez que  $(-1)^{k-1} \lfloor \frac{k+1}{2} \rfloor \lfloor \frac{k+2}{2} \rfloor$  est aussi une solution.

57 Utilisez la méthode de Gosper pour trouver une constante  $\theta$  telle que

$$\sum \binom{n}{k} z^k (k + \theta) \delta k$$

soit sommable en termes hypergéométriques.

58 Si  $m$  et  $n$  sont des entiers tels que  $0 \leq m \leq n$ , soit

$$T_{m,n} = \sum_{0 \leq k < n} \binom{k}{m} \frac{1}{n-k}.$$

Trouvez une relation entre  $T_{m,n}$  et  $T_{m-1,n-1}$ , puis résolvez votre récurrence en appliquant un facteur de sommation.

### Problèmes d'examen

59 Trouvez une forme close pour

$$\sum_{k \geq 1} \binom{n}{[\log_m k]}$$

lorsque  $m$  et  $n$  sont des entiers strictement positifs.

- 60 Utilisez la formule d'approximation de Stirling (4.23) pour estimer  $\binom{m+n}{n}$  lorsque  $m$  et  $n$  sont grands tous les deux. A quoi se ramène votre formule lorsque  $m = n$  ?

- 61 Démontrez que si  $p$  est premier, alors

$$\binom{n}{m} \equiv \left( \frac{\lfloor n/p \rfloor}{\lfloor m/p \rfloor} \right) \left( \frac{n \bmod p}{m \bmod p} \right) \pmod{p},$$

pour tous entiers positifs ou nuls  $m$  et  $n$ .

- 62 En supposant que  $p$  est premier et que  $m$  et  $n$  sont des entiers strictement positifs, déterminez la valeur de  $\binom{np}{mp} \bmod p^2$ . *Suggestion :* vous pourriez utiliser la généralisation suivante de la convolution de Vandermonde :

$$\sum_{k_1+k_2+\dots+k_m=n} \binom{r_1}{k_1} \binom{r_2}{k_2} \dots \binom{r_m}{k_m} = \binom{r_1+r_2+\dots+r_m}{n}.$$

- 63 Trouvez une forme close pour

$$\sum_{k=0}^n (-4)^k \binom{n+k}{2k},$$

où  $n \geq 0$  est un entier donné.

- 64 Calculez  $\sum_{k=0}^n \binom{n}{k} / \left\lceil \frac{k+1}{2} \right\rceil$ , où  $n \geq 0$  est un entier donné.

- 65 Démontrez que

$$\sum_k \binom{n-1}{k} n^{-k} (k+1)! = n.$$

- 66 Calculez la “double somme de Harry”

$$\sum_{0 \leq j \leq k} \binom{-1}{j - \lfloor \sqrt{k-j} \rfloor} \binom{j}{m} \frac{1}{2^j}, \quad m \geq 0 \text{ entier},$$

qui est une fonction de  $m$  (la somme est à la fois sur  $j$  et  $k$ ).

- 67 Trouvez une forme close pour

$$\sum_{k=0}^n \binom{\binom{k}{2}}{2} \binom{2n-k}{n}, \quad n \geq 0 \text{ entier}.$$

- 68 Trouvez une forme close pour

$$\sum_k \binom{n}{k} \min(k, n-k), \quad n \geq 0 \text{ entier}.$$

**69** Trouvez une forme close pour

$$\min_{\substack{k_1, \dots, k_m \geq 0 \\ k_1 + \dots + k_m = n}} \sum_{j=1}^m \binom{k_j}{2}$$

en tant que fonction de  $m$  et  $n$ .

**70** Trouvez une forme close pour

$$\sum_k \binom{n}{k} \binom{2k}{k} \left(\frac{-1}{2}\right)^k, \quad n \geq 0 \text{ entier.}$$

**71** Soit  $S_n = \sum_{k \geq 0} \binom{n+k}{m+2k} a_k$ , où  $m$  et  $n$  sont des entiers positifs ou nuls, et soit  $A(z) = \sum_{k \geq 0} a_k z^k$  la fonction génératrice de la suite  $(a_0, a_1, a_2, \dots)$ .

- a Exprimez la fonction génératrice  $S(z) = \sum_{n \geq 0} S_n z^n$  en fonction de  $A(z)$ .
- b Utilisez cette technique pour résoudre le problème 7 de la section 5.2.

**72** Montrez que, si  $m$ ,  $n$  et  $k$  sont des entiers et si  $n > 0$ , alors

$$\binom{m/n}{k} n^{2k - \nu(k)} \text{ est un entier,}$$

où  $\nu(k)$  est le nombre de chiffres 1 dans la représentation binaire de  $k$ .

**73** Utilisez la méthode du répertoire pour résoudre la récurrence

$$\begin{aligned} X_0 &= \alpha; & X_1 &= \beta; \\ X_n &= (n-1)(X_{n-1} + X_{n-2}), & \text{pour } n > 1. \end{aligned}$$

*Suggestion :*  $n!$  et  $n!$  satisfont tous deux cette récurrence.

**74** Considérons une variante du triangle de Pascal, dans laquelle les côtés sont constitués des nombres 1, 2, 3, 4, ... au lieu d'une suite de 1. Toutefois, les autres nombres du triangle satisfont toujours la formule d'addition :

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & & 2 & 2 & & \\ & & & 3 & 4 & 3 & \\ & & & 4 & 7 & 7 & 4 \\ & & & 5 & 11 & 14 & 11 & 5 \\ & \cdot \end{array}$$

Si  $\binom{n}{k}$  désigne le  $k$ ème nombre de la ligne  $n$ , pour  $1 \leq k \leq n$ , on a  $\binom{n}{1} = \binom{n}{n} = n$ , et  $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$  pour  $1 < k < n$ . Donnez une forme close de  $\binom{n}{k}$ .

**75** Trouvez une relation entre les fonctions

$$S_0(n) = \sum_k \binom{n}{3k},$$

$$S_1(n) = \sum_k \binom{n}{3k+1},$$

$$S_2(n) = \sum_k \binom{n}{3k+2}$$

et les quantités  $\lfloor 2^n/3 \rfloor$  et  $\lceil 2^n/3 \rceil$ .

**76** Résolvez la récurrence suivante pour  $n, k \geq 0$  :

$$Q_{n,0} = 1; \quad Q_{0,k} = [k=0];$$

$$Q_{n,k} = Q_{n-1,k} + Q_{n-1,k-1} + \binom{n}{k}, \quad \text{pour } n, k > 0.$$

**77** Quelle est la valeur de

$$\sum_{0 \leq k_1, \dots, k_m \leq n} \prod_{1 \leq j \leq m} \binom{k_{j+1}}{k_j}, \quad \text{si } m > 1 ?$$

**78** En supposant que  $m$  est un entier strictement positif, trouvez une forme close pour

$$\sum_{k=0}^{2m^2} \binom{k \bmod m}{(2k+1) \bmod (2m+1)}.$$

**79 a** Quel est le plus grand commun diviseur de  $\binom{2n}{1}, \binom{2n}{3}, \dots, \binom{2n}{2n-1}$  ?

*Suggestion :* considérez la somme de ces  $n$  nombres.

**b** Montrez que le plus petit commun multiple de  $\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n}$  est égal à  $L(n+1)/(n+1)$ , où  $L(n) = \text{ppcm}(1, 2, \dots, n)$ .

*Bon à savoir.*

**80** Montrez que  $\binom{n}{k} \leq (en/k)^k$  pour tous entiers  $k, n \geq 0$ .

**81** Si  $0 < \theta < 1$  et  $0 \leq x \leq 1$ , et si  $l, m, n$  sont des entiers positifs ou nuls tels que  $m < n$ , prouvez l'inégalité

$$(-1)^{n-m-1} \sum_k \binom{l}{k} \binom{m+\theta}{n+k} x^k > 0.$$

*Suggestion :* pensez à dériver par rapport à  $x$ .

**Questions subsidiaires**

- 82** Montrez que le triangle de Pascal satisfait une propriété de l'hexagone plus surprenante encore que celle citée dans le texte du chapitre :

$$\operatorname{pgcd}\left(\binom{n-1}{k-1}, \binom{n}{k+1}, \binom{n+1}{k}\right) = \operatorname{pgcd}\left(\binom{n-1}{k}, \binom{n+1}{k+1}, \binom{n}{k-1}\right),$$

si  $0 < k < n$ . Par exemple,  $\operatorname{pgcd}(56, 36, 210) = \operatorname{pgcd}(28, 120, 126) = 2$ .

- 83** Démontrez l'identité à cinq paramètres et double somme (5.32).

- 84** Montrez que la seconde paire d'identités (5.61) découle de la première paire (5.60). *Suggestion* : dérivez par rapport à  $z$ .

- 85** Prouvez que

$$\begin{aligned} \sum_{m=1}^n (-1)^m \sum_{1 \leq k_1 < k_2 < \dots < k_m \leq n} \binom{k_1^3 + k_2^3 + \dots + k_m^3 + 2^n}{n} \\ = (-1)^n n!^3 - \binom{2^n}{n}. \end{aligned}$$

(Le membre gauche est une somme de  $2^n - 1$  termes). *Suggestion* : on peut prouver beaucoup plus.

- 86** Soient  $a_1, \dots, a_n$  des entiers positifs ou nuls, et soit  $C(a_1, \dots, a_n)$  le coefficient du terme constant  $z_1^{a_1} \dots z_n^{a_n}$  lorsque les  $n(n-1)$  facteurs

$$\prod_{\substack{1 \leq i, j \leq n \\ i \neq j}} \left(1 - \frac{z_i}{z_j}\right)^{a_i}$$

sont complètement développés en puissances positives et négatives des variables complexes  $z_1, \dots, z_n$ .

- a Prouvez que  $C(a_1, \dots, a_n)$  est égal au membre gauche de (5.31).  
b Prouvez que si  $z_1, \dots, z_n$  sont des nombres complexes distincts, alors le polynôme

$$f(z) = \sum_{k=1}^n \prod_{\substack{1 \leq j \leq n \\ j \neq k}} \frac{z - z_j}{z_k - z_j}$$

est identiquement égal à 1.

- c Multipliez le produit original de  $n(n-1)$  facteurs par  $f(0)$  et déduisez-en que  $C(a_1, a_2, \dots, a_n)$  est égal à

$$\begin{aligned} C(a_1 - 1, a_2, \dots, a_n) + C(a_1, a_2 - 1, \dots, a_n) \\ + \dots + C(a_1, a_2, \dots, a_n - 1). \end{aligned}$$

(Cette récurrence définit des coefficients multinomiaux, de sorte que  $C(a_1, \dots, a_n)$  est égal au membre droit de (5.31)).

87 Soient  $m$  un entier strictement positif et  $\zeta = e^{\pi i/m}$ . Montrez que

$$\begin{aligned} & \sum_{k \leq n/m} \binom{n-mk}{k} z^{mk} \\ &= \frac{\mathcal{B}_{-m}(z^m)^{n+1}}{(1+m)\mathcal{B}_{-m}(z^m) - m} \\ &- \sum_{0 \leq j < m} \frac{(\zeta^{2j+1}z\mathcal{B}_{1+1/m}(\zeta^{2j+1}z)^{1/m})^{n+1}}{(m+1)\mathcal{B}_{1+1/m}(\zeta^{2j+1}z)^{-1} - 1}. \end{aligned}$$

(Lorsque  $m = 1$ , on retrouve (5.74)).

88 Montrez que les coefficients  $s_k$  de (5.47) sont égaux à

$$(-1)^k \int_0^\infty e^{-t} (1 - e^{-t})^{k-1} \frac{dt}{t},$$

pour tout  $k > 1$ ; ainsi,  $|s_k| < 1/(k-1)$ .

89 Prouvez qu'il existe une somme infinie analogue à (5.19), à savoir

$$\sum_{k>m} \binom{m+r}{k} x^k y^{m-k} = \sum_{k>m} \binom{-r}{k} (-x)^k (x+y)^{m-k}, \quad m \text{ entier},$$

si  $|x| < |y|$  et  $|x| < |x+y|$ . Dérivez cette identité  $n$  fois par rapport à  $y$  et exprimez le résultat sous forme hypergéométrique. Quelle relation obtenez-vous?

90 Dans le premier problème de la section 5.2, on considère la somme  $\sum_{k \geq 0} \binom{r}{k} / \binom{s}{k}$ , où  $r$  et  $s$  sont des entiers tels que  $s \geq r \geq 0$ . Que vaut cette somme si  $r$  et  $s$  ne sont pas des entiers?

91 Prouvez l'*identité de Whipple*,

$$\begin{aligned} & F\left(\begin{matrix} \frac{1}{2}a, \frac{1}{2}a+\frac{1}{2}, 1+a-b-c \\ 1+a-b, 1+a-c \end{matrix} \middle| \frac{-4z}{(1-z)^2}\right) \\ &= (1-z)^a F\left(\begin{matrix} a, b, c \\ 1+a-b, 1+a-c \end{matrix} \middle| z\right), \end{aligned}$$

en montrant que les deux membres satisfont la même équation différentielle.

92 Prouvez les *identités de Clausen*

$$\begin{aligned} F\left(\begin{matrix} a, b \\ a+b+\frac{1}{2} \end{matrix} \middle| z\right)^2 &= F\left(\begin{matrix} 2a, a+b, 2b \\ 2a+2b, a+b+\frac{1}{2} \end{matrix} \middle| z\right); \\ F\left(\begin{matrix} \frac{1}{4}+a, \frac{1}{4}+b \\ 1+a+b \end{matrix} \middle| z\right) F\left(\begin{matrix} \frac{1}{4}-a, \frac{1}{4}-b \\ 1-a-b \end{matrix} \middle| z\right) &= F\left(\begin{matrix} \frac{1}{2}, \frac{1}{2}+a-b, \frac{1}{2}-a+b \\ 1+a+b, 1-a-b \end{matrix} \middle| z\right). \end{aligned}$$

## 270 COEFFICIENTS BINOMIAUX

Quelles identités trouve-t-on en mettant en équations les coefficients de  $z^n$  dans chacune de ces formules ?

- 93 Montrez que la somme indéfinie

$$\sum \left( \prod_{j=1}^{k-1} (f(j) + \alpha) \Big/ \prod_{j=1}^k f(j) \right) \delta k$$

admet une forme (assez) simple, quelles que soient la fonction  $f$  et la constante  $\alpha$  données.

- 94 Calculez  $\sum \binom{a}{k} \binom{-a}{n-k} \delta k$  lorsque  $n$  est un entier strictement positif.

- 95 Quelles conditions faut-il ajouter à (5.118) pour que les polynômes  $p$ ,  $q$  et  $r$  de (5.117) soient déterminés de façon unique ?

- 96 Prouvez que si l'algorithme de Gosper ne trouve pas de solution à (5.120), pour un terme hypergéométrique  $t(k)$  donné, alors l'équation plus générale suivante, où  $T_1(k), \dots, T_m(k)$  sont des termes hypergéométriques, n'admet pas de solution :

$$t(k) = (T_1(k+1) + \dots + T_m(k+1)) - (T_1(k) + \dots + T_m(k)).$$

- 97 Trouvez tous les nombres complexes  $z$  pour lesquels  $k!^2 / \prod_{j=1}^k (j^2 + jz + 1)$  est sommable en termes géométriques.

- 98 Quelle récurrence la méthode de Gosper-Zeilberger donne-t-elle pour la somme  $S_n = \sum_k \binom{n}{2k}$  ?

- 99 Utilisez la méthode de Gosper-Zeilberger pour trouver une forme close de  $\sum_k t(n, k)$  lorsque  $t(n, k) = (n+a+b+c+k)!/(n+k)! (c+k)! (b-k)! (a-k)! k!$ , en supposant que  $a$  est un entier positif ou nul.

- 100 Trouvez une relation de récurrence pour la somme

$$S_n = \sum_{k=0}^n \frac{1}{\binom{n}{k}},$$

puis utilisez cette récurrence pour trouver une autre formule pour  $S_n$ .

- 101 Donnez des relations de récurrence satisfaites par les sommes

a     $S_{m,n}(z) = \sum_k \binom{m}{k} \binom{n}{k} z^k;$

b     $S_n(z) = S_{n,n}(z) = \sum_k \binom{n}{k}^2 z^k.$

*Il est conseillé d'utiliser un logiciel de calcul formel pour résoudre celui-ci (ainsi que les suivants).*

**102** Utilisez la procédure de Gosper-Zeilberger pour généraliser l'identité “inutile” (5.113) : trouvez d'autres valeurs de  $a$ ,  $b$  et  $z$  telles que

$$\sum_k \binom{n}{k} \binom{\frac{1}{3}n - a}{k} z^k / \binom{\frac{4}{3}n - b}{k}$$

possède une forme close simple.

**103** Soit  $t(n, k)$  le terme propre (5.143). Quels sont les degrés en  $k$  de  $\hat{p}(n, k)$ ,  $q(n, k)$  et  $r(n, k)$  lorsqu'on applique la procédure de Gosper-Zeilberger à  $\hat{t}(n, k) = \beta_0(n)t(n, k) + \dots + \beta_l(n)t(n+l, k)$ ? Ne vous préoccupez pas des cas exceptionnels.

**104** Vérifiez, en appliquant la procédure de Gosper-Zeilberger, la remarquable identité

$$\sum_k (-1)^k \binom{r-s-k}{k} \binom{r-2k}{n-k} \frac{1}{r-n-k+1} = \binom{s}{n} \frac{1}{r-2n+1}.$$

Expliquez pourquoi l'algorithme ne trouve pas la récurrence la plus simple.

**105** Montrez que si  $\omega = e^{2\pi i/3}$ , alors

$$\sum_{k+l+m=3n} \binom{3n}{k, l, m}^2 \omega^{l-m} = \binom{4n}{n, n, 2n}, \quad n \geq 0 \text{ entier.}$$

**106** Prouvez la surprenante identité (5.32). Pour cela, Divisez le terme général de la somme par le membre droit de l'égalité, et notez  $t(r, j, k)$  ce rapport ; puis montrez qu'il existe des fonctions  $T(r, j, k)$  et  $U(r, j, k)$  telles que  $t(r+1, j, k) - t(r, j, k) = T(r, j+1, k) - T(r, j, k) + U(r, j, k+1) - U(r, j, k)$ .

**107** Prouvez que  $1/(nk+1)$  n'est pas un terme propre.

**108** Montrez que les nombres d'Apéry  $A_n$  de (5.141) sont les éléments diagonaux  $A_{n,n}$  de la matrice des nombres définis par

$$A_{m,n} = \sum_{j,k} \binom{m}{j}^2 \binom{m}{k}^2 \binom{2m+n-j-k}{2m}:$$

Prouvez en fait que cette matrice est symétrique et que

$$\begin{aligned} A_{m,n} &= \sum_k \binom{m+n-k}{k}^2 \binom{m+n-2k}{m-k}^2 \\ &= \sum_k \binom{m}{k} \binom{n}{k} \binom{m+k}{k} \binom{n+k}{k}. \end{aligned}$$

## 272 COEFFICIENTS BINOMIAUX

**109** Montrez que les nombres d'Apéry (5.141) satisfont

$$A_n \equiv A_{\lfloor n/p \rfloor} A_{n \bmod p} \pmod{p}$$

pour tout nombre premier  $p$  et tout entier  $n \geq 0$ .

### Problèmes de recherche

**110** Pour quelles valeurs de  $n$  la congruence  $\binom{2n}{n} \equiv (-1)^n \pmod{(2n+1)}$  est-elle valide ?

**111** Soit  $q(n)$  le plus petit facteur premier impair du coefficient binomial central  $\binom{2n}{n}$ . Selon l'exercice 36, les nombres premiers impairs qui ne divisent pas  $\binom{2n}{n}$  sont ceux pour lesquels tous les chiffres dans la représentation de  $n$  en base  $p$  sont inférieurs ou égaux à  $(p-1)/2$ . Des calculs sur ordinateur ont montré que  $q(n) \leq 11$  pour  $1 < n < 10^{10000}$ , à l'exception de  $q(3160) = 13$ .

a Est-ce que  $q(n) \leq 11$  pour tout  $n > 3160$  ?

b Est-ce que  $q(n) = 11$  pour une infinité de  $n$  ?

La somme de \$7 · 11 · 13 sera offerte pour une solution de (a) ou de (b).

**112** Est-il vrai que le binomial  $\binom{2n}{n}$  est toujours divisible soit par 4 soit par 9 pour tout  $n > 4$  sauf  $n = 64$  et  $n = 256$  ?

**113** Si  $t(n+1, k)/t(n, k)$  et  $t(n, k+1)/t(n, k)$  sont des fonctions rationnelles de  $n$  et  $k$ , et s'il existe un opérateur de différence linéaire non nul  $H(N, K, n)$  tel que  $H(N, K, n)t(n, k) = 0$ , est-il vrai que  $t(n, k)$  est un terme propre ?

**114** Soit  $m$  un entier strictement positif, et soit la suite  $c_n^{(m)}$  définie par la récurrence

$$\sum_k \binom{n}{k}^m \binom{n+k}{k}^m = \sum_k \binom{n}{k} \binom{n+k}{k} c_k^{(m)}.$$

Les nombres  $c_n^{(m)}$  sont-ils entiers ?

# 6

## Nombres remarquables

CERTAINES SUITES de nombres apparaissent si souvent en mathématiques qu'on les reconnaît instantanément. Par exemple, quiconque fait de l'arithmétique connaît la suite des nombres carrés  $\langle 1, 4, 9, 16, \dots \rangle$ . Dans le chapitre 1, nous avons rencontré les nombres triangulaires  $\langle 1, 3, 6, 10, \dots \rangle$  ; au chapitre 4, nous avons étudié les nombres premiers  $\langle 2, 3, 5, 7, \dots \rangle$  ; au chapitre 5, nous avons eu un aperçu des nombres de Catalan  $\langle 1, 2, 5, 14, \dots \rangle$ .

Dans le présent chapitre, nous allons découvrir quelques autres suites importantes, en commençant par les nombres de Stirling  $\{n\}_k$  et  $[n]_k$ , et les nombres eulériens  $\langle n \rangle_k$ . Ces trois suites de nombres forment des motifs triangulaires, tout comme les coefficients binomiaux  $\binom{n}{k}$  dans le triangle de Pascal. Puis nous examinerons les nombres harmoniques  $H_n$  et jetterons un coup d'oeil sur les nombres de Bernoulli  $B_n$ . Ces deux suites diffèrent des précédentes du fait que leurs éléments sont non pas des entiers, mais des fractions. Pour finir, nous étudierons la fascinante suite des nombres de Fibonacci  $F_n$  et quelques-unes de leurs généralisations.

### 6.1 NOMBRES DE STIRLING

Commençons donc par des nombres proches des coefficients binomiaux, les nombres de Stirling, du nom de James Stirling (1692–1770). Il en existe en fait deux types, traditionnellement appelés “nombres de Stirling de première espèce” et “nombres de Stirling de deuxième espèce”. Bien qu'ils soient chargés d'histoire et apparaissent dans de très nombreuses applications, ils ne disposent pas d'une notation standard. En ce qui nous concerne, nous allons suivre l'exemple de Jovan Karamata, et écrire  $\{n\}_k$  pour les nombres de Stirling de deuxième espèce et  $[n]_k$  pour les nombres de Stirling de première espèce. Ces notations sont en effet plus pratiques que celles qui ont été proposées par maints autres mathématiciens.

“... par cette notation, les formules deviennent plus symétriques.”

—J. Karamata [199]

Les tables 274 et 275 montrent que  $\{n\}_k$  et  $[n]_k$  sont proches lorsque  $n$  et  $k$  sont petits. Lorsqu'on voit apparaître les nombres “1, 4, 6, 4, 1”

Table 274 Le triangle de Stirling pour les sous-ensembles.

$n$	$\{ \begin{smallmatrix} n \\ 0 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 1 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 2 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 3 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 4 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 5 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 6 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 7 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 8 \end{smallmatrix} \}$	$\{ \begin{smallmatrix} n \\ 9 \end{smallmatrix} \}$
0	1									
1	0	1								
2	0	1	1							
3	0	1	3	1						
4	0	1	7	6	1					
5	0	1	15	25	10	1				
6	0	1	31	90	65	15	1			
7	0	1	63	301	350	140	21	1		
8	0	1	127	966	1701	1050	266	28	1	
9	0	1	255	3025	7770	6951	2646	462	36	1

dans un problème, on peut légitimement soupçonner qu'il existe un lien avec les coefficients binomiaux  $\binom{n}{k}$ . De la même façon, un problème où apparaissent les nombres "1, 7, 6, 1" est probablement lié aux  $\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \}$ , tandis qu'un problème où apparaissent "6, 11, 6, 1" semble plutôt lié aux  $\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \}$ . Ces trois suites correspondent à  $n = 4$ .

Les nombres de Stirling de deuxième espèce apparaissent en fait plus souvent que les autres. C'est pourquoi nous allons les étudier en premier. Le symbole  $\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \}$  représente le nombre de façons de partager un ensemble de  $n$  éléments en  $k$  sous-ensembles non vides. Un tel partage s'appelle une *partition d'ensemble*. Par exemple, il existe sept partitions d'un ensemble de quatre éléments en deux sous-ensembles :

$$\begin{aligned} & \{1, 2, 3\} \cup \{4\}, \quad \{1, 2, 4\} \cup \{3\}, \quad \{1, 3, 4\} \cup \{2\}, \quad \{2, 3, 4\} \cup \{1\}, \\ & \{1, 2\} \cup \{3, 4\}, \quad \{1, 3\} \cup \{2, 4\}, \quad \{1, 4\} \cup \{2, 3\}; \end{aligned} \quad (6.1)$$

donc  $\{ \begin{smallmatrix} 4 \\ 2 \end{smallmatrix} \} = 7$ . Remarquez que les accolades servent à représenter les ensembles aussi bien que les nombres  $\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \}$ . Cette similarité de notations nous aide à retenir la signification de  $\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \}$ . Regardons un peu ces nombres lorsque  $k$  est petit. Il y a exactement une façon de mettre  $n$  éléments dans un unique ensemble non vide ; donc  $\{ \begin{smallmatrix} n \\ 1 \end{smallmatrix} \} = 1$ , pour tout  $n > 0$ . En revanche,  $\{ \begin{smallmatrix} 0 \\ 1 \end{smallmatrix} \} = 0$ , du fait qu'un ensemble à 0 élément ne peut être que vide.

Le cas  $k = 0$  est un peu délicat. Pour que tout se passe bien, il nous faut convenir qu'il y a une et une seule façon de partager un ensemble vide en zéro part non vide, de sorte que  $\{ \begin{smallmatrix} 0 \\ 0 \end{smallmatrix} \} = 1$ . Par contre, un ensemble non vide requiert au moins une part, donc  $\{ \begin{smallmatrix} n \\ 0 \end{smallmatrix} \} = 0$  pour tout  $n > 0$ .

Que se passe-t-il lorsque  $k = 2$  ? Ce qui est sûr, c'est que  $\{ \begin{smallmatrix} 0 \\ 2 \end{smallmatrix} \} = 0$ . Si on divise un ensemble de  $n > 0$  objets en deux parts non vides, l'une de ces parts contient le dernier objet et un sous-ensemble d'objets choisis parmi

(Stirling lui-même considéra d'abord la deuxième espèce dans son livre [343]).

Table 275 Le triangle de Stirling pour les cycles.

$n$	$\begin{bmatrix} n \\ 0 \end{bmatrix}$	$\begin{bmatrix} n \\ 1 \end{bmatrix}$	$\begin{bmatrix} n \\ 2 \end{bmatrix}$	$\begin{bmatrix} n \\ 3 \end{bmatrix}$	$\begin{bmatrix} n \\ 4 \end{bmatrix}$	$\begin{bmatrix} n \\ 5 \end{bmatrix}$	$\begin{bmatrix} n \\ 6 \end{bmatrix}$	$\begin{bmatrix} n \\ 7 \end{bmatrix}$	$\begin{bmatrix} n \\ 8 \end{bmatrix}$	$\begin{bmatrix} n \\ 9 \end{bmatrix}$
0	1									
1	0	1								
2	0	1	1							
3	0	2	3	1						
4	0	6	11	6	1					
5	0	24	50	35	10	1				
6	0	120	274	225	85	15	1			
7	0	720	1764	1624	735	175	21	1		
8	0	5040	13068	13132	6769	1960	322	28	1	
9	0	40320	109584	118124	67284	22449	4536	546	36	1

les  $n - 1$  autres. Il y a  $2^{n-1}$  façons de choisir ce sous-ensemble, car chacun des  $n - 1$  premiers objets est soit dedans, soit dehors. Toutefois, il ne faut pas y mettre la totalité des  $n - 1$  objets, car sinon l'autre part serait vide. C'est pourquoi on soustrait 1 :

$$\begin{Bmatrix} n \\ 2 \end{Bmatrix} = 2^{n-1} - 1, \quad n > 0 \text{ entier.} \quad (6.2)$$

Ceci correspond bien au calcul de  $\begin{Bmatrix} 4 \\ 2 \end{Bmatrix} = 7 = 2^3 - 1$  effectué plus haut.

Généralisons cet argument pour aboutir à une récurrence qui nous permettra de calculer  $\begin{Bmatrix} n \\ k \end{Bmatrix}$  pour tout  $k$ . Soit un ensemble de  $n > 0$  objets à partager en  $k$  parts non vides. Pour ce faire, soit on met le dernier objet tout seul dans un sous-ensemble (de  $\begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix}$  façons possibles), soit on l'ajoute à un sous-ensemble non vide des  $n - 1$  premiers objets. Il y a  $k \begin{Bmatrix} n-1 \\ k \end{Bmatrix}$  possibilités dans ce dernier cas, car chacune des  $\begin{Bmatrix} n-1 \\ k \end{Bmatrix}$  façons de répartir les  $n - 1$  premiers objets dans  $k$  sous-ensembles non vides donne  $k$  choix possibles pour placer le dernier objet.

$$\begin{Bmatrix} n \\ k \end{Bmatrix} = k \begin{Bmatrix} n-1 \\ k \end{Bmatrix} + \begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix}, \quad n > 0 \text{ entier.} \quad (6.3)$$

C'est avec cette règle qu'on construit la table 274. Le facteur  $k$  fait la différence avec la formule d'addition (5.8) du triangle de Pascal.

Passons maintenant aux nombres de Stirling de première espèce :  $\begin{bmatrix} n \\ k \end{bmatrix}$  est le nombre de façons de répartir  $n$  objets dans  $k$  cycles. Un cycle est une configuration cyclique, comme les colliers que nous avons considérés au

## chapitre 4. Le cycle



peut s'écrire, de façon plus compacte, “[A, B, C, D]”, en convenant que

$$[A, B, C, D] = [B, C, D, A] = [C, D, A, B] = [D, A, B, C].$$

Ainsi, un cycle reste le même si on lui fait ainsi subir des décalages circulaires, car son début et sa fin sont reliés. Par contre, le cycle [A, B, C, D] est différent des cycles [A, B, D, C] et [D, C, B, A].

Il y a onze façons différentes de faire deux cycles avec quatre éléments :

$$\begin{array}{llll} [1, 2, 3] [4], & [1, 2, 4] [3], & [1, 3, 4] [2], & [2, 3, 4] [1], \\ [1, 3, 2] [4], & [1, 4, 2] [3], & [1, 4, 3] [2], & [2, 4, 3] [1], \\ [1, 2] [3, 4], & [1, 3] [2, 4], & [1, 4] [2, 3]. \end{array}$$

(6.4)

*"There are nine and sixty ways of constructing tribal lays, And every-single-one-of-them-is-right."*

—Rudyard Kipling

Par conséquent,  $\left[\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right] = 11$ .

Un cycle singleton (c'est-à-dire un cycle composé d'un unique élément) représente exactement la même chose qu'un ensemble singleton. De manière similaire, un cycle à deux éléments est identique à un ensemble à deux éléments, car  $[A, B] = [B, A]$ , tout comme  $\{A, B\} = \{B, A\}$ . En revanche, il existe deux cycles *differents* à trois éléments  $[A, B, C]$  et  $[A, C, B]$ . Remarquez, par exemple, qu'on peut obtenir les onze paires de cycles de (6.4) à partir des sept paires d'ensembles de (6.1) en construisant deux cycles à partir de chaque ensemble à trois éléments.

Pour tout  $n > 0$ , il existe  $n!/n = (n-1)!$  cycles de  $n$  éléments distincts. En effet, il y a  $n!$  permutations de  $n$  éléments, et chaque cycle correspond à  $n$  d'entre elles car chacun des éléments du cycle peut être le début d'une permutation. Par conséquent,

$$\left[\begin{smallmatrix} n \\ 1 \end{smallmatrix}\right] = (n-1)!, \quad n > 0 \text{ entier.} \quad (6.5)$$

C'est bien plus que la valeur  $\left\{\begin{smallmatrix} n \\ 1 \end{smallmatrix}\right\} = 1$  que nous avions pour le nombre de partitions en un sous-ensemble. En fait, il est facile de voir que tout nombre de Stirling de première espèce est supérieur ou égal au nombre de Stirling de deuxième espèce correspondant,

$$\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right] \geq \left\{\begin{smallmatrix} n \\ k \end{smallmatrix}\right\}, \quad n, k \geq 0 \text{ entiers,} \quad (6.6)$$

car à chaque partition d'un ensemble correspond au moins une configuration de cycles.

Il y a égalité en (6.6) lorsque tous les cycles sont des singletons ou des doubletons, car dans ce cas un cycle équivaut à un ensemble. Cela arrive lorsque  $k = n$  et lorsque  $k = n - 1$ . Ainsi,

$$\begin{bmatrix} n \\ n \end{bmatrix} = \left\{ \begin{matrix} n \\ n \end{matrix} \right\}; \quad \begin{bmatrix} n \\ n-1 \end{bmatrix} = \left\{ \begin{matrix} n \\ n-1 \end{matrix} \right\}.$$

En fait, on voit facilement que

$$\begin{bmatrix} n \\ n \end{bmatrix} = \left\{ \begin{matrix} n \\ n \end{matrix} \right\} = 1; \quad \begin{bmatrix} n \\ n-1 \end{bmatrix} = \left\{ \begin{matrix} n \\ n-1 \end{matrix} \right\} = \binom{n}{2}. \quad (6.7)$$

En effet, le nombre de façons de répartir  $n$  objets dans  $n - 1$  cycles ou sous-ensembles est égal au nombre de façons de choisir les deux objets qui seront dans le même cycle ou sous-ensemble. C'est pourquoi les nombres triangulaires  $\binom{n}{2} = 1, 3, 6, 10, \dots$  sont bien en évidence dans les tables 274 et 275.

Il suffit de modifier légèrement l'argument que nous avons utilisé dans le cas de  $\left\{ \begin{matrix} n \\ k \end{matrix} \right\}$  pour trouver une récurrence sur les  $\begin{bmatrix} n \\ k \end{bmatrix}$ . Dans toute répartition de  $n$  objets dans  $k$  cycles, soit le dernier objet constitue un cycle à lui seul (de  $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}$  façons différentes), soit il est inséré dans un cycle de l'une des  $\begin{bmatrix} n-1 \\ k \end{bmatrix}$  répartitions possibles des  $n - 1$  premiers objets. Dans ce dernier cas, il y a  $n - 1$  façons d'effectuer l'insertion. En effet, il est facile de vérifier qu'il y a  $j$  façons d'ajouter un élément dans un cycle de longueur  $j$  pour former un cycle de longueur  $j + 1$ . Si  $j = 3$  par exemple, le cycle [A, B, C] devient

$$[A, B, C, D], \quad [A, B, D, C], \quad \text{ou} \quad [A, D, B, C]$$

si on y insère un nouvel élément D ; ce sont les seules possibilités. En sommant sur tous les  $j$ , on obtient bien un total de  $n - 1$  façons d'insérer un  $n$ ième objet dans une décomposition en cycles de  $n$  objets. Voici donc la récurrence cherchée :

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n - 1) \begin{bmatrix} n - 1 \\ k \end{bmatrix} + \begin{bmatrix} n - 1 \\ k - 1 \end{bmatrix}, \quad n > 0 \text{ entier.} \quad (6.8)$$

C'est avec cet analogue de la formule d'addition que l'on construit la table 275.

Si on compare (6.8) et (6.3), on voit que le premier terme du membre droit est multiplié par son index du haut ( $n - 1$ ) dans le cas des cycles, et par son index du bas dans le cas des sous-ensembles. Il nous sera donc possible, lorsque nous ferons des preuves par induction notamment, de pratiquer l'absorption lorsque nous rencontrerons des termes tels que  $n \begin{bmatrix} n \\ k \end{bmatrix}$  ou  $k \left\{ \begin{matrix} n \\ k \end{matrix} \right\}$ .

Toute permutation est équivalente à un ensemble de cycles. Considérons par exemple la permutation qui transforme 123456789 en 384729156.

Une bonne façon de la représenter consiste à l'écrire sur deux lignes,

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 8 & 4 & 7 & 2 & 9 & 1 & 5 & 6, \end{array}$$

ce qui permet de bien voir que 1 devient 3, 2 devient 8 etc. La structure en cycles apparaît lorsqu'on dit que 1 devient 3, qui devient 4, qui devient 7, qui devient l'élément d'origine 1. Cela forme le cycle [1, 3, 4, 7]. Il y a deux autres cycles dans cette permutation, [2, 8, 5] et [6, 9]. Par conséquent, la permutation 384729156 est équivalente à la répartition en cycles suivante :

$$[1, 3, 4, 7] [2, 8, 5] [6, 9].$$

Tout élément de toute permutation  $\pi_1\pi_2\dots\pi_n$  de  $\{1, 2, \dots, n\}$  se trouve dans un unique cycle. En effet, si on part de  $m_0 = m$  et si on calcule  $m_1 = \pi_{m_0}$ ,  $m_2 = \pi_{m_1}$  etc, on finit forcément par revenir à  $m_k = m_0$  (on doit arriver tôt ou tard à un nombre déjà vu, et ce ne peut être que  $m_0$  car on connaît déjà l'unique prédécesseur de chacun des autres nombres  $m_1, m_2, \dots, m_{k-1}$ ). Donc, toute permutation définit un ensemble de cycles. Réciproquement, il suffit d'inverser la construction pour vérifier que tout ensemble de cycles définit une permutation. Nous pouvons donc en déduire que permutations et arrangements de cycles sont en fait les mêmes objets.

Par conséquent,  $\begin{bmatrix} n \\ k \end{bmatrix}$  représente le nombre de permutations de  $n$  objets qui contiennent exactement  $k$  cycles. Si on somme  $\begin{bmatrix} n \\ k \end{bmatrix}$  sur tout  $k$ , on obtient donc le nombre total de permutations :

$$\sum_{k=0}^n \begin{bmatrix} n \\ k \end{bmatrix} = n!, \quad n \geq 0 \text{ entier.} \tag{6.9}$$

Par exemple,  $6 + 11 + 6 + 1 = 24 = 4!$ .

Il s'avère que les relations de récurrence (6.3) et (6.8) apparaissent dans beaucoup de problèmes ; c'est pour cela que les nombres de Stirling sont utiles. Par exemple, si on veut représenter les puissances ordinaires  $x^n$  par des sommes de puissances descendantes  $x^n$  voici ce qu'on trouve pour premières valeurs :

$$\begin{aligned} x^0 &= x^0; \\ x^1 &= x^1; \\ x^2 &= x^2 + x^1; \\ x^3 &= x^3 + 3x^2 + x^1; \\ x^4 &= x^4 + 6x^3 + 7x^2 + x^1. \end{aligned}$$

Ces coefficients ressemblent comme deux gouttes d'eau aux nombres de la table 274, lus de droite à gauche. On peut donc raisonnablement penser

que la formule générale doit être

Il aurait mieux valu définir

$$\{n\} = [n] = 0$$

lorsque  $k < 0$  et  $n \geq 0$ .

$$x^n = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} x^k, \quad n \geq 0 \text{ entier.} \quad (6.10)$$

Il suffit d'une simple preuve par induction pour le confirmer. On a  $x \cdot x^k = x^{k+1} + kx^k$ , car  $x^{k+1} = x^k(x - k)$ ; donc  $x \cdot x^{n-1}$  est égal à

$$\begin{aligned} x \sum_k \begin{Bmatrix} n-1 \\ k \end{Bmatrix} x^k &= \sum_k \begin{Bmatrix} n-1 \\ k \end{Bmatrix} x^{k+1} + \sum_k \begin{Bmatrix} n-1 \\ k \end{Bmatrix} kx^k \\ &= \sum_{k-1} \begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix} x^k + \sum_k \begin{Bmatrix} n-1 \\ k \end{Bmatrix} kx^k \\ &= \sum_k \left( k \begin{Bmatrix} n-1 \\ k \end{Bmatrix} + \begin{Bmatrix} n-1 \\ k-1 \end{Bmatrix} \right) x^k = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} x^k. \end{aligned}$$

En d'autres termes, les nombres de Stirling de deuxième espèce sont les coefficients à appliquer aux puissances factorielles pour obtenir des puissances ordinaires.

On peut aussi faire l'opération inverse, et montrer que les nombres de Stirling de première espèce sont les coefficients à appliquer aux puissances ordinaires pour obtenir des puissances factorielles :

$$\begin{aligned} x^{\bar{0}} &= x^0; \\ x^{\bar{1}} &= x^1; \\ x^{\bar{2}} &= x^2 + x^1; \\ x^{\bar{3}} &= x^3 + 3x^2 + 2x^1; \\ x^{\bar{4}} &= x^4 + 6x^3 + 11x^2 + 6x^1. \end{aligned}$$

Comme on a  $(x + n - 1) \cdot x^k = x^{k+1} + (n - 1)x^k$ , une preuve similaire à la précédente aboutit à

$$(x + n - 1)x^{\overline{n-1}} = (x + n - 1) \sum_k \begin{bmatrix} n-1 \\ k \end{bmatrix} x^k = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} x^k.$$

Nous avons donc une preuve par induction de la formule générale

$$x^{\bar{n}} = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} x^k, \quad n \geq 0 \text{ entier.} \quad (6.11)$$

Remarquez que si on pose  $x = 1$ , on retrouve (6.9).

Un instant s'il vous plaît ! Cette équation concerne les puissances factorielles montantes  $x^{\bar{n}}$ , tandis que (6.10) concerne les puissances descendantes  $x^{\underline{n}}$ . Que se passe-t-il si on veut exprimer  $x^{\underline{n}}$  avec des puissances

**Table 280** Nombres de Stirling : identités de base, pour  $n \geq 0$  entier.

Réurrences :

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = k \left\{ \begin{matrix} n-1 \\ k \end{matrix} \right\} + \left\{ \begin{matrix} n-1 \\ k-1 \end{matrix} \right\}.$$

$$\left[ \begin{matrix} n \\ k \end{matrix} \right] = (n-1) \left[ \begin{matrix} n-1 \\ k \end{matrix} \right] + \left[ \begin{matrix} n-1 \\ k-1 \end{matrix} \right].$$

Valeurs particulières :

$$\left\{ \begin{matrix} n \\ 0 \end{matrix} \right\} = \left[ \begin{matrix} n \\ 0 \end{matrix} \right] = [n=0].$$

$$\left\{ \begin{matrix} n \\ 1 \end{matrix} \right\} = [n>0]; \quad \left[ \begin{matrix} n \\ 1 \end{matrix} \right] = (n-1)! [n>0].$$

$$\left\{ \begin{matrix} n \\ 2 \end{matrix} \right\} = (2^{n-1} - 1)[n>0]; \quad \left[ \begin{matrix} n \\ 2 \end{matrix} \right] = (n-1)! H_{n-1} [n>0].$$

$$\left\{ \begin{matrix} n \\ n-1 \end{matrix} \right\} = \left[ \begin{matrix} n \\ n-1 \end{matrix} \right] = \binom{n}{2}.$$

$$\left\{ \begin{matrix} n \\ n \end{matrix} \right\} = \left[ \begin{matrix} n \\ n \end{matrix} \right] = \binom{n}{n} = 1.$$

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \left[ \begin{matrix} n \\ k \end{matrix} \right] = \binom{n}{k} = 0, \quad \text{if } k > n.$$

Conversions entre puissances :

$$x^n = \sum_k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} x^k = \sum_k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^{n-k} x^k.$$

$$x^{\frac{n}{k}} = \sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] (-1)^{n-k} x^k;$$

$$x^{\overline{n}} = \sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] x^k.$$

Formules d'inversion :

$$\sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] \left\{ \begin{matrix} k \\ m \end{matrix} \right\} (-1)^{n-k} = [m=n];$$

$$\sum_k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \left[ \begin{matrix} k \\ m \end{matrix} \right] (-1)^{n-k} = [m=n].$$

**Table 281** Nombres de Stirling : autres identités, pour  $l, m, n \geq 0$  entiers.

$$\left\{ \begin{matrix} n+1 \\ m+1 \end{matrix} \right\} = \sum_k \binom{n}{k} \left\{ \begin{matrix} k \\ m \end{matrix} \right\}. \quad (6.12)$$

$$\left[ \begin{matrix} n+1 \\ m+1 \end{matrix} \right] = \sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] \left( \begin{matrix} k \\ m \end{matrix} \right). \quad (6.13)$$

$$\left\{ \begin{matrix} n \\ m \end{matrix} \right\} = \sum_k \binom{n}{k} \left\{ \begin{matrix} k+1 \\ m+1 \end{matrix} \right\} (-1)^{n-k}. \quad (6.14)$$

$$\begin{aligned} n^m (-1)^{n-m} \left[ \begin{matrix} n \\ m \end{matrix} \right] \\ = \sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] \left( \begin{matrix} -m \\ k-m \end{matrix} \right) n^k. \end{aligned} \quad (6.15)$$

$$m! \left\{ \begin{matrix} n \\ m \end{matrix} \right\} = \sum_k \binom{m}{k} k^n (-1)^{m-k}. \quad (6.16)$$

$$\left\{ \begin{matrix} n+1 \\ m+1 \end{matrix} \right\} = \sum_{k=0}^n \left\{ \begin{matrix} k \\ m \end{matrix} \right\} (m+1)^{n-k}. \quad (6.17)$$

$$\left[ \begin{matrix} n+1 \\ m+1 \end{matrix} \right] = \sum_{k=0}^n \left[ \begin{matrix} k \\ m \end{matrix} \right] n^{\underline{n-k}} = n! \sum_{k=0}^n \left[ \begin{matrix} k \\ m \end{matrix} \right] / k!. \quad (6.18)$$

$$\left\{ \begin{matrix} m+n+1 \\ m \end{matrix} \right\} = \sum_{k=0}^m k \left\{ \begin{matrix} n+k \\ k \end{matrix} \right\}. \quad (6.19)$$

$$\left[ \begin{matrix} m+n+1 \\ m \end{matrix} \right] = \sum_{k=0}^m (n+k) \left[ \begin{matrix} n+k \\ k \end{matrix} \right]. \quad (6.20)$$

$$\binom{n}{m} = \sum_k \left\{ \begin{matrix} n+1 \\ k+1 \end{matrix} \right\} \left[ \begin{matrix} k \\ m \end{matrix} \right] (-1)^{m-k}. \quad (6.21)$$

$$On a aussi \quad n^{\underline{n-m}} [n \geq m] = \sum_k \left[ \begin{matrix} n+1 \\ k+1 \end{matrix} \right] \left\{ \begin{matrix} k \\ m \end{matrix} \right\} (-1)^{m-k}. \quad (6.22)$$

$$\left\{ \begin{matrix} n \\ n-m \end{matrix} \right\} = \sum_k \binom{m-n}{m+k} \binom{m+n}{n+k} \left[ \begin{matrix} m+k \\ k \end{matrix} \right]. \quad (6.23)$$

$$\left[ \begin{matrix} n \\ n-m \end{matrix} \right] = \sum_k \binom{m-n}{m+k} \binom{m+n}{n+k} \left\{ \begin{matrix} m+k \\ k \end{matrix} \right\}. \quad (6.24)$$

$$\left\{ \begin{matrix} n \\ l+m \end{matrix} \right\} \binom{l+m}{l} = \sum_k \left\{ \begin{matrix} k \\ l \end{matrix} \right\} \left\{ \begin{matrix} n-k \\ m \end{matrix} \right\} \binom{n}{k}. \quad (6.25)$$

$$\left[ \begin{matrix} n \\ l+m \end{matrix} \right] \binom{l+m}{l} = \sum_k \left[ \begin{matrix} k \\ l \end{matrix} \right] \left[ \begin{matrix} n-k \\ m \end{matrix} \right] \binom{n}{k}. \quad (6.26)$$

On a aussi

$$\binom{n}{m} (n-1)^{\underline{n-m}} = \sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] \left\{ \begin{matrix} k \\ m \end{matrix} \right\},$$

une généralisation  
de (6.9).

## 282 NOMBRES REMARQUABLES

ordinaires ou  $x^n$  avec des puissances montantes ? Facile, on n'a qu'à ajouter quelques signes moins :

$$x^n = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^{n-k} x^k, \quad n \geq 0 \text{ entier}; \quad (6.27)$$

$$x^{\bar{n}} = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^{n-k} x^k, \quad n \geq 0 \text{ entier}. \quad (6.28)$$

Ça marche parce que, par exemple, la formule

$$x^4 = x(x-1)(x-2)(x-3) = x^4 - 6x^3 + 11x^2 - 6x$$

est identique à la formule

$$x^{\bar{4}} = x(x+1)(x+2)(x+3) = x^4 + 6x^3 + 11x^2 + 6x$$

à ceci près que les signes alternent. L'identité générale

$$x^{\bar{n}} = (-1)^n (-x)^{\bar{n}} \quad (6.29)$$

de l'exercice 2.17 permet de passer de (6.10) à (6.27) et de (6.11) à (6.28) en modifiant le signe de  $x$ .

Pour savoir où mettre le facteur  $(-1)^{n-k}$  dans les formules du genre de (6.27), il suffit d'avoir en tête l'ordre naturel des puissances :

$$x^{\bar{n}} > x^n > x^{\bar{m}}, \quad \text{pour tout } x > n > 1. \quad (6.30)$$

Comme les nombres de Stirling  $\begin{Bmatrix} n \\ k \end{Bmatrix}$  et  $\begin{Bmatrix} n \\ k \end{Bmatrix}$  sont positifs ou nuls, il faut mettre des signes moins lorsqu'on exprime une "petite" puissance en fonction d'une plus "grande".

En combinant (6.11) et (6.27), on obtient une somme double :

$$x^n = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^{n-k} x^k = \sum_{k,m} \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{Bmatrix} k \\ m \end{Bmatrix} (-1)^{n-k} x^m.$$

Puisque cela est vrai pour tout  $x$ , les coefficients de  $x^0, x^1, \dots, x^{n-1}, x^{n+1}, x^{n+2}, \dots$  du membre droit sont nécessairement tous nuls, et donc

$$\sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{Bmatrix} k \\ m \end{Bmatrix} (-1)^{n-k} = [m=n], \quad m, n \geq 0 \text{ entiers}. \quad (6.31)$$

Les nombres de Stirling, tout comme les coefficients binomiaux, satisfont un grand nombre de surprenantes identités. Cependant, les applications de ces identités sont moins variées que celles concernant les binomiaux que nous avons vues au chapitre 5. C'est pourquoi nous nous contenterons

de noter les plus simples d'entre elles, pour pouvoir nous y référer en cas de besoin. Les tables 280 et 281 présentent les formules les plus utilisées, y compris celles que nous avons déjà calculées.

Au chapitre 5, nous avons trouvé intéressant d'étendre les coefficients binomiaux  $\binom{n}{k}$  aux entiers  $n$  négatifs, de sorte que l'identité  $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$  soit valide sans aucune restriction. Cela a donné lieu à la table 175, qui nous a permis de découvrir que le triangle de Pascal s'auto-reproduit, à une rotation près, lorsqu'on l'étend vers le haut. Essayons donc de faire de même avec les triangles de Stirling : que se passe-t-il si on décide que les récurrences de base

$$\begin{aligned}\left\{\begin{matrix} n \\ k \end{matrix}\right\} &= k \left\{\begin{matrix} n-1 \\ k \end{matrix}\right\} + \left\{\begin{matrix} n-1 \\ k-1 \end{matrix}\right\} \\ \left[\begin{matrix} n \\ k \end{matrix}\right] &= (n-1) \left[\begin{matrix} n-1 \\ k \end{matrix}\right] + \left[\begin{matrix} n-1 \\ k-1 \end{matrix}\right]\end{aligned}$$

deviennent valables pour tous entiers  $n$  et  $k$  ? La solution est unique à condition qu'on pose deux conditions tout à fait raisonnables :

$$\left\{\begin{matrix} 0 \\ k \end{matrix}\right\} = \left[\begin{matrix} 0 \\ k \end{matrix}\right] = [k=0] \quad \text{et} \quad \left\{\begin{matrix} n \\ 0 \end{matrix}\right\} = \left[\begin{matrix} n \\ 0 \end{matrix}\right] = [n=0]. \quad (6.32)$$

Le résultat que l'on obtient est tout à fait surprenant : le triangle de Stirling pour les cycles apparaît au dessus du triangle de Stirling pour les sous-ensembles, et vice versa ! Les deux espèces de nombres de Stirling sont en fait liées par une formule extrêmement simple [220, 221] :

$$\left[\begin{matrix} n \\ k \end{matrix}\right] = \left\{\begin{matrix} -k \\ -n \end{matrix}\right\}, \quad k, n \text{ entiers.} \quad (6.33)$$

C'est une relation de "dualité", un peu comme entre min et max,  $[x]$  et  $\lceil x \rceil$ ,  $x^n$  et  $x^{\bar{n}}$ , ou encore pgcd et ppcm. Au vu de cette correspondance, les deux récurrences  $\left[\begin{matrix} n \\ k \end{matrix}\right] = (n-1) \left[\begin{matrix} n-1 \\ k \end{matrix}\right] + \left[\begin{matrix} n-1 \\ k-1 \end{matrix}\right]$  and  $\left\{\begin{matrix} n \\ k \end{matrix}\right\} = k \left\{\begin{matrix} n-1 \\ k \end{matrix}\right\} + \left\{\begin{matrix} n-1 \\ k-1 \end{matrix}\right\}$  comptent en fait exactement la même chose.

## 6.2 NOMBRES EULÉRIENS

Nous allons maintenant voir un autre triangle que l'on est souvent amené à rencontrer. Celui-ci est dû à Euler [104, §13; 110, page 485], et ses éléments sont notés  $\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \rangle$ . Les crochets de cette expression sont sensés suggérer les signes "supérieur à" et "inférieur à", car  $\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \rangle$  représente le nombre de permutations  $\pi_1 \pi_2 \dots \pi_n$  de  $\{1, 2, \dots, n\}$  qui possèdent  $k$  montées, c'est-à-dire  $k$  endroits où  $\pi_j < \pi_{j+1}$ . Bien que cette notation soit moins standard que les notations  $\left[\begin{matrix} n \\ k \end{matrix}\right]$ ,  $\left\{\begin{matrix} n \\ k \end{matrix}\right\}$  des nombres de Stirling, nous verrons qu'elle est bien justifiée.

(Knuth [209, première édition] écrivait  $\langle \begin{smallmatrix} n \\ k+1 \end{smallmatrix} \rangle$  pour  $\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \rangle$ ).

Table 284 Le tandem des triangles de Stirling.

$n$	$\binom{n}{-5}$	$\binom{n}{-4}$	$\binom{n}{-3}$	$\binom{n}{-2}$	$\binom{n}{-1}$	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$
-5	1										
-4	10	1									
-3	35	6	1								
-2	50	11	3	1							
-1	24	6	2	1	1						
0	0	0	0	0	0	1					
1	0	0	0	0	0	0	1				
2	0	0	0	0	0	0	1	1			
3	0	0	0	0	0	0	1	3	1		
4	0	0	0	0	0	0	1	7	6	1	
5	0	0	0	0	0	0	1	15	25	10	1

Il existe par exemple onze permutations de  $\{1, 2, 3, 4\}$  qui ont trois montées :

$$\begin{aligned} & 1324, \quad 1423, \quad 2314, \quad 2413, \quad 3412; \\ & 1243, \quad 1342, \quad 2341; \quad 2134, \quad 3124, \quad 4123 \end{aligned}$$

(la première ligne contient les permutations telles que  $\pi_1 < \pi_2 > \pi_3 < \pi_4$ , tandis que la seconde présente celles telles que  $\pi_1 < \pi_2 < \pi_3 > \pi_4$  ou  $\pi_1 > \pi_2 < \pi_3 < \pi_4$ ). Donc  $\binom{4}{2} = 11$ . On peut voir dans la table 285 les premiers nombres eulériens. La suite 1, 11, 11, 1 est représentative des nombres eulériens, comme la suite 1, 7, 6, 1 est représentative des nombres de Stirling de deuxième espèce. Pour tout  $n > 0$ , il ne peut y avoir que  $n - 1$  montées au plus ; c'est pourquoi on a  $\binom{n}{n} = [n=0]$  sur la diagonale du triangle.

Les lignes du triangle d'Euler, comme celles du triangle de Pascal, sont symétriques. Toutefois, la loi de symétrie est légèrement différente dans le cas présent :

$$\binom{n}{k} = \binom{n}{n-1-k}, \quad n > 0 \text{ entier}; \quad (6.34)$$

La permutation  $\pi_1 \pi_2 \dots \pi_n$  a  $n - 1 - k$  montées si et seulement si son image-miroir  $\pi_n \dots \pi_2 \pi_1$  a  $k$  montées.

Essayons maintenant de trouver une récurrence pour  $\binom{n}{k}$ . Toute permutation  $\rho = \rho_1 \dots \rho_{n-1}$  de  $\{1, \dots, n-1\}$  donne lieu à  $n$  permutations de  $\{1, 2, \dots, n\}$ , en insérant le nouvel élément  $n$  de toutes les façons possibles. Supposons que  $n$  soit mis en position  $j$ , obtenant ainsi la permutation  $\pi = \rho_1 \dots \rho_{j-1} n \rho_j \dots \rho_{n-1}$ . Si  $j = 1$  ou  $\rho_{j-1} < \rho_j$ , alors  $\pi$  a le même

**Table 285** Le triangle d'Euler.

$n$	$\langle n \rangle_0$	$\langle n \rangle_1$	$\langle n \rangle_2$	$\langle n \rangle_3$	$\langle n \rangle_4$	$\langle n \rangle_5$	$\langle n \rangle_6$	$\langle n \rangle_7$	$\langle n \rangle_8$	$\langle n \rangle_9$
0	1									
1	1	0								
2	1	1	0							
3	1	4	1	0						
4	1	11	11	1	0					
5	1	26	66	26	1	0				
6	1	57	302	302	57	1	0			
7	1	120	1191	2416	1191	120	1	0		
8	1	247	4293	15619	15619	4293	247	1	0	
9	1	502	14608	88234	156190	88234	14608	502	1	0

nombre de montées que  $\rho$ ; si  $\rho_{j-1} > \rho_j$  ou  $j = n$ , alors  $\pi$  a une montée de plus que  $\rho$ . Par conséquent,  $\pi$  possède  $k$  montées si  $\rho$  est choisie parmi  $(k+1)\langle n-1 \rangle_k$  permutations à  $k$  montées ou parmi  $((n-2)-(k-1)+1)\langle n-1 \rangle_{k-1}$  permutations à  $k-1$  montées. Voici donc la récurrence cherchée :

$$\langle n \rangle_k = (k+1)\langle n-1 \rangle_k + (n-k)\langle n-1 \rangle_{k-1}, \quad n > 0 \text{ entier.} \quad (6.35)$$

le départ de cette récurrence est donné par

$$\langle 0 \rangle_k = [k=0], \quad k \text{ entier,} \quad (6.36)$$

et nous supposerons que  $\langle n \rangle_k = 0$  lorsque  $k < 0$ .

Les nombres eulériens permettent d'exprimer une relation insolite entre les puissances ordinaires et les coefficients binomiaux :

$$x^n = \sum_k \langle n \rangle_k \binom{x+k}{n}, \quad n \geq 0 \text{ entier.} \quad (6.37)$$

C'est dans un important ouvrage chinois de Li Shan-Lan [249; 265, pages 320-325], publié en 1867, que la formule (6.37) apparaît pour la première fois.

Cette formule est appelée l'identité de Worlitzky [378]. par exemple, on a

$$x^2 = \binom{x}{2} + \binom{x+1}{2},$$

$$x^3 = \binom{x}{3} + 4\binom{x+1}{3} + \binom{x+2}{3},$$

$$x^4 = \binom{x}{4} + 11\binom{x+1}{4} + 11\binom{x+2}{4} + \binom{x+3}{4},$$

et ainsi de suite. La formule (6.37) se prouve aisément par induction (exercice 14).

Remarquons que (6.37) nous fournit un nouveau moyen de calculer la somme des  $n$  premiers carrés :  $k^2 = \binom{2}{0} \binom{k}{2} + \binom{2}{1} \binom{k+1}{2} = \binom{k}{2} + \binom{k+1}{2}$ , donc

$$\begin{aligned} 1^2 + 2^2 + \cdots + n^2 &= \left( \binom{1}{2} + \binom{2}{2} + \cdots + \binom{n}{2} \right) + \left( \binom{2}{2} + \binom{3}{2} + \cdots + \binom{n+1}{2} \right) \\ &= \binom{n+1}{3} + \binom{n+2}{3} = \frac{1}{6}(n+1)n((n-1)+(n+2)). \end{aligned}$$

La récurrence eulérienne (6.35) est un peu plus compliquée que les récurrences de Stirling (6.3) et (6.8). Il n'est donc pas étonnant que les nombres  $\binom{n}{m}$  satisfassent moins d'identités simples que les nombres de Stirling. On peut toutefois en trouver quelques-unes :

$$\binom{n}{m} = \sum_{k=0}^m \binom{n+1}{k} (m+1-k)^n (-1)^k; \quad (6.38)$$

$$m! \left\{ \binom{n}{m} \right\} = \sum_k \binom{n}{k} \binom{k}{n-m}; \quad (6.39)$$

$$\binom{n}{m} = \sum_k \left\{ \binom{n}{k} \right\} \binom{n-k}{m} (-1)^{n-k-m} k!. \quad (6.40)$$

En multipliant (6.39) par  $z^{n-m}$  puis en sommant sur  $m$ , on obtient l'égalité  $\sum_m \left\{ \binom{n}{m} \right\} m! z^{n-m} = \sum_k \binom{n}{k} (z+1)^k$ . Il suffit alors de remplacer  $z$  par  $z-1$  et de mettre les coefficients en équations pour retrouver (6.40). Les deux dernières identités sont donc équivalentes. La première, (6.38), donne des valeurs particulières pour des petits  $m$  :

$$\binom{n}{0} = 1; \quad \binom{n}{1} = 2^n - n - 1; \quad \binom{n}{2} = 3^n - (n+1)2^n + \binom{n+1}{2}.$$

Inutile de nous étendre davantage sur les nombres eulériens. Nous savons qu'ils existent, et nous disposons d'un certain nombre d'identités pour les manipuler lorsque nous les rencontrons. Avant de changer totalement de sujet, nous allons quand même jeter un coup d'œil à la table 287, qui présente ce qu'on appelle les "nombres eulériens du second ordre"  $\langle\langle \binom{n}{k} \rangle\rangle$ . Ils sont ainsi nommés car ils satisfont une récurrence similaire à (6.35), mais dans laquelle une des occurrences de  $n$  est remplacée par  $2n-1$  :

$$\langle\langle \binom{n}{k} \rangle\rangle = (k+1) \langle\langle \binom{n-1}{k} \rangle\rangle + (2n-1-k) \langle\langle \binom{n-1}{k-1} \rangle\rangle. \quad (6.41)$$

Ces nombres ont une curieuse interprétation combinatoire, due à Gessel et Stanley [147]. Considérons les permutations du multi-ensemble  $\{1, 1, 2, 2\}$ ,

**Table 287** Le triangle eulérien du second ordre.

$n$	$\langle\langle n \rangle\rangle_0$	$\langle\langle n \rangle\rangle_1$	$\langle\langle n \rangle\rangle_2$	$\langle\langle n \rangle\rangle_3$	$\langle\langle n \rangle\rangle_4$	$\langle\langle n \rangle\rangle_5$	$\langle\langle n \rangle\rangle_6$	$\langle\langle n \rangle\rangle_7$	$\langle\langle n \rangle\rangle_8$
0	1								
1	1	0							
2	1	2	0						
3	1	8	6	0					
4	1	22	58	24	0				
5	1	52	328	444	120	0			
6	1	114	1452	4400	3708	720	0		
7	1	240	5610	32120	58140	33984	5040	0	
8	1	494	19950	195800	644020	785304	341136	40320	0

$\dots, n, n\}$  telles que, pour tout  $1 \leq m \leq n$ , tous les nombres situés entre deux occurrences de  $m$  sont strictement supérieurs à  $m$ . Alors  $\langle\langle n \rangle\rangle_k$  compte le nombre de ces permutations qui ont  $k$  montées. Par exemple, il existe huit telles permutations de  $\{1, 1, 2, 2, 3, 3\}$  ayant une seule montée :

$$113322, 133221, 221331, 221133, 223311, 233211, 331122, 331221.$$

Ainsi,  $\langle\langle 3 \rangle\rangle_1 = 8$ . Le nombre total de permutations du multi-ensemble  $\{1, 1, 2, 2, \dots, n, n\}$  qui satisfont la condition donnée ci-dessus est

$$\sum_k \langle\langle n \rangle\rangle_k = (2n-1)(2n-3)\dots(1) = \frac{(2n)^n}{2^n}. \quad (6.42)$$

En effet, les deux occurrences de  $n$  doivent être adjacentes et il y a  $2n-1$  façons de les insérer dans une permutation de  $\{1, 1, 2, 2, \dots, n-1, n-1\}$ . Par exemple, pour  $n=3$ , la permutation 1221 donne naissance à 331221, 133221, 123321, 122331 et 122133. La récurrence (6.41) se démontre en étendant l'argument que nous avons utilisé pour les nombres eulériens ordinaires.

Les nombres eulériens du second ordre sont importants principalement en raison de leur lien avec les nombres de Stirling [148]: on peut prouver par induction sur  $n$  que

$$\left\{ \begin{matrix} x \\ x-n \end{matrix} \right\} = \sum_k \langle\langle n \rangle\rangle \binom{x+n-1-k}{2n}, \quad n \geq 0 \text{ entier}; \quad (6.43)$$

$$\left[ \begin{matrix} x \\ x-n \end{matrix} \right] = \sum_k \langle\langle n \rangle\rangle \binom{x+k}{2n}, \quad n \geq 0 \text{ entier}. \quad (6.44)$$

Par exemple,

$$\begin{aligned}\left\{\begin{array}{l}x \\ x-1\end{array}\right\} &= \binom{x}{2}, & \left[\begin{array}{l}x \\ x-1\end{array}\right] &= \binom{x}{2}; \\ \left\{\begin{array}{l}x \\ x-2\end{array}\right\} &= \binom{x+1}{4} + 2\binom{x}{4}, & \left[\begin{array}{l}x \\ x-2\end{array}\right] &= \binom{x}{4} + 2\binom{x+1}{4}; \\ \left\{\begin{array}{l}x \\ x-3\end{array}\right\} &= \binom{x+2}{6} + 8\binom{x+1}{6} + 6\binom{x}{6}, & \left[\begin{array}{l}x \\ x-3\end{array}\right] &= \binom{x}{6} + 8\binom{x+1}{6} + 6\binom{x+2}{6}.\end{aligned}$$

(Nous avons déjà vu le cas  $n = 1$  en (6.7)). Ces identités sont vraies pour tout entier  $x$  et tout entier positif ou nul  $n$ . Du fait que les membres droits sont des polynômes en  $x$ , les formules (6.43) et (6.44) permettent de définir le nombre de Stirling  $\left\{\begin{array}{l}x \\ x-n\end{array}\right\}$  et  $\left[\begin{array}{l}x \\ x-n\end{array}\right]$  pour tout  $x$  réel ou même complexe.

Pour tout  $n > 0$ , les polynômes  $\left\{\begin{array}{l}x \\ x-n\end{array}\right\}$  et  $\left[\begin{array}{l}x \\ x-n\end{array}\right]$  sont nuls si  $x = 0, x = 1, \dots$ , ou  $x = n$ . Ils sont donc divisibles par  $(x-0), (x-1), \dots$  et  $(x-n)$ . Essayons de voir ce qui reste de ces polynômes une fois qu'on les a divisés par tous ces facteurs. Soient les *polynômes de Stirling*  $\sigma_n(x)$  définis par la formule

$$\sigma_n(x) = \left[\begin{array}{l}x \\ x-n\end{array}\right] / (x(x-1)\dots(x-n)). \quad (6.45)$$

Le polynôme  $\sigma_n(x)$  est donc de degré  $n - 1$ . Voici les premiers polynômes de Stirling :

$$\begin{aligned}\sigma_0(x) &= 1/x; \\ \sigma_1(x) &= 1/2; \\ \sigma_2(x) &= (3x-1)/24; \\ \sigma_3(x) &= (x^2-x)/48; \\ \sigma_4(x) &= (15x^3-30x^2+5x+2)/5760.\end{aligned}$$

*1/x est donc un polynôme ?*  
*(Désolés).*

On peut aussi les calculer à l'aide des nombres eulériens du second ordre. Par exemple,  $\sigma_3(x) = ((x-4)(x-5) + 8(x-4)(x+1) + 6(x+2)(x+1))/6!$ .

Ces polynômes satisfont deux jolies identités :

$$\left(\frac{ze^z}{e^z - 1}\right)^x = x \sum_n \sigma_n(x) z^n; \quad (6.46)$$

$$\left(\frac{1}{z} \ln \frac{1}{1-z}\right)^x = x \sum_n \sigma_n(x+n) z^n. \quad (6.47)$$

**Table 289** Convolutions de Stirling.

$$rs \sum_{k=0}^n \sigma_k(r+tk) \sigma_{n-k}(s+t(n-k)) = (r+s)\sigma_n(r+s+tn) \quad (6.48)$$

$$s \sum_{k=0}^n k\sigma_k(r+tk) \sigma_{n-k}(s+t(n-k)) = n\sigma_n(r+s+tn) \quad (6.49)$$

$$\left\{ \begin{matrix} n \\ m \end{matrix} \right\} = (-1)^{n-m+1} \frac{n!}{(m-1)!} \sigma_{n-m}(-m) \quad (6.50)$$

$$\left[ \begin{matrix} n \\ m \end{matrix} \right] = \frac{n!}{(m-1)!} \sigma_{n-m}(n) \quad (6.51)$$

Plus généralement, si  $\mathcal{S}_t(z)$  est une série qui satisfait

$$\ln(1 - z\mathcal{S}_t(z)^{t-1}) = -z\mathcal{S}_t(z)^t, \quad (6.52)$$

alors

$$\mathcal{S}_t(z)^x = x \sum_n \sigma_n(x+tn) z^n. \quad (6.53)$$

On peut ainsi obtenir les formules de convolutions pour les nombres de Stirling qui sont présentées dans la table 289, comme celles que nous avions pour les coefficients binomiaux en table 216. Lorsque les identités des tables 280 et 281 ne suffisent pas à venir à bout d'une somme, celles de la table 289 peuvent avantageusement venir à la rescousse (nous en verrons un exemple un peu plus loin, après l'équation (6.100)). L'exercice 7.19 porte sur les principes généraux des convolutions basées sur les identités du genre de (6.46) et (6.53).

### 6.3 NOMBRES HARMONIQUES

Il est temps maintenant de regarder de plus près les nombres harmoniques que nous avons rencontrés au chapitre 2 :

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k}, \quad n \geq 0 \text{ entier.} \quad (6.54)$$

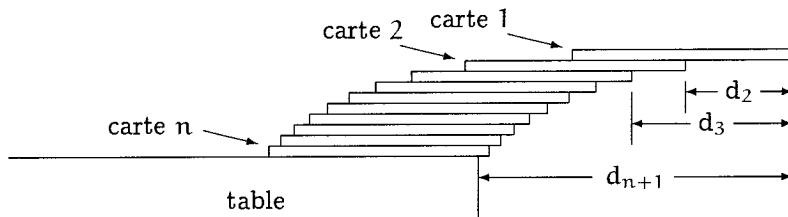
C'est parce qu'ils apparaissent très souvent lors de l'analyse d'algorithmes qu'on leur a accordé une notation spécifique. On les note donc  $H_n$ , où le "H" signifie "harmonique" en référence au monde musical : une note de longueur d'onde  $1/n$  est appelée une harmonique d'une note de longueur

d'où 1. Voici les premiers nombres harmoniques :

n	0	1	2	3	4	5	6	7	8	9	10
$H_n$	0	1	$\frac{3}{2}$	$\frac{11}{6}$	$\frac{25}{12}$	$\frac{137}{60}$	$\frac{49}{20}$	$\frac{363}{140}$	$\frac{761}{280}$	$\frac{7129}{2520}$	$\frac{7381}{2520}$

On montre dans l'exercice 21 que  $H_n$  n'est jamais un entier lorsque  $n > 1$ .

Voici maintenant un tour de cartes, imaginé par R. T. Sharp [325], qui montre comment les nombres harmoniques apparaissent naturellement dans des situations simples. Etant données  $n$  cartes et une table, on veut empiler les cartes au bord de la table de façon que la carte du haut soit la plus éloignée possible de ce bord, tout en respectant bien sûr la loi de la gravitation.



C'est la table 290.

Pour être tout à fait précis, ajoutons que les côtés des cartes doivent être parallèles aux côtés de la table. Enfin, pour que la réponse au problème soit simple, nous supposerons que la longueur de chaque carte est égale à 2.

Si on a une seule carte, on obtient le dépassement maximum en positionnant son centre de gravité juste à l'aplomb du bord de la table. Dans ce cas, puisque le centre de gravité est situé au milieu de la carte, le dépassement est égal à 1. Dans le cas de deux cartes, il n'est pas difficile de se convaincre qu'on obtient le dépassement maximum lorsque le centre de gravité de la carte du haut est juste au-dessus du bord de l'autre carte, et le centre de gravité de l'ensemble des deux cartes est positionné exactement à l'aplomb du bord de la table. Comme ce centre de gravité commun est situé exactement au milieu de l'intersection des deux cartes, le dépassement vaut maintenant  $1 + 1/2 = 3/2$ .

Ces premières expériences nous suggèrent une méthode générale, selon laquelle on place les cartes de façon que le centre de gravité des  $k$  cartes du haut soit situé exactement au-dessus du bord de la  $k + 1$ ème (celle qui supporte les  $k$  précédentes). Pour exprimer cette condition en termes algébriques, notons  $d_k$  la distance entre le bord extérieur de la première carte (celle du haut) et le bord extérieur de la  $k$ ème carte à partir de celle du haut. Alors  $d_1 = 0$ , et nous voulons que  $d_{k+1}$  soit le centre de gravité des  $k$  premières cartes :

$$d_{k+1} = \frac{(d_1 + 1) + (d_2 + 1) + \cdots + (d_k + 1)}{k}, \text{ pour } 1 \leq k \leq n. \quad (6.55)$$

En effet, le centre de gravité de  $k$  objets de poids respectifs  $w_1, \dots, w_k$  et de centres de gravité respectifs placés en  $p_1, \dots, p_k$ , se trouve à la position  $(w_1p_1 + \dots + w_kp_k)/(w_1 + \dots + w_k)$ . Nous pouvons réécrire cette récurrence sous deux formes équivalentes :

$$\begin{aligned} kd_{k+1} &= k + d_1 + \dots + d_{k-1} + d_k, & k \geq 0; \\ (k-1)d_k &= k - 1 + d_1 + \dots + d_{k-1}, & k \geq 1. \end{aligned}$$

En faisant la différence de ces deux équations, on obtient

$$kd_{k+1} - (k-1)d_k = 1 + d_k, \quad k \geq 1;$$

par conséquent  $d_{k+1} = d_k + 1/k$ . La seconde carte sera décalée d'une demi-unité par rapport à la troisième, qui sera décalée d'un tiers par rapport à la quatrième, et ainsi de suite. On obtient la formule générale

$$d_{k+1} = H_k \tag{6.56}$$

par induction, et si on pose  $k = n$  on trouve que le tas de  $n$  cartes dépasse du bord de la table de  $d_{n+1} = H_n$ .

Aurions-nous pu mieux faire en retenant un peu certaines cartes, en n'éloignant pas chacune d'entre elles le plus possible de la table, pour garder une "réserve d'énergie gravitationnelle" à récupérer plus tard ? Eh bien non : pour tout positionnement correct des cartes, on a

$$d_{k+1} \leq \frac{(1+d_1) + (1+d_2) + \dots + (1+d_k)}{k}, \quad 1 \leq k \leq n$$

et  $d_1 = 0$ . On en déduit, par induction, que  $d_{k+1} \leq H_k$ .

Remarquez qu'on n'a pas besoin de beaucoup de cartes pour que celle du dessus dépasse complètement le bord de la table. Cela correspond en effet à un dépassement de taille 2, et le premier nombre harmonique supérieur à 2 est  $H_4 = \frac{25}{12}$ . Quatre cartes suffisent donc.

Avec 52 cartes, on obtient un dépassement de  $H_{52}$  unités, soit  $H_{52}/2 \approx 2.27$  longueurs de cartes (nous apprendrons bientôt à approximer  $H_n$  sans avoir besoin d'additionner un grand nombre de fractions).

*Pour espérer faire ça avec 52 cartes, il faut être soit le roi des idiots, soit un as.*

Le problème du "ver sur l'élastique" met aussi en pratique les nombres harmoniques. Un ver  $V$ , lent mais têtu, part d'une extrémité d'un élastique d'un mètre de long et rampe à la vitesse d'un centimètre par minute vers l'autre bout. Cet élastique appartient à un propriétaire  $P$ , têtu lui aussi, et dont le seul plaisir dans la vie est d'enquiquiner  $V$ . Aussi, à la fin de chaque minute, il allonge l'élastique d'un mètre. Donc, après une minute de reptation,  $V$  se trouve à 1 centimètre du départ et 99 centimètres de l'arrivée ; puis  $P$  allonge le parcours d'un mètre. Pendant l'allongement,  $V$  garde sa position relative, à 1% du départ et 99% de l'arrivée ; donc  $V$  se

situe maintenant à 2 cm du départ et 198 cm du but. Une minute après, il a parcouru 3 cm et il lui en reste 197 à faire. A ce moment, P étire encore l'élastique et les distances deviennent respectivement 4,5 et 295,5. Et cela continue ainsi. Le ver finira-t-il par arriver au but ? A chaque fois qu'il approche du but, celui-ci semble s'éloigner. Nous supposons bien évidemment que P et V sont immortels, que l'élastique est infiniment étirable et que le ver est infiniment petit.

Réfléchissons un peu. Lorsque P étire l'élastique, la fraction de celui-ci que V a déjà parcourue reste la même. Donc il en parcourt 1/100ème la première minute, 1/200ème la deuxième, 1/300ème la troisième etc. Voici donc la fraction de l'élastique parcourue après  $n$  minutes :

$$\frac{1}{100} \left( \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right) = \frac{H_n}{100}. \quad (6.57)$$

Le ver arrivera donc au but s'il existe un entier  $n$  tel que  $H_n \geq 100$ .

Nous verrons bientôt comment estimer  $H_n$  lorsque  $n$  est grand. Pour l'instant, contentons-nous de vérifier notre analyse en regardant comment "Superver" s'en tirerait dans la même situation. Superver peut parcourir 50 cm par minute ; il aura donc parcouru  $H_n/2$  de l'élastique au bout de  $n$  minutes, selon l'argument donné plus haut. Si notre raisonnement est correct, Superver doit finir avant que  $n$  n'atteigne 4, car  $H_4 > 2$ . C'est en effet vrai : un calcul simple montre qu'il ne reste à Superver que  $33\frac{1}{3}$  cm à faire au bout de trois minutes. Il atteint donc le but en 3 minutes et 40 secondes exactement.

Les nombres harmoniques apparaissent aussi dans le triangle de Stirling. Voyons cela en essayant de trouver une forme close pour  $\begin{bmatrix} n \\ 2 \end{bmatrix}$ , le nombre de permutations de  $n$  objets contenant exactement deux cycles. Selon la récurrence (6.8),

$$\begin{aligned} \begin{bmatrix} n+1 \\ 2 \end{bmatrix} &= n \begin{bmatrix} n \\ 2 \end{bmatrix} + \begin{bmatrix} n \\ 1 \end{bmatrix} \\ &= n \begin{bmatrix} n \\ 2 \end{bmatrix} + (n-1)!, \quad \text{si } n > 0. \end{aligned}$$

Cette récurrence est un candidat idéal pour la technique du facteur de sommation du chapitre 2 :

$$\frac{1}{n!} \begin{bmatrix} n+1 \\ 2 \end{bmatrix} = \frac{1}{(n-1)!} \begin{bmatrix} n \\ 2 \end{bmatrix} + \frac{1}{n}.$$

Si on développe cette formule, on trouve que  $\frac{1}{n!} \begin{bmatrix} n+1 \\ 2 \end{bmatrix} = H_n$ , donc

$$\begin{bmatrix} n+1 \\ 2 \end{bmatrix} = n! H_n. \quad (6.58)$$

*Ce problème est bien plus scientifique que le précédent : on y utilise les unités métriques.*

*Ce n'est pas Superver pépère, alors.*

Nous avons démontré au chapitre 2 que la série harmonique  $\sum_k 1/k$  diverge, ce qui signifie que  $H_n$  n'est pas borné lorsque  $n \rightarrow \infty$ . Cependant, notre preuve était indirecte : elle était basée sur la somme (2.58), qui donne des résultats différents selon la façon dont on arrange ses termes. Le fait que  $H_n \rightarrow \infty$  choque l'intuition, car il implique notamment qu'avec assez de cartes on peut faire un tas qui dépasse le bord de la table d'un kilomètre ou plus, et que le ver V arrivera à coup sûr au bout de l'élastique. Regardons donc d'un peu plus près ce que donne  $H_n$  lorsque  $n$  est très grand.

Pour voir que  $H_n \rightarrow \infty$ , la chose la plus simple à faire est probablement de grouper ses termes par puissances de 2 : on met un terme dans le premier groupe, deux dans le deuxième, quatre dans le troisième, huit dans le quatrième et ainsi de suite :

$$\underbrace{\frac{1}{1}}_{\text{groupe 1}} + \underbrace{\frac{1}{2} + \frac{1}{3}}_{\text{groupe 2}} + \underbrace{\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}}_{\text{groupe 3}} + \underbrace{\frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15}}_{\text{groupe 4}} + \dots$$

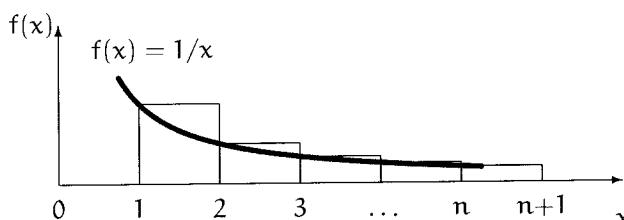
Les deux termes du deuxième groupe sont compris entre  $\frac{1}{4}$  et  $\frac{1}{2}$ , donc leur somme est comprise entre  $2 \cdot \frac{1}{4} = \frac{1}{2}$  et  $2 \cdot \frac{1}{2} = 1$ . Les quatre termes du troisième groupe sont entre  $\frac{1}{8}$  et  $\frac{1}{4}$ , donc leur somme est aussi comprise entre  $\frac{1}{2}$  et 1. Il s'avère que chacun des  $2^{k-1}$  termes du  $k$ ème groupe est entre  $2^{-k}$  et  $2^{1-k}$ , donc la somme de tous les termes de  $n$ 'importe quel groupe est entre  $\frac{1}{2}$  et 1.

Nous pouvons en déduire, par induction sur  $k$ , que si  $n$  est dans le  $k$ ème groupe, alors  $H_n > k/2$  et  $H_n \leq k$ . Par conséquent,  $H_n \rightarrow \infty$ , et de plus

$$\frac{\lfloor \lg n \rfloor + 1}{2} < H_n \leq \lfloor \lg n \rfloor + 1. \quad (6.59)$$

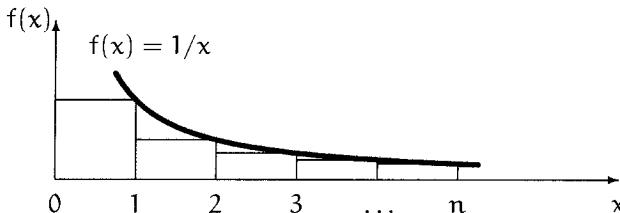
Nous connaissons maintenant  $H_n$  à un facteur 2 près. Les nombres harmoniques tendent bien vers l'infini, mais ils s'en approchent à vitesse logarithmique, autant dire très lentement.

Avec un tout petit peu plus de travail et une pincée de calcul infinitésimal, on peut trouver une meilleure approximation. Nous avons vu au chapitre 2 que  $H_n$  est l'analogie discret de la fraction continue  $\ln n$ . Le logarithme népérien étant défini comme l'aire sous une courbe, voyons ce que donne une comparaison géométrique :



*Ils sont si lents  
qu'on devrait les  
appeler les "nombres  
du ver".*

L'aire sous la courbe entre 1 et  $n$ , qui est égale à  $\int_1^n dx/x = \ln n$ , est plus petite que l'aire des  $n$  rectangles, qui vaut  $\sum_{k=1}^n 1/k = H_n$ . Par conséquent,  $\ln n < H_n$ . Ce minorant est meilleur que celui que nous avions en (6.59). Il suffit de placer les rectangles un peu différemment pour trouver un majorant similaire :



Cette fois, l'aire des  $n$  rectangles,  $H_n$ , est plus petite que la somme de celle du premier rectangle et de celle sous la courbe. Nous avons ainsi prouvé que

$$\ln n < H_n < \ln n + 1, \quad \text{pour } n > 1. \quad (6.60)$$

Nous connaissons maintenant la valeur de  $H_n$  à 1 près.

Lorsqu'on somme les carrés des inverses au lieu des inverses simplement, on trouve les nombres harmoniques du "second ordre"  $H_n^{(2)}$  :

$$H_n^{(2)} = 1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} = \sum_{k=1}^n \frac{1}{k^2}.$$

On définit de même les nombres harmoniques d'ordre  $r$  en sommant les puissances négatives de  $r$  :

$$H_n^{(r)} = \sum_{k=1}^n \frac{1}{k^r}. \quad (6.61)$$

Si  $r > 1$ , ces nombres tendent vers une limite finie lorsque  $n \rightarrow \infty$ . Nous avons vu dans l'exercice 2.31 que cette limite est traditionnellement appelée la fonction zêta de Riemann :

$$\zeta(r) = H_\infty^{(r)} = \sum_{k \geq 1} \frac{1}{k^r}. \quad (6.62)$$

On doit à Euler [103] un procédé plaisant pour approximer les nombres harmoniques ordinaires  $H_n^{(1)}$  par des nombres harmoniques généralisés. Soit la série infinie

$$\ln\left(\frac{k}{k-1}\right) = \frac{1}{k} + \frac{1}{2k^2} + \frac{1}{3k^3} + \frac{1}{4k^4} + \cdots, \quad (6.63)$$

*"I now see a way too how y<sup>e</sup> aggregate of y<sup>e</sup> termes of Musicall progressions may bee found (much after y<sup>e</sup> same manner) by Logarithms, but y<sup>e</sup> calculations for finding out those rules would bee still more troublesome."*

— I. Newton [280]

qui converge lorsque  $k > 1$ . Le membre gauche est égal à  $\ln k - \ln(k-1)$  ; donc, si on somme les deux membres pour  $2 \leq k \leq n$ , les membres gauches se télescopent pour donner

$$\begin{aligned}\ln n - \ln 1 &= \sum_{k=2}^n \left( \frac{1}{k} + \frac{1}{2k^2} + \frac{1}{3k^3} + \frac{1}{4k^4} + \dots \right) \\ &= (H_n - 1) + \frac{1}{2}(H_n^{(2)} - 1) + \frac{1}{3}(H_n^{(3)} - 1) + \frac{1}{4}(H_n^{(4)} - 1) + \dots\end{aligned}$$

En réarrangeant le termes, on obtient une expression de la différence entre  $H_n$  et  $\ln n$  :

$$H_n - \ln n = 1 - \frac{1}{2}(H_n^{(2)} - 1) - \frac{1}{3}(H_n^{(3)} - 1) - \frac{1}{4}(H_n^{(4)} - 1) - \dots.$$

Lorsque  $n \rightarrow \infty$ , le membre gauche tend vers la valeur limite

$$1 - \frac{1}{2}(\zeta(2) - 1) - \frac{1}{3}(\zeta(3) - 1) - \frac{1}{4}(\zeta(4) - 1) - \dots,$$

qui est connue sous le nom de *constante d'Euler* et traditionnellement désignée par la lettre grecque  $\gamma$ . Comme  $\zeta(r) - 1$  vaut à peu près  $1/2^r$ , cette série infinie converge assez vite, ce qui permet de calculer une bonne approximation de  $\gamma$  :

*"Huius igitur quantitatis constantis C valorem deteximus, quippe est C = 0,577218."*

—L. Euler [103]

$$\gamma = 0,5772156649\dots. \quad (6.64)$$

L'argument d'Euler permet d'établir que

$$\lim_{n \rightarrow \infty} (H_n - \ln n) = \gamma, \quad (6.65)$$

ce qui indique que  $H_n$  se situe à peu près à 58% de la distance qui sépare les deux bornes de (6.60). Nous cernons sa valeur de plus en plus près.

Nous verrons au chapitre 9 qu'on peut raffiner encore l'estimation. Nous montrerons notamment que

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{\epsilon_n}{120n^4}, \quad 0 < \epsilon_n < 1. \quad (6.66)$$

Grâce à cette formule, nous pouvons dire que le millionième nombre harmonique est

$$H_{1000000} \approx 14,3927267228657236313811275$$

sans pour cela additionner un million de fractions. Ce résultat implique, entre autres choses, qu'un tas d'un million de cartes peut dépasser le bord d'une table de plus de sept longueurs de carte.

Que nous dit (6.66) sur le problème du ver sur l'élastique ? Comme  $H_n$  n'est pas borné, le ver atteindra son but dès que  $H_n$  atteindra 100. Selon notre approximation de  $H_n$ , ceci arrivera pour  $n$  à peu près égal à

$$e^{100-\gamma} \approx e^{99.423}.$$

L'exercice 9.49 montre que la valeur critique de  $n$  est en fait soit  $\lfloor e^{100-\gamma} \rfloor$ , soit  $\lceil e^{100-\gamma} \rceil$ . Imaginons le triomphe de V en franchissant enfin la ligne d'arrivée, au grand dépit de P, après environ  $287 \times 10^{60}$  siècles de reptation. A ce moment, l'élastique aura été étiré de plus de  $10^{27}$  années-lumière et ses molécules seront assez clairsemées.

*En fait, le ver ne pourra pas ramper aussi longtemps, car la fin du monde sera arrivée bien plus tôt, dès que la Tour de Brahma aura été entièrement déplacée.*

## 6.4 SOMMATION HARMONIQUE

Nous allons examiner maintenant quelques sommes où apparaissent les nombres harmoniques. Commençons par appliquer quelques résultats que nous avons appris au chapitre 2. Nous avons prouvé en (2.36) et (2.57) que

$$\sum_{0 \leq k < n} H_k = nH_n - n; \quad (6.67)$$

$$\sum_{0 \leq k < n} kH_k = \frac{n(n-1)}{2}H_n - \frac{n(n-1)}{4}. \quad (6.68)$$

Attaquons-nous hardiment à une somme plus générale qui admet les deux précédentes comme cas particuliers : quelle est la valeur de

$$\sum_{0 \leq k < n} \binom{k}{m} H_k$$

lorsque  $m$  est un entier positif ou nul ?

La méthode qui a le mieux fonctionné pour (6.67) et (6.68) au chapitre 2 est la *sommation par parties*. Il s'agit d'écrire le terme général sous la forme  $u(k)\Delta v(k)$  et d'appliquer l'identité

$$\sum_a^b u(x)\Delta v(x) \delta x = u(x)v(x)|_a^b - \sum_a^b v(x+1)\Delta u(x) \delta x. \quad (6.69)$$

Cette méthode marche bien avec la somme à laquelle nous sommes confrontés maintenant,  $\sum_{0 \leq k < n} \binom{k}{m} H_k$ , car nous pouvons poser

$$u(k) = H_k, \quad \Delta u(k) = H_{k+1} - H_k = \frac{1}{k+1};$$

$$v(k) = \binom{k}{m+1}, \quad \Delta v(k) = \binom{k+1}{m+1} - \binom{k}{m+1} = \binom{k}{m}.$$

En transportant tout cela dans (6.69), on obtient

$$\begin{aligned}\sum_{0 \leq k < n} \binom{k}{m} H_k &= \sum_0^n \binom{x}{m} H_x \delta x \\ &= \left( \binom{x}{m+1} H_x \right) \Big|_0^n - \sum_0^n \binom{x+1}{m+1} \frac{\delta x}{x+1} \\ &= \binom{n}{m+1} H_n - \sum_{0 \leq k < n} \binom{k+1}{m+1} \frac{1}{k+1}.\end{aligned}$$

Pour évaluer la somme qui reste, il suffit de faire absorber le  $(k+1)^{-1}$  au moyen de notre bonne vieille équation (5.5) :

$$\sum_{0 \leq k < n} \binom{k+1}{m+1} \frac{1}{k+1} = \sum_{0 \leq k < n} \binom{k}{m} \frac{1}{m+1} = \binom{n}{m+1} \frac{1}{m+1}.$$

Voici donc la réponse à notre problème :

$$\sum_{0 \leq k < n} \binom{k}{m} H_k = \binom{n}{m+1} \left( H_n - \frac{1}{m+1} \right). \quad (6.70)$$

Remarquez que cela colle bien avec (6.67) et (6.68) lorsque  $m = 0$  et  $m = 1$  respectivement.

Voyons maintenant un exemple avec des divisions à la place des multiplications : essayons de calculer

$$S_n = \sum_{k=1}^n \frac{H_k}{k}.$$

Commençons par remplacer  $H_k$  par sa définition. Nous obtenons ainsi une somme double,

$$S_n = \sum_{1 \leq j \leq k \leq n} \frac{1}{j \cdot k}.$$

Il ne reste plus qu'à appeler à la rescousse une autre méthode du chapitre 2 : selon l'équation (2.33),

$$S_n = \frac{1}{2} \left( \left( \sum_{k=1}^n \frac{1}{k} \right)^2 + \sum_{k=1}^n \frac{1}{k^2} \right) = \frac{1}{2} (H_n^2 + H_n^{(2)}). \quad (6.71)$$

Nous aurions aussi pu obtenir le même résultat en sommant par parties (voir l'exercice 26).

Voici un problème plus difficile [354], qui est réfractaire à la sommation par parties :

$$u_n = \sum_{k \geq 1} \binom{n}{k} \frac{(-1)^{k-1}}{k} (n-k)^n, \quad n \geq 1 \text{ entier.}$$

On ne voit pas de nombres harmoniques dans cette somme, direz-vous ; mais qui sait ?

Nous allons résoudre ce problème de deux façons différentes : d'abord avec beaucoup d'efforts, sans faire appel à l'intuition ; puis par une approche qui doit une bonne part à l'intuition, pour ne pas dire à la chance. Voici donc pour commencer la méthode "ingrate". Développons  $(n-k)^n$  avec la formule du binôme, de façon que le  $k$  qui nous gène dans le dénominateur se combine avec le numérateur :

$$\begin{aligned} u_n &= \sum_{k \geq 1} \binom{n}{k} \frac{(-1)^{k-1}}{k} \sum_j \binom{n}{j} (-k)^j n^{n-j} \\ &= \sum_j \binom{n}{j} (-1)^{j-1} n^{n-j} \sum_{k \geq 1} \binom{n}{k} (-1)^k k^{j-1}. \end{aligned}$$

Malgré les apparences, ce n'est pas tout à fait la pagaille. En effet, le  $k^{j-1}$  de la somme interne est un polynôme en  $k$ , et l'identité (5.40) nous indique que, à quelques détails près, nous sommes simplement en présence de la  $j$ ème différence de ce polynôme. A quelques détails près, parce que, premièrement,  $k^{j-1}$  n'est pas un polynôme lorsque  $j = 0$ , donc il nous faut traiter le terme correspondant séparément ; deuxièmement, il nous manque le terme  $k = 0$  pour retrouver exactement la formule de la  $j$ ème différence, et comme ce terme est non nul lorsque  $j = 1$ , nous devons l'ajouter à la somme (et le soustraire en même temps à l'extérieur). Voici le résultat de ces opérations :

$$\begin{aligned} u_n &= \sum_{j \geq 1} \binom{n}{j} (-1)^{j-1} n^{n-j} \sum_{k \geq 0} \binom{n}{k} (-1)^k k^{j-1} \\ &\quad - \sum_{j \geq 1} \binom{n}{j} (-1)^{j-1} n^{n-j} \binom{n}{0} 0^{j-1} \\ &\quad - \binom{n}{0} n^n \sum_{k \geq 1} \binom{n}{k} (-1)^k k^{-1}. \end{aligned}$$

Parfait. Maintenant, la ligne du haut (la seule somme double qui reste) est nulle, car c'est une somme de multiples de différences  $j$ èmes de polynômes de degrés inférieurs à  $n$ . La seconde ligne est nulle sauf pour  $j = 1$ , auquel cas elle vaut  $-n^n$ . La seule difficulté réside maintenant dans la troisième

ligne. Nous avons réduit le problème original à une somme bien plus simple :

$$U_n = n^n(T_n - 1), \quad \text{où } T_n = \sum_{k \geq 1} \binom{n}{k} \frac{(-1)^{k-1}}{k}. \quad (6.72)$$

Par exemple,  $U_3 = \binom{3}{1} \frac{8}{1} - \binom{3}{2} \frac{1}{2} = \frac{45}{2}$ ;  $T_3 = \binom{3}{1} \frac{1}{1} - \binom{3}{2} \frac{1}{2} + \binom{3}{3} \frac{1}{3} = \frac{11}{6}$ ; donc  $U_3 = 27(T_3 - 1)$  comme annoncé.

Comment évaluer  $T_n$ ? On pourrait remplacer  $\binom{n}{k}$  par  $\binom{n-1}{k} + \binom{n-1}{k-1}$  pour obtenir une récurrence donnant  $T_n$  en fonction de  $T_{n-1}$ . Nous allons cependant procéder autrement, ce qui s'avèrera plus instructif : nous avons rencontré une formule similaire en (5.41), à savoir

$$\sum_k \binom{n}{k} \frac{(-1)^k}{x+k} = \frac{n!}{x(x+1)\dots(x+n)}.$$

Si on soustrait le terme pour lequel  $k=0$  et si on pose  $x=0$ , on obtient  $-T_n$ . Faisons le donc :

$$\begin{aligned} T_n &= \left( \frac{1}{x} - \frac{n!}{x(x+1)\dots(x+n)} \right) \Big|_{x=0} \\ &= \left( \frac{(x+1)\dots(x+n) - n!}{x(x+1)\dots(x+n)} \right) \Big|_{x=0} \\ &= \left( \frac{x^n \binom{n+1}{n+1} + \dots + x^{\binom{n+1}{2}} + \binom{n+1}{1} - n!}{x(x+1)\dots(x+n)} \right) \Big|_{x=0} = \frac{1}{n!} \binom{n+1}{2}. \end{aligned}$$

Dans ce calcul, nous avons utilisé le développement (6.11) de  $(x+1)\dots(x+n) = x^{\binom{n+1}{2}}/x$ , et nous avons pu diviser le numérateur par  $x$  car  $\binom{n+1}{1} = n!$ . Revenons à nos moutons : nous savons, grâce à (6.58), que  $\binom{n+1}{2} = n! H_n$ . Par conséquent,  $T_n = H_n$ , et voici la réponse à notre problème :

$$U_n = n^n(H_n - 1). \quad (6.73)$$

C'était la première approche. L'autre méthode consiste à essayer de calculer une somme bien plus générale,

$$U_n(x, y) = \sum_{k \geq 1} \binom{n}{k} \frac{(-1)^{k-1}}{k} (x + ky)^n, \quad n \geq 0 \text{ entier.} \quad (6.74)$$

On y retrouve le  $U_n$  original dans le cas particulier  $U_n(n, -1)$ .

Nous pourrions refaire, à peu de chose près, le calcul précédent et trouver la valeur de  $U_n(x, y)$ . Nous pourrions aussi remplacer  $(x + ky)^n$  par  $(x + ky)^{n-1}(x + ky)$  puis  $\binom{n}{k}$  par  $\binom{n-1}{k} + \binom{n-1}{k-1}$ , ce qui nous mènerait

à la récurrence

$$U_n(x, y) = xU_{n-1}(x, y) + x^n/n + yx^{n-1}, \quad (6.75)$$

que nous saurions résoudre sans problème avec un facteur de sommation (voir l'exercice 5).

Ce que nous allons faire est encore plus facile que cela. Nous allons utiliser quelque chose qui nous a déjà bien aidés au chapitre 2 : la dérivation. En dérivant  $U_n(x, y)$  par rapport à  $y$ , on amène un facteur  $k$  qui se simplifie avec celui du dénominateur, et la somme qui en résulte est triviale :

$$\begin{aligned} \frac{\partial}{\partial y} U_n(x, y) &= \sum_{k \geq 1} \binom{n}{k} (-1)^{k-1} n(x + ky)^{n-1} \\ &= \binom{n}{0} nx^{n-1} - \sum_{k \geq 0} \binom{n}{k} (-1)^k n(x + ky)^{n-1} \\ &= nx^{n-1}. \end{aligned}$$

C'est encore, la  $n$ ième différence d'un polynôme de degré  $< n$  qui simplifie bien les choses.

Nous avons démontré que la dérivée de  $U_n(x, y)$  par rapport à  $y$  est égale à  $nx^{n-1}$ , donc indépendante de  $y$ . De façon générale, si  $f'(y) = c$ , alors  $f(y) = f(0) + cy$  ; donc nécessairement

$$U_n(x, y) = U_n(x, 0) + nx^{n-1}y.$$

Il nous reste à déterminer  $U_n(x, 0)$ . Il suffit de remarquer que  $U_n(x, 0)$  est exactement égal à  $x^n$  fois la somme  $T_n = H_n$  de l'équation (6.72) pour conclure que la somme générale (6.74) peut s'écrire

$$U_n(x, y) = x^n H_n + nx^{n-1}y. \quad (6.76)$$

En particulier, la solution de notre problème d'origine est  $U_n(n, -1) = n^n(H_n - 1)$ .

## 6.5 NOMBRES DE BERNOULLI

La suite de nombres que nous allons voir maintenant porte le nom de Jakob Bernoulli (1654–1705), qui découvrit de curieuses relations en résolvant des sommes de puissances mièmes [26]. Soit

$$S_m(n) = 0^m + 1^m + \cdots + (n-1)^m = \sum_{k=0}^{n-1} k^m = \sum_0^n x^m dx \quad (6.77)$$

Remarquons que si  $m > 0$ , alors on retrouve les nombres harmoniques généralisés :  $S_m(n) = H_{n-1}^{(-m)}$ . Bernoulli, en observant la suite de formules ci-dessous, y distingua un motif :

$$\begin{aligned}
 S_0(n) &= n \\
 S_1(n) &= \frac{1}{2}n^2 - \frac{1}{2}n \\
 S_2(n) &= \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n \\
 S_3(n) &= \frac{1}{4}n^4 - \frac{1}{2}n^3 + \frac{1}{4}n^2 \\
 S_4(n) &= \frac{1}{5}n^5 - \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n \\
 S_5(n) &= \frac{1}{6}n^6 - \frac{1}{2}n^5 + \frac{5}{12}n^4 - \frac{1}{12}n^2 \\
 S_6(n) &= \frac{1}{7}n^7 - \frac{1}{2}n^6 + \frac{1}{2}n^5 - \frac{1}{6}n^3 + \frac{1}{42}n \\
 S_7(n) &= \frac{1}{8}n^8 - \frac{1}{2}n^7 + \frac{7}{12}n^6 - \frac{7}{24}n^4 + \frac{1}{12}n^2 \\
 S_8(n) &= \frac{1}{9}n^9 - \frac{1}{2}n^8 + \frac{2}{3}n^7 - \frac{7}{15}n^5 + \frac{2}{9}n^3 - \frac{1}{30}n \\
 S_9(n) &= \frac{1}{10}n^{10} - \frac{1}{2}n^9 + \frac{3}{4}n^8 - \frac{7}{10}n^6 + \frac{1}{2}n^4 - \frac{3}{20}n^2 \\
 S_{10}(n) &= \frac{1}{11}n^{11} - \frac{1}{2}n^{10} + \frac{5}{6}n^9 - n^7 + n^5 - \frac{1}{2}n^3 + \frac{5}{66}n
 \end{aligned}$$

L'avez-vous deviné aussi ? Le coefficient de  $n^{m+1}$  dans  $S_m(n)$  vaut toujours  $1/(m+1)$ . Le coefficient de  $n^m$  vaut toujours  $-1/2$ . Celui de  $n^{m-1}$  vaut toujours ... voyons ...  $m/12$ . Celui de  $n^{m-2}$  vaut toujours zéro. Celui de  $n^{m-3}$  vaut toujours ... réfléchissons un peu ... hmmm ... oui, il vaut  $-m(m-1)(m-2)/720$ . Celui de  $n^{m-4}$  est toujours nul. On dirait bien que cela continue ainsi, que le coefficient de  $n^{m-k}$  est toujours égal à une constante multipliée par  $m^k$ . C'est ce que Bernoulli découvrit empiriquement, sans en donner la preuve. Avec les notations modernes, les sommes s'écrivent

$$\begin{aligned}
 S_m(n) &= \frac{1}{m+1} \left( B_0 n^{m+1} + \binom{m+1}{1} B_1 n^m + \cdots + \binom{m+1}{m} B_m n \right) \\
 &= \frac{1}{m+1} \sum_{k=0}^m \binom{m+1}{k} B_k n^{m+1-k}.
 \end{aligned} \tag{6.78}$$

Les nombres de Bernoulli sont définis par la relation de récurrence implicite suivante :

$$\sum_{j=0}^m \binom{m+1}{j} B_j = [m=0], \quad \text{pour tout } m \geq 0. \tag{6.79}$$

Par exemple,  $\binom{2}{0} B_0 + \binom{2}{1} B_1 = 0$ . Voici les premiers de ces nombres :

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12
$B_n$	1	$-\frac{1}{2}$	$\frac{1}{6}$	0	$-\frac{1}{30}$	0	$\frac{1}{42}$	0	$-\frac{1}{30}$	0	$\frac{5}{66}$	0	$-\frac{691}{2730}$

L'apparition de l'étrange fraction  $-691/2730$  semble bien anéantir tout espoir raisonnable de trouver une forme close simple pour les  $B_n$ .

On peut prouver la formule de Bernoulli (6.78) par induction sur  $m$ , en appliquant la méthode de perturbation (elle nous a déjà permis de calculer  $S_2(n) = \square_n$  au chapitre 2) :

$$\begin{aligned} S_{m+1}(n) + n^{m+1} &= \sum_{k=0}^{n-1} (k+1)^{m+1} \\ &= \sum_{k=0}^{n-1} \sum_{j=0}^{m+1} \binom{m+1}{j} k^j = \sum_{j=0}^{m+1} \binom{m+1}{j} S_j(n). \end{aligned} \quad (6.80)$$

Soit  $\widehat{S}_m(n)$  le membre droit de (6.78). Nous voulons montrer que  $S_m(n) = \widehat{S}_m(n)$ , en supposant que  $S_j(n) = \widehat{S}_j(n)$  pour tout  $0 \leq j < m$ . Commençons comme nous avons commencé au chapitre 2 pour  $m = 2$ , en soustrayant  $S_{m+1}(n)$  des deux membres de (6.80). Puis appliquons (6.78) pour développer chacun des  $S_j(n)$ , et regroupons le tout de sorte que les puissances de  $n$  du membre droit se simplifient :

$$\begin{aligned} n^{m+1} &= \sum_{j=0}^m \binom{m+1}{j} S_j(n) = \sum_{j=0}^m \binom{m+1}{j} \widehat{S}_j(n) + \binom{m+1}{m} \Delta \\ &= \sum_{j=0}^m \binom{m+1}{j} \frac{1}{j+1} \sum_{k=0}^j \binom{j+1}{k} B_k n^{j+1-k} + (m+1) \Delta \\ &= \sum_{0 \leq k \leq j \leq m} \binom{m+1}{j} \binom{j+1}{k} \frac{B_k}{j+1} n^{j+1-k} + (m+1) \Delta \\ &= \sum_{0 \leq k \leq j \leq m} \binom{m+1}{j} \binom{j+1}{j-k} \frac{B_{j-k}}{j+1} n^{k+1} + (m+1) \Delta \\ &= \sum_{0 \leq k \leq j \leq m} \binom{m+1}{j} \binom{j+1}{k+1} \frac{B_{j-k}}{j+1} n^{k+1} + (m+1) \Delta \\ &= \sum_{0 \leq k \leq m} \frac{n^{k+1}}{k+1} \sum_{k \leq j \leq m} \binom{m+1}{j} \binom{j}{k} B_{j-k} + (m+1) \Delta \\ &= \sum_{0 \leq k \leq m} \frac{n^{k+1}}{k+1} \binom{m+1}{k} \sum_{k \leq j \leq m} \binom{m+1-k}{j-k} B_{j-k} + (m+1) \Delta \\ &= \sum_{0 \leq k \leq m} \frac{n^{k+1}}{k+1} \binom{m+1}{k} \sum_{0 \leq j \leq m-k} \binom{m+1-k}{j} B_j + (m+1) \Delta \\ &= \sum_{0 \leq k \leq m} \frac{n^{k+1}}{k+1} \binom{m+1}{k} [m-k=0] + (m+1) \Delta \end{aligned}$$

$$\begin{aligned}
 &= \frac{n^{m+1}}{m+1} \binom{m+1}{m} + (m+1)\Delta \\
 &= n^{m+1} + (m+1)\Delta, \quad \text{avec } \Delta = S_m(n) - \widehat{S}_m(n).
 \end{aligned}$$

Notons en passant que ce calcul met en pratique un bon nombre de manipulations standard que nous avons apprises au chapitre 5. Nous obtenons donc  $\Delta = 0$  et  $S_m(n) = \widehat{S}_m(n)$ . CQFD.

*Encore une fois,  
vous pouvez sau-  
ter ce qui suit en  
première lecture.*

— Votre  
sympathique chargé  
de TD

↓  
Sautez  
à partir d'ici

Dans le chapitre 7, nous utiliserons les fonctions génératrices pour obtenir une preuve bien plus simple de (6.78). L'idée maîtresse consistera à prouver que les nombres de Bernoulli sont les coefficients de la série

$$\cdot \frac{z}{e^z - 1} = \sum_{n \geq 0} B_n \frac{z^n}{n!}. \quad (6.81)$$

Admettons pour l'instant ce résultat, et observons quelques-unes de ses conséquences. Si on ajoute  $\frac{1}{2}z$  aux deux membres de (6.81), on supprime le terme  $B_1 z / 1! = -\frac{1}{2}z$  du membre droit et on obtient

$$\frac{z}{e^z - 1} + \frac{z}{2} = \frac{z}{2} \frac{e^z + 1}{e^z - 1} = \frac{z}{2} \frac{e^{z/2} + e^{-z/2}}{e^{z/2} - e^{-z/2}} = \frac{z}{2} \coth \frac{z}{2}, \quad (6.82)$$

où  $\coth$  désigne la cotangente hyperbolique, fonction bien connue définie par l'expression  $\cosh z / \sinh z$ , où

$$\sinh z = \frac{e^z - e^{-z}}{2}; \quad \cosh z = \frac{e^z + e^{-z}}{2}. \quad (6.83)$$

En remplaçant  $z$  par  $-z$ , on obtient  $(-\frac{z}{2}) \coth(-\frac{z}{2}) = \frac{z}{2} \coth \frac{z}{2}$ , donc tous les coefficients d'indice impair de  $\frac{z}{2} \coth \frac{z}{2}$  sont nuls, et

$$B_3 = B_5 = B_7 = B_9 = B_{11} = B_{13} = \cdots = 0. \quad (6.84)$$

On peut aussi déduire de (6.82) une forme close pour les coefficients de  $\coth$  :

$$\begin{aligned}
 z \coth z &= \frac{2z}{e^{2z} - 1} + \frac{2z}{2} \\
 &= \sum_{n \geq 0} B_{2n} \frac{(2z)^{2n}}{(2n)!} = \sum_{n \geq 0} 4^n B_{2n} \frac{z^{2n}}{(2n)!}.
 \end{aligned} \quad (6.85)$$

Cependant, les fonctions hyperboliques ne sont pas très populaires ; les gens préfèrent en général les "vraies" fonctions trigonométriques. Celles-ci peuvent s'exprimer en fonction de leurs cousines hyperboliques :

$$\sin z = -i \sinh iz, \quad \cos z = \cosh iz. \quad (6.86)$$

Voici leurs développements en séries :

$$\begin{aligned}\sin z &= \frac{z^1}{1!} - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots, & \sinh z &= \frac{z^1}{1!} + \frac{z^3}{3!} + \frac{z^5}{5!} + \dots; \\ \cos z &= \frac{z^0}{0!} - \frac{z^2}{2!} + \frac{z^4}{4!} - \dots, & \cosh z &= \frac{z^0}{0!} + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots.\end{aligned}$$

Par conséquent,  $\cot z = \cos z / \sin z = i \cosh iz / \sinh iz = i \coth iz$ , et on a

$$z \cot z = \sum_{n \geq 0} B_{2n} \frac{(2iz)^{2n}}{(2n)!} = \sum_{n \geq 0} (-4)^n B_{2n} \frac{z^{2n}}{(2n)!}. \quad (6.87)$$

*Je vois. On obtient des "vraies" fonctions en utilisant des nombres imaginaires.*

Il existe une autre remarquable formule pour  $z \cot z$  due à Euler (exercice 73) :

$$z \cot z = 1 - 2 \sum_{k \geq 1} \frac{z^2}{k^2 \pi^2 - z^2}. \quad (6.88)$$

On peut la développer en puissances de  $z^2$  pour obtenir

$$\begin{aligned}z \cot z &= 1 - 2 \sum_{k \geq 1} \left( \frac{z^2}{k^2 \pi^2} + \frac{z^4}{k^4 \pi^4} + \frac{z^6}{k^6 \pi^6} + \dots \right) \\ &= 1 - 2 \left( \frac{z^2}{\pi^2} H_{\infty}^{(2)} + \frac{z^4}{\pi^4} H_{\infty}^{(4)} + \frac{z^6}{\pi^6} H_{\infty}^{(6)} + \dots \right).\end{aligned}$$

En mettant les coefficients de  $z^{2n}$  en équations avec ceux de la formule (6.87), on aboutit à une forme close quasi-miraculeuse pour une infinité de sommes :

$$\zeta(2n) = H_{\infty}^{(2n)} = (-1)^{n-1} \frac{2^{2n-1} \pi^{2n} B_{2n}}{(2n)!}, \quad n > 0 \text{ entier.} \quad (6.89)$$

Par exemple,

$$\zeta(2) = H_{\infty}^{(2)} = 1 + \frac{1}{4} + \frac{1}{9} + \dots = \pi^2 B_2 = \pi^2/6; \quad (6.90)$$

$$\zeta(4) = H_{\infty}^{(4)} = 1 + \frac{1}{16} + \frac{1}{81} + \dots = -\pi^4 B_4/3 = \pi^4/90. \quad (6.91)$$

La formule (6.89) n'est pas seulement une forme close pour  $H_{\infty}^{(2n)}$ ; elle nous donne aussi une valeur approchée de  $B_{2n}$ , car  $H_{\infty}^{(2n)}$  est très proche de 1 lorsque  $n$  est grand. D'autre part, elle nous apprend que  $(-1)^{n-1} B_{2n} > 0$  pour tout  $n > 0$ , donc que les nombres de Bernoulli non nuls sont à signes alternants.

Ce n'est pas tout. Les nombres de Bernoulli apparaissent aussi dans les coefficients de la fonction tangente,

$$\tan z = \frac{\sin z}{\cos z} = \sum_{n \geq 0} (-1)^{n-1} 4^n (4^n - 1) B_{2n} \frac{z^{2n-1}}{(2n)!}, \quad (6.92)$$

↓  
Salez  
encore plus  
haut

tout comme dans les autres fonctions trigonométriques (exercice 72). La formule (6.92) entraîne un autre résultat important sur les nombres de Bernoulli, à savoir que

$$T_{2n-1} = (-1)^{n-1} \frac{4^n (4^n - 1)}{2n} B_{2n} \text{ est un entier } > 0. \quad (6.93)$$

Voici les premiers nombres  $T$ , que l'on appelle *nombres tangents* :

n	1	3	5	7	9	11	13
$T_n$	1	2	16	272	7936	353792	22368256

Pour prouver (6.93), on peut, suivant en cela une idée de B. F. Logan, considérer la série

$$\begin{aligned} \frac{\sin z + x \cos z}{\cos z - x \sin z} &= x + (1+x^2)z + (2x^3+2x)\frac{z^2}{2} + (6x^4+8x^2+2)\frac{z^3}{6} + \dots \\ &= \sum_{n \geq 0} T_n(x) \frac{z^n}{n!}, \end{aligned} \quad (6.94)$$

Lorsque  $x = \tan w$ , c'est égal à  $\tan(z+w)$ . Donc, d'après la formule de Taylor, la dérivée  $n$ ième de  $\tan w$  est  $T_n(\tan w)$ .

tandis qu'en dérivant par rapport à  $z$ , on a

$$\frac{1+x^2}{(\cos z - x \sin z)^2} = \sum_{n \geq 1} T_n(x) \frac{z^{n-1}}{(n-1)!} = \sum_{n \geq 0} T_{n+1}(x) \frac{z^n}{n!},$$

(essayez, vous verrez que les simplifications sont très jolies à voir). Par conséquent,

$$T_{n+1}(x) = (1+x^2)T'_n(x), \quad T_0(x) = x. \quad (6.95)$$

On déduit de cette récurrence toute simple que les coefficients de  $T_n(x)$  sont des entiers positifs ou nuls. De plus, on peut aisément démontrer que  $T_n(x)$  est de degré  $n+1$  et que ses coefficients sont alternativement nuls

ou strictement positifs. Par conséquent  $T_{2n+1}(0) = T_{2n+1}$  est un entier strictement positif, ce qui confirme bien (6.93).

La récurrence (6.95) nous fournit un moyen simple de calculer les nombres de Bernoulli, en passant par les nombres tangents et en n'effectuant que des opérations simples sur des entiers. C'est bien plus efficace que la récurrence de départ (6.79), qui donne lieu à des opérations difficiles sur des fractions.

Si on veut calculer la somme des puissances nièmes de  $a$  à  $b - 1$  plutôt que de 0 to  $n - 1$ , il suffit d'appliquer la théorie du chapitre 2 :

$$\sum_{k=a}^{b-1} k^m = \sum_a^b x^m dx = S_m(b) - S_m(a). \quad (6.96)$$

Cette identité a d'intéressantes conséquences lorsqu'on donne des valeurs négatives à  $k$  :

$$\sum_{k=-n+1}^{-1} k^m = (-1)^m \sum_{k=0}^{n-1} k^m, \quad \text{pour } m > 0,$$

donc

$$S_m(0) - S_m(-n + 1) = (-1)^m (S_m(n) - S_m(0)).$$

Or,  $S_m(0) = 0$ , ce qui donne l'identité

$$S_m(1 - n) = (-1)^{m+1} S_m(n), \quad m > 0. \quad (6.97)$$

Donc  $S_m(1) = 0$ . Si on écrit le polynôme  $S_m(n)$  sous forme factorisée, il contiendra toujours les facteurs  $n$  et  $(n - 1)$ , car il admet les racines 0 et 1. On peut dire, plus généralement, que  $S_m(n)$  est un polynôme de degré  $m + 1$  et de coefficient directeur  $\frac{1}{m+1}$ . De plus, si on pose  $n = \frac{1}{2}$  dans (6.97), on obtient  $S_m(\frac{1}{2}) = (-1)^{m+1} S_m(\frac{1}{2})$ ; donc, si  $m$  est pair,  $S_m(\frac{1}{2}) = 0$ , et  $(n - \frac{1}{2})$  constitue un facteur de plus. Tout ceci explique pourquoi nous sommes tombés sur la factorisation toute simple

$$S_2(n) = \frac{1}{3}n(n - \frac{1}{2})(n - 1)$$

au chapitre 2. Nous aurions ainsi pu la trouver sans la calculer, simplement avec un raisonnement de ce genre ! Remarquons encore que (6.97) implique que le polynôme  $\hat{S}_m(n) = S_m(n)/(n - \frac{1}{2})$  satisfait

$$\hat{S}_m(1 - n) = \hat{S}_m(n), \quad m \text{ pair}, \quad m > 0.$$

*Johann Faulhaber utilisa implicitement (6.97) en 1635 [119] pour exprimer simplement  $S_m(n)$  comme un polynôme en  $n(n + 1)/2$  lorsque  $m \leq 17$  (voir [222])).*

Il s'ensuit que  $S_m(n)$  peut toujours s'écrire sous la forme factorisée suivante :

$$S_m(n) = \begin{cases} \frac{1}{m+1} \prod_{k=1}^{\lceil m/2 \rceil} (n - \frac{1}{2} - \alpha_k)(n - \frac{1}{2} + \alpha_k), & m \text{ impair;} \\ \frac{(n - \frac{1}{2})}{m+1} \prod_{k=1}^{m/2} (n - \frac{1}{2} - \alpha_k)(n - \frac{1}{2} + \alpha_k), & m \text{ pair.} \end{cases} \quad (6.98)$$

Dans cette formule,  $\alpha_1 = \frac{1}{2}$  et  $\alpha_2, \dots, \alpha_{\lceil m/2 \rceil}$  sont des nombres complexes qui dépendent de  $m$ . Par exemple,

$$\begin{aligned} S_3(n) &= n^2(n-1)^2/4; \\ S_4(n) &= n(n-\frac{1}{2})(n-1)(n-\frac{1}{2}+\sqrt{7/12})(n-\frac{1}{2}-\sqrt{7/12})/5; \\ S_5(n) &= n^2(n-1)^2(n-\frac{1}{2}+\sqrt{3/4})(n-\frac{1}{2}-\sqrt{3/4})/6; \\ S_6(n) &= n(n-\frac{1}{2})(n-1)(n-\frac{1}{2}+\alpha)(n-\frac{1}{2}-\alpha)(n-\frac{1}{2}+\bar{\alpha})(n-\frac{1}{2}-\bar{\alpha})/7, \\ &\text{avec } \alpha = 2^{-3/2} 3^{-1/4} (\sqrt{\sqrt{31}+\sqrt{27}} + i\sqrt{\sqrt{31}-\sqrt{27}}). \end{aligned}$$

Si  $m$  est impair et  $> 1$ , on a  $B_m = 0$ , donc  $S_m(n)$  n'est pas divisible par  $n^2$  (ni par  $(n-1)^2$ ). A part cela, les racines de  $S_m(n)$  n'ont pas l'air d'obéir à une règle simple.

Pour terminer notre étude des nombres de Bernoulli, observons leurs relations avec le nombres de Stirling. Pour calculer  $S_m(n)$ , on peut avantageusement transformer les puissances ordinaires en puissances descendantes, car ces dernières sont faciles à sommer. Une fois les sommes faites, on n'a plus qu'à reconvertir en puissances ordinaires :

$$\begin{aligned} S_m(n) &= \sum_{k=0}^{n-1} k^m = \sum_{k=0}^{n-1} \sum_{j \geq 0} \left\{ \begin{matrix} m \\ j \end{matrix} \right\} k^j = \sum_{j \geq 0} \left\{ \begin{matrix} m \\ j \end{matrix} \right\} \sum_{k=0}^{n-1} k^j \\ &= \sum_{j \geq 0} \left\{ \begin{matrix} m \\ j \end{matrix} \right\} \frac{n^{j+1}}{j+1} \\ &= \sum_{j \geq 0} \left\{ \begin{matrix} m \\ j \end{matrix} \right\} \frac{1}{j+1} \sum_{k \geq 0} (-1)^{j+1-k} \binom{j+1}{k} n^k. \end{aligned}$$

En mettant les coefficients en équations avec ceux de (6.78) on trouve l'identité

$$\sum_{j \geq 0} \left\{ \begin{matrix} m \\ j \end{matrix} \right\} \binom{j+1}{k} \frac{(-1)^{j+1-k}}{j+1} = \frac{1}{m+1} \binom{m+1}{k} B_{m+1-k}, \quad k > 0. \quad (6.99)$$

Commencer à redescendre

Nous aimerais bien pouvoir démontrer cette relation directement, pour ainsi découvrir les nombres de Bernoulli d'une autre manière. Cependant, il n'y a rien dans les tables 280 et 281 qui nous donne un moyen évident de prouver par induction que la somme du membre gauche de (6.99) est égale à une constante multipliée par  $m^{k-1}$ . Si  $k = m + 1$ , c'est facile, car la somme vaut  $\{m\}_{m+1} / (m+1) = 1/(m+1)$ . Si  $k = m$ , elle vaut  $\{m\}_{m-1} \{m\}_m m^{-1} - \{m\}_{m-1} \{m+1\}_m (m+1)^{-1} = \frac{1}{2}(m-1) - \frac{1}{2}m = -\frac{1}{2}$ , donc ce cas n'offre pas de difficulté non plus. Par contre, si  $k < m$ , la somme devient franchement compliquée. Bernoulli n'aurait probablement pas découvert ses nombres s'il avait emprunté ce chemin.

Ce que nous pouvons faire, c'est remplacer  $\{m\}_j$  par  $\{m+1\}_{j+1} - (j+1)\{m\}_{j+1}$ . Le facteur  $(j+1)$  a le bon goût de se simplifier avec le dénominateur qui nous gène, et le membre gauche devient

$$\sum_{j \geq 0} \left\{ \begin{matrix} m+1 \\ j+1 \end{matrix} \right\} \left[ \begin{matrix} j+1 \\ k \end{matrix} \right] \frac{(-1)^{j+1-k}}{j+1} - \sum_{j \geq 0} \left\{ \begin{matrix} m \\ j+1 \end{matrix} \right\} \left[ \begin{matrix} j+1 \\ k \end{matrix} \right] (-1)^{j+1-k}.$$

D'après (6.31) la seconde somme est nulle lorsque  $k < m$ . Il nous reste la première, qui a désespérément besoin d'un changement de notation. Renommons donc toutes les variables afin que l'indice de sommation s'appelle  $k$  et que les autres variables soient  $m$  et  $n$ . Dès lors, l'identité (6.99) est équivalente à

$$\sum_k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \left[ \begin{matrix} k \\ m \end{matrix} \right] \frac{(-1)^{k-m}}{k} = \frac{1}{n} \binom{n}{m} B_{n-m} + [m=n-1], \quad m > 0. \quad (6.100)$$

C'est quand même plus agréable à voir. Cependant, la table 281 n'a toujours pas de suggestion évidente à nous proposer pour la suite.

Ce sont finalement les convolutions de la table 289 qui viennent à notre secours. Nous pouvons en effet utiliser (6.51) et (6.50) pour réécrire le terme général en fonction des polynômes de Stirling :

$$\begin{aligned} \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \left[ \begin{matrix} k \\ m \end{matrix} \right] &= (-1)^{n-k+1} \frac{n!}{(k-1)!} \sigma_{n-k}(-k) \cdot \frac{k!}{(m-1)!} \sigma_{k-m}(k); \\ \left\{ \begin{matrix} n \\ k \end{matrix} \right\} \left[ \begin{matrix} k \\ m \end{matrix} \right] \frac{(-1)^{k-m}}{k} &= (-1)^{n+1-m} \frac{n!}{(m-1)!} \sigma_{n-k}(-k) \sigma_{k-m}(k). \end{aligned}$$

Tout s'arrange : la convolution de (6.48), avec  $t = 1$ , entraîne que

$$\begin{aligned} \sum_{k=0}^n \sigma_{n-k}(-k) \sigma_{k-m}(k) &= \sum_{k=0}^{n-m} \sigma_{n-m-k}(-n + (n-m-k)) \sigma_k(m+k) \\ &= \frac{m-n}{(m)(-n)} \sigma_{n-m}(m-n+(n-m)). \end{aligned}$$

Atterrir  
ici.

La formule (6.100) est maintenant prouvée, et nous voyons que les nombres de Bernoulli sont liés aux termes constants des polynômes de Stirling :

$$\frac{B_m}{m!} = -m\sigma_m(0). \quad (6.101)$$

## 6.6 NOMBRES DE FIBONACCI

Nous voici arrivés à la suite de nombres qui est peut-être la plus agréable de toutes, la suite de Fibonacci  $\langle F_n \rangle$  :

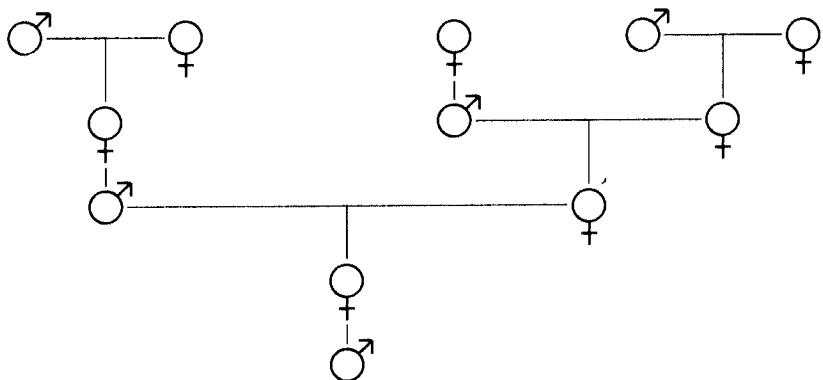
n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
F <sub>n</sub>	0	1	1	2	3	5	8	13	21	34	55	89	144	233	377

A la différence des nombres harmoniques et des nombres de Bernoulli, les nombres de Fibonacci sont des entiers. Ils sont définis par la récurrence

$$\begin{aligned} F_0 &= 0; & F_1 &= 1; \\ F_n &= F_{n-1} + F_{n-2}, & \text{pour } n > 1. \end{aligned} \quad (6.102)$$

C'est parce que cette règle est très simple — la plus simple possible dans laquelle chaque nombre dépend des deux précédents — que les nombres de Fibonacci apparaissent naturellement dans des situations très variées.

Les arbres généalogiques d'abeilles en sont un bon exemple. Considérons l'ascendance d'une abeille mâle. Chaque mâle (qu'on appelle aussi faux bourdon) est produit de façon asexuée par une femelle (qu'on appelle reine). Par contre, chaque femelle a deux parents, un mâle et une femelle. Voici les premiers niveaux de l'arbre :

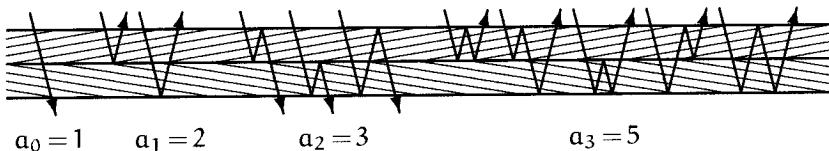


Le faux bourdon a un grand-père et une grand-mère, un arrière grand-père et deux arrière-grand-mères, deux arrière-arrière-grands-pères et trois

arrière-arrière-grand-mères. On montre facilement par induction qu'il a en fait exactement  $F_{n+1}$  arrière<sup>n</sup>-grands-pères et  $F_{n+2}$  arrière<sup>n</sup>-grand-mères.

Les nombres de Fibonacci sont très présents dans la nature. Par exemple, les fleurs de tournesol forment des sortes de spirales compactes, ayant généralement 34 rayons dans un sens et 55 dans l'autre, ou 21 et 34 respectivement, ou encore 13 et 21. On a même découvert un jour en Angleterre un tournesol géant avec 89 et 144 spirales. On trouve des motifs similaires dans les fruits de certains conifères.

Voici un exemple d'un autre genre [277]. Supposons que nous mettions deux plaques de verre face contre face. De combien de façons possibles  $a_n$  un rayon de lumière peut-il traverser les deux plaques ou être réfléchi après avoir changé de direction  $n$  fois ? Un petit dessin valant mieux qu'un long discours, voici les premiers cas possibles :



Lorsque  $n$  est pair, le nombre de réflexions est pair et le rayon franchit les deux plaques. Lorsque  $n$  est impair, le rayon finit par retourner du côté d'où il est parti. On dirait bien que les  $a_n$  sont des nombres de Fibonacci. Pour confirmer cela, regardons d'un peu plus près le parcours d'un rayon sujet à  $n$  réflexions, lorsque  $n \geq 2$  : soit il est réfléchi d'abord par la deuxième surface (la plus éloignée de son point de départ) et peut continuer son parcours de  $a_{n-1}$  façons différentes, soit il est réfléchi d'abord par la surface du milieu, puis encore une fois pour finir son trajet de  $a_{n-2}$  manières possibles. On retrouve ainsi la récurrence de Fibonacci  $a_n = a_{n-1} + a_{n-2}$ . Les conditions initiales sont juste un peu différentes, car  $a_0 = 1 = F_2$  et  $a_1 = 2 = F_3$  ; tout est donc simplement décalé de deux places, et  $a_n = F_{n+2}$ .

C'est Leonardo Fibonacci qui introduisit ces nombres en 1202. Par la suite, les mathématiciens se mirent à découvrir des choses de plus en plus intéressantes à leur propos. Edouard Lucas, l'homme de la Tour de Hanoi que nous avons étudiée au chapitre 1, les étudia intensivement au cours de la deuxième moitié du dix-neuvième siècle (c'est même lui qui popularisa le nom de "nombres de Fibonacci"). Il obtint des résultats étonnans : par exemple, il utilisa les propriétés des nombres de Fibonacci pour démontrer que le nombre de Mersenne à 39 chiffres  $2^{127} - 1$  est premier.

L'identité suivante, que l'astronome français Jean-Dominique Cassini publia en 1680 [51], est l'un des plus anciens théorèmes sur les nombres de Fibonacci (en fait, Johannes Kepler le connaissait déjà en 1608 [202]) :

$$F_{n+1} F_{n-1} - F_n^2 = (-1)^n, \quad \text{pour } n > 0. \quad (6.103)$$

*Phyllotaxie, n. f.  
Attirance pour les  
taxis.*

*"La suite de Fibonacci possède des propriétés nombreuses fort intéressantes."*

— E. Lucas [259]

Lorsque  $n = 6$  par exemple, l'identité de Cassini prétend, avec raison, que  $13 \cdot 5 - 8^2$  est égal à 1.

Toute formule polynomiale qui contient des nombres de Fibonacci de la forme  $F_{n\pm k}$ , pour des petites valeurs de  $k$ , peut être convertie en une formule où n'apparaissent que des  $F_n$  et des  $F_{n+1}$ . A cette fin, on utilise la règle

$$F_m = F_{m+2} - F_{m+1} \quad (6.104)$$

pour exprimer  $F_m$  en fonction de nombres de Fibonacci plus grands lorsque  $m < n$ , et

$$F_m = F_{m-2} + F_{m-1} \quad (6.105)$$

pour remplacer  $F_m$  par des nombres de Fibonacci plus petits lorsque  $m > n + 1$ . Ainsi, par exemple, on peut remplacer  $F_{n-1}$  par  $F_{n+1} - F_n$  dans (6.103) pour obtenir une autre forme de l'identité de Cassini :

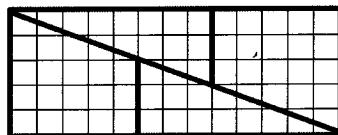
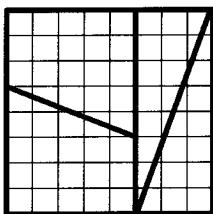
$$F_{n+1}^2 - F_{n+1} F_n - F_n^2 = (-1)^n. \quad (6.106)$$

Remarquons que l'identité originale de Cassini s'écrit

$$F_{n+2} F_n - F_{n+1}^2 = (-1)^{n+1}$$

si on remplace  $n$  par  $n + 1$  ; c'est équivalent à la formule  $(F_{n+1} + F_n)F_n - F_{n+1}^2 = (-1)^{n+1}$ , qui est elle-même équivalente à (6.106). Par conséquent, Cassini( $n$ ) est vrai si et seulement si Cassini( $n+1$ ) est vrai, et donc l'équation (6.103) est vraie pour tout  $n$ , par induction.

L'identité de Cassini est à la base d'un paradoxe géométrique qui était l'un des casse-tête favoris de Lewis Carroll [63], [319], [364]. Il s'agit de prendre un échiquier et de le couper en quatre parts, comme sur le dessin, puis de disposer ces parts pour former un rectangle :



Problème : l'échiquier comportait  $8 \times 8 = 64$  cases et le rectangle en contient  $5 \times 13 = 65$  ! De façon similaire, on peut découper n'importe quel carré  $F_n \times F_n$  en quatre parts ayant  $F_{n+1}$ ,  $F_n$ ,  $F_{n-1}$ , et  $F_{n-2}$  dans leurs dimensions (celles de l'exemple sont 13, 8, 5 et 3). On obtient un rectangle  $F_{n-1} \times F_{n+1}$ .

D'après (6.103), on a ainsi gagné ou perdu une case, selon que  $n$  est pair ou impair.

A strictement parler, l'identité (6.105) ne peut s'appliquer que si  $m \geq 2$ , car nous n'avons pas défini  $F_n$  pour  $n$  strictement négatif. Pour faciliter les choses, supprimons donc cette condition et définissons les nombres de Fibonacci négatifs au moyen des expressions (6.104) et (6.105). Par exemple,  $F_{-1}$  est égal à  $F_1 - F_0 = 1$  et  $F_{-2}$  vaut  $F_0 - F_{-1} = -1$ . Au vu des premières valeurs,

$n$	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
$F_n$	0	1	-1	2	-3	5	-8	13	-21	34	-55	89

il est clair (et cela se prouve par induction) que

$$F_{-n} = (-1)^{n-1} F_n, \quad n \text{ entier.} \quad (6.107)$$

Ainsi, l'identité de Cassini (6.103) est vraie pour *tout* entier  $n$ , et non plus seulement pour  $n > 0$ .

Si on applique de façon répétée les équations (6.105) et (6.104), on obtient des expressions de  $F_{n \pm k}$  en fonction de  $F_n$  et  $F_{n+1}$ :

$$\begin{array}{ll} F_{n+2} = F_{n+1} + F_n & F_{n-1} = F_{n+1} - F_n \\ F_{n+3} = 2F_{n+1} + F_n & F_{n-2} = -F_{n+1} + 2F_n \\ F_{n+4} = 3F_{n+1} + 2F_n & F_{n-3} = 2F_{n+1} - 3F_n \\ F_{n+5} = 5F_{n+1} + 3F_n & F_{n-4} = -3F_{n+1} + 5F_n \end{array}$$

On y distingue facilement le motif

$$F_{n+k} = F_k F_{n+1} + F_{k-1} F_n, \quad (6.108)$$

valable pour tous entiers  $k$  et  $n$ , et facilement démontrable par induction.

En posant  $k = n$  dans (6.108), on trouve que

$$F_{2n} = F_n F_{n+1} + F_{n-1} F_n, \quad (6.109)$$

donc que  $F_{2n}$  est un multiple de  $F_n$ . De même,

$$F_{3n} = F_{2n} F_{n+1} + F_{2n-1} F_n,$$

et  $F_{3n}$  est aussi un multiple de  $F_n$ . Par induction encore, on prouve que

$$F_{kn} \text{ est un multiple de } F_n, \quad (6.110)$$

pour tous entiers  $k$  et  $n$ . Ceci explique pourquoi  $F_{15}$  (qui vaut 610) est à la fois un multiple de  $F_3$  et de  $F_5$  (qui valent respectivement 2 et 5). Il y a

*Le paradoxe s'explique par le fait que... non, un magicien ne dévoile jamais ses trucs.*

même plus encore : l'exercice 27 montre que

$$\text{pgcd}(F_m, F_n) = F_{\text{pgcd}(m, n)}. \quad (6.111)$$

Par exemple,  $\text{pgcd}(F_{12}, F_{18}) = \text{pgcd}(144, 2584) = 8 = F_6$ .

Nous pouvons maintenant démontrer une proposition presque réciproque de (6.110) : si  $n > 2$  et  $F_m$  est un multiple de  $F_n$ , alors  $m$  est un multiple de  $n$ . En effet, si  $F_n \nmid F_m$ , alors  $F_n \nmid \text{pgcd}(F_m, F_n) = F_{\text{pgcd}(m, n)} \leq F_n$ . Cela n'est possible que si  $F_{\text{pgcd}(m, n)} = F_n$  ; donc  $\text{pgcd}(m, n) = n$  car  $n > 2$ . Yuri Matijasevich se sert d'une extension de ce résultat dans sa célèbre preuve [266] du fait qu'il n'existe pas d'algorithme permettant de décider si une équation polynomiale à plusieurs variables et coefficients entiers a une solution en nombres entiers. Le lemme qu'il utilise est le suivant : pour tout  $n > 2$ , le nombre de Fibonacci  $F_m$  est un multiple de  $F_n^2$  si et seulement si  $m$  est un multiple de  $nF_n$ .

Nous allons le démontrer en observant la suite  $\langle F_{kn} \bmod F_n^2 \rangle$  pour  $k = 1, 2, 3, \dots$  pour voir à quels moments  $F_{kn} \bmod F_n^2 = 0$  (nous savons déjà que  $m$  doit être un multiple de  $n$  pour que  $F_m \bmod F_n = 0$ ). D'abord nous avons  $F_n \bmod F_n^2 = F_n$  ; ce n'est pas nul. Puis

$$F_{2n} = F_n F_{n+1} + F_{n-1} F_n \equiv 2F_n F_{n+1} \pmod{F_n^2},$$

d'après (6.108), car  $F_{n+1} \equiv F_{n-1} \pmod{F_n}$ . De même,

$$F_{2n+1} = F_{n+1}^2 + F_n^2 \equiv F_{n+1}^2 \pmod{F_n^2}.$$

Cette congruence nous permet de calculer

$$\begin{aligned} F_{3n} &= F_{2n+1} F_n + F_{2n} F_{n-1} \\ &\equiv F_{n+1}^2 F_n + (2F_n F_{n+1}) F_{n-1} = 3F_{n+1}^2 F_n \pmod{F_n^2}; \end{aligned}$$

$$\begin{aligned} F_{3n+1} &= F_{2n+1} F_{n+1} + F_{2n} F_n \\ &\equiv F_{n+1}^3 + (2F_n F_{n+1}) F_n \equiv F_{n+1}^3 \pmod{F_n^2}. \end{aligned}$$

Plus généralement, on prouve par induction sur  $k$  que

$$F_{kn} \equiv kF_n F_{n+1}^{k-1} \quad \text{et} \quad F_{kn+1} \equiv F_{n+1}^k \pmod{F_n^2}.$$

Comme  $F_{n+1}$  est premier par rapport à  $F_n$ , on a

$$\begin{aligned} F_{kn} \equiv 0 \pmod{F_n^2} &\iff kF_n \equiv 0 \pmod{F_n^2} \\ &\iff k \equiv 0 \pmod{F_n}. \end{aligned}$$

Nous avons ainsi démontré le lemme de Matijasevich.

Voici maintenant l'une des propriétés les plus importantes des nombres de Fibonacci. Il s'agit d'un système de représentation des nombres entiers. Convenons de la notation suivante :

$$j \gg k \iff j \geq k+2. \quad (6.112)$$

Alors tout entier strictement positif peut s'écrire de façon unique sous la forme

$$n = F_{k_1} + F_{k_2} + \cdots + F_{k_r}, \quad k_1 \gg k_2 \gg \cdots \gg k_r \gg 0. \quad (6.113)$$

Cette propriété constitue le "théorème de Zeckendorf" [246], [381]. Par exemple, le nombre un million s'écrit

$$\begin{aligned} 1000000 &= 832040 + 121393 + 46368 + 144 + 55 \\ &= F_{30} + F_{26} + F_{24} + F_{12} + F_{10}. \end{aligned}$$

Pour trouver la représentation d'un nombre, on peut toujours opérer de manière "gloutonne", en prenant pour  $F_{k_1}$  le plus grand nombre de Fibonacci  $\leq n$ , puis pour  $F_{k_2}$  le plus grand qui est  $\leq n - F_{k_1}$ , et ainsi de suite. Voyons cela plus précisément : supposons que  $F_k \leq n < F_{k+1}$  ; donc  $0 \leq n - F_k < F_{k+1} - F_k = F_{k-1}$ . Si  $n$  est un nombre de Fibonacci, alors  $r = 1$ ,  $k_1 = k$  et (6.113) est respectée. Sinon, par induction sur  $n$ ,  $n - F_k$  peut s'écrire  $F_{k_2} + \cdots + F_{k_r}$  ; et (6.113) est respectée si on pose  $k_1 = k$ , car les inégalités  $F_{k_2} \leq n - F_k < F_{k-1}$  entraînent que  $k \gg k_2$ .

Réiproquement, si  $n$  s'écrit comme en (6.113), alors

$$F_{k_1} \leq n < F_{k_1+1},$$

car la plus grande valeur possible de  $F_{k_2} + \cdots + F_{k_r}$  lorsque  $k \gg k_2 \gg \cdots \gg k_r \gg 0$  est

$$F_{k-2} + F_{k-4} + \cdots + F_{k \bmod 2+2} = F_{k-1} - 1, \quad \text{si } k \geq 2. \quad (6.114)$$

Cette formule se prouve facilement par induction sur  $k$  (le membre gauche est nul lorsque  $k$  égale 2 ou 3). Par conséquent,  $k_1$  est bien la valeur choisie par l'approche gloutonne décrite plus haut, et la représentation est unique.

Tout système qui donne une représentation unique des nombres est un système de numération. Le théorème de Zeckendorf conduit donc naturellement au *système de numération de Fibonacci*. Tout entier positif ou nul  $n$  peut être représenté par une suite de 0 et de 1, en convenant que

$$n = (b_m b_{m-1} \dots b_2)_F \iff n = \sum_{k=2}^m b_k F_k. \quad (6.115)$$

Ce système de numération rappelle le système binaire, mais ici on ne pourra jamais trouver deux chiffres 1 adjacents. Voici par exemple les nombres de 1 à 20 selon notre nouveau système :

$1 = (000001)_F$	$6 = (001001)_F$	$11 = (010100)_F$	$16 = (100100)_F$
$2 = (000010)_F$	$7 = (001010)_F$	$12 = (010101)_F$	$17 = (100101)_F$
$3 = (000100)_F$	$8 = (010000)_F$	$13 = (100000)_F$	$18 = (101000)_F$
$4 = (000101)_F$	$9 = (010001)_F$	$14 = (100001)_F$	$19 = (101001)_F$
$5 = (001000)_F$	$10 = (010010)_F$	$15 = (100010)_F$	$20 = (101010)_F$

On peut comparer la représentation de Fibonacci du nombre un million, que nous avons vue il y a un instant, avec sa représentation binaire  $2^{19} + 2^{18} + 2^{17} + 2^{16} + 2^{14} + 2^9 + 2^6$ :

$$\begin{aligned}(1000000)_{10} &= (10001010000000000001010000000)_F \\ &= (11110100001001000000)_2.\end{aligned}$$

La représentation de Fibonacci est un peu plus longue, du fait qu'on n'y autorise pas les 1 adjacents. A ceci près, les deux représentations sont analogues.

Pour ajouter 1 à un nombre dans le système de Fibonacci, deux cas se présentent. Si le "chiffre des unités" est 0, on le remplace par 1 ; on ajoute ainsi  $F_2 = 1$ , car le chiffre des unités correspond à  $F_2$ . Sinon, on remplace les deux chiffres de droite, qui sont forcément 01, par 10, ce qui a pour effet d'ajouter  $F_3 - F_2 = 1$ . Pour finir, il faut faire autant de "retenues" que nécessaire, en remplaçant le motif "011" par "100", pour supprimer les éventuels chiffres 1 consécutifs dans la représentation. Faire une telle retenue équivaut en fait à remplacer  $F_{m+1} + F_m$  par  $F_{m+2}$ . Par exemple, pour passer de  $5 = (1000)_F$  à  $6 = (1001)_F$  ou de  $6 = (1001)_F$  à  $7 = (1010)_F$ , pas besoin de retenue ; par contre, il en faut deux pour passer de  $7 = (1010)_F$  à  $8 = (10000)_F$ .

Nous avons présenté jusqu'ici beaucoup de propriétés des nombres de Fibonacci, mais nous n'avons pas dit un mot sur une éventuelle forme close. Nous n'en avons pas trouvé pour les nombres de Stirling, pas plus que pour les nombres eulériens, ni pour les nombres de Bernoulli ; en revanche, nous avons su découvrir la forme close  $H_n = \left[ \frac{n+1}{2} \right] / n!$  pour les nombres harmoniques. Alors, existe-t-il une relation entre les  $F_n$  et d'autres quantités que nous connaissons ? Est-il possible de "résoudre" la récurrence qui définit  $F_n$  ?

La réponse est oui. En fait, la récurrence est facile à résoudre si on utilise l'idée de *fonction génératrice* que nous avons vue rapidement au

chapitre 5. Considérons la série infinie

$$F(z) = F_0 + F_1 z + F_2 z^2 + \cdots = \sum_{n \geq 0} F_n z^n. \quad (6.116)$$

Si nous trouvons une formule simple pour  $F(z)$ , il y a de bonnes chances pour que nous puissions trouver aussi une formule simple pour ses coefficients  $F_n$ .

Nous verrons les fonctions génératrices en détail au chapitre 7, mais l'exemple que nous allons étudier maintenant nous sera bien utile d'ici là. Si on regarde attentivement ce qui se passe lorsqu'on multiplie la série  $F(z)$  par  $z$  et par  $z^2$ , on découvre une bien jolie propriété :

$$\begin{aligned} F(z) &= F_0 + F_1 z + F_2 z^2 + F_3 z^3 + F_4 z^4 + F_5 z^5 + \cdots, \\ zF(z) &= F_0 z + F_1 z^2 + F_2 z^3 + F_3 z^4 + F_4 z^5 + \cdots, \\ z^2 F(z) &= F_0 z^2 + F_1 z^3 + F_2 z^4 + F_3 z^5 + \cdots. \end{aligned}$$

En soustrayant les deux dernières équations de la première, tous les termes en  $z^2, z^3, z^4, \dots$  disparaissent à cause de la récurrence de Fibonacci. Tout ce qui reste alors, c'est  $(F_1 - F_0)z = z$ , car le coefficient  $F_0$  est nul. En d'autres termes,

$$F(z) - zF(z) - z^2 F(z) = z,$$

ce qui donne la formule

$$F(z) = \frac{z}{1 - z - z^2}. \quad (6.117)$$

Toute l'information sur la suite de Fibonacci se trouve maintenant concentrée dans la simple expression  $z/(1-z-z^2)$ . Bien que cette expression ne nous dise rien a priori, nous avons fait un grand pas. En effet, en factorisant le dénominateur puis en développant en éléments simples, nous pouvons aboutir à une formule qu'il nous sera facile de développer en série. L'expression des coefficients de cette série constituera alors une forme close des nombres de Fibonacci.

Le plan d'attaque que nous venons de concocter sera peut-être plus clair si nous le détaillons en commençant par la fin. Supposons que nous ayons une fonction génératrice plus simple, disons  $1/(1-\alpha z)$  où  $\alpha$  est une constante. Dans ce cas, nous connaissons tous les coefficients des puissances de  $z$  car

$$\frac{1}{1 - \alpha z} = 1 + \alpha z + \alpha^2 z^2 + \alpha^3 z^3 + \cdots.$$

De même, si la fonction génératrice est de la forme  $A/(1-\alpha z) + B/(1-\beta z)$ ,

*"Sit 1 + x + 2xx + 3x<sup>3</sup> + 5x<sup>4</sup> + 8x<sup>5</sup> + 13x<sup>6</sup> + 21x<sup>7</sup> + 34x<sup>8</sup> &c Series nata ex divisione Unitatis per Trinomium 1 - x - xx."*

— A. de Moivre [76]

*"The quantities r, s, t, which show the relation of the terms, are the same as those in the denominator of the fraction. This property, howsoever obvious it may be, M. DeMoivre was the first that applied it to use, in the solution of problems about infinite series, which otherwise would have been very intricate."*

— J. Stirling [343]

les coefficients sont faciles à déterminer du fait que

$$\begin{aligned} \frac{A}{1-\alpha z} + \frac{B}{1-\beta z} &= A \sum_{n \geq 0} (\alpha z)^n + B \sum_{n \geq 0} (\beta z)^n \\ &= \sum_{n \geq 0} (A\alpha^n + B\beta^n)z^n. \end{aligned} \quad (6.118)$$

Par conséquent, il nous suffit de trouver des constantes  $A$ ,  $B$ ,  $\alpha$  et  $\beta$  telles que

$$\frac{A}{1-\alpha z} + \frac{B}{1-\beta z} = \frac{z}{1-z-z^2},$$

pour obtenir une forme close  $A\alpha^n + B\beta^n$  des coefficients  $F_n$  de  $z^n$  dans  $F(z)$ . Le membre gauche peut se réécrire

$$\frac{A}{1-\alpha z} + \frac{B}{1-\beta z} = \frac{A - A\beta z + B - B\alpha z}{(1-\alpha z)(1-\beta z)}.$$

On en déduit que les quatre constantes recherchées sont solutions des deux équations polynomiales suivantes :

$$(1-\alpha z)(1-\beta z) = 1-z-z^2; \quad (6.119)$$

$$(A+B) - (A\beta + B\alpha)z = z. \quad (6.120)$$

Nous devons donc factoriser le dénominateur de  $F(z)$  sous la forme  $(1-\alpha z)(1-\beta z)$ . Alors nous pourrons exprimer  $F(z)$  comme une somme de deux fractions dans laquelle les facteurs  $(1-\alpha z)$  et  $(1-\beta z)$  seront convenablement séparés l'un de l'autre.

Notez que nous avons écrit les facteurs du dénominateur de (6.119) sous la forme  $(1-\alpha z)(1-\beta z)$ , et non sous la forme usuelle  $c(z-\rho_1)(z-\rho_2)$ , où  $\rho_1$  et  $\rho_2$  sont les racines. La raison en est que la forme que nous avons retenue se prête mieux au développement en série.

Il y a plusieurs façons de trouver  $\alpha$  and  $\beta$ . L'une d'elles fait appel à une petite astuce : introduisons une nouvelle variable  $w$  et cherchons la factorisation

$$w^2 - wz - z^2 = (w - \alpha z)(w - \beta z).$$

Celle-ci une fois trouvée, il suffit de poser  $w = 1$  pour obtenir nos facteurs de  $1-z-z^2$ . Voici les racines de  $w^2 - wz - z^2 = 0$  :

$$\frac{z \pm \sqrt{z^2 + 4z^2}}{2} = \frac{1 \pm \sqrt{5}}{2} z.$$

Ainsi,

$$w^2 - wz - z^2 = \left(w - \frac{1 + \sqrt{5}}{2}z\right) \left(w - \frac{1 - \sqrt{5}}{2}z\right)$$

et nous avons nos constantes  $\alpha$  et  $\beta$ .

Le nombre  $(1 + \sqrt{5})/2 \approx 1,61803$  occupe une place importante dans bien des domaines des mathématiques aussi bien que dans le monde des arts, car il est considéré depuis l'antiquité comme le rapport le plus esthétique dans toutes sortes de constructions. A ce titre, il a droit à un nom particulier : c'est le *nombre d'or*. On le désigne par la lettre grecque  $\phi$ , en l'honneur du sculpteur Phidias qui en a, paraît-il, fait consciencieusement usage dans ses œuvres. L'autre racine,  $(1 - \sqrt{5})/2 = -1/\phi \approx -0,61803$  partage bien des propriétés de  $\phi$ , c'est pourquoi elle est désignée par la symbole  $\hat{\phi}$ , "phi chapeau". Comme ces nombres sont racines de l'équation  $w^2 - w - 1 = 0$ , on a

$$\phi^2 = \phi + 1; \quad \hat{\phi}^2 = \hat{\phi} + 1. \quad (6.121)$$

Nous en apprendrons davantage sur  $\phi$  et  $\hat{\phi}$  dans quelque temps.

Nous avons donc trouvé les constantes  $\alpha = \phi$  et  $\beta = \hat{\phi}$  de l'équation (6.119) ; il ne nous reste plus qu'à trouver  $A$  et  $B$  qui satisfont (6.120). Si on pose  $z = 0$  dans cette équation, on trouve  $B = -A$ , et on se ramène à résoudre

$$-\hat{\phi}A + \phi A = 1.$$

La solution est  $A = 1/(\phi - \hat{\phi}) = 1/\sqrt{5}$ . Voici par conséquent le développement en éléments simples de (6.117) :

$$F(z) = \frac{1}{\sqrt{5}} \left( \frac{1}{1 - \phi z} - \frac{1}{1 - \hat{\phi} z} \right). \quad (6.122)$$

Bien. Nous avons notre  $F(z)$  comme nous le voulions. Nous pouvons maintenant faire un développement en série de chacune des fractions pour trouver une forme close du coefficient de  $z^n$  :

$$F_n = \frac{1}{\sqrt{5}} (\phi^n - \hat{\phi}^n). \quad (6.123)$$

C'est Leonhard Euler [113] qui, le premier, publia cette formule en 1765, mais elle fut oubliée jusqu'à ce que Jacques Binet [31] la redécouvrit en 1843.

Avant de nous extasier devant notre résultat, vérifions qu'il est correct. Pour  $n = 0$ , la formule donne bien  $F_0 = 0$ ; pour  $n = 1$ , elle donne  $F_1 = (\phi - \hat{\phi})/\sqrt{5}$ , ce qui vaut exactement 1. Pour les puissances plus élevées, les équations (6.121) montrent que les nombres définis par (6.123) satisfont

*D'après les observations de savants européens, le rapport entre la hauteur d'une personne et la hauteur de son nombril est à peu près égal à 1,618 [136].*

la récurrence de Fibonacci ; on déduit donc par induction que ce sont des nombres de Fibonacci. Pour faire cette vérification, nous aurions aussi pu développer  $\phi^n$  et  $\hat{\phi}^n$  avec la formule du binôme, mais cela aurait été plutôt fastidieux. Le but d'une forme close n'est pas forcément de nous fournir un moyen de calcul rapide, mais plutôt de nous faire connaître les liens d'une quantité donnée (ici  $F_n$ ) avec les autres quantités mathématiques.

Avec un peu d'intuition, nous aurions pu simplement deviner la formule (6.123) et la prouver par induction. Ceci n'empêche pas la méthode des séries génératrices d'être un outil très puissant pour ce genre de problème. Nous verrons au chapitre 7 que cette même méthode peut résoudre des récurrences beaucoup plus difficiles. A propos, nous ne nous sommes pas préoccupés de savoir si les sommes que nous manipulions dans notre calcul de (6.123) convergeaient ou pas. En fait, la plupart des opérations que l'on fait sur ces sommes sont rigoureusement justifiées, qu'elles convergent ou non [182]. Il y a pourtant certainement des lecteurs qui se méfient encore des raisonnements sur des sommes infinies. Qu'ils se rassurent en sachant que (6.123) peut aussi se prouver sans problème par induction.

La formule (6.123) a plusieurs conséquences particulièrement intéressantes. Parmi celles-ci, on trouve le fait que l'entier  $F_n$  est extrêmement proche du nombre irrationnel  $\phi^n/\sqrt{5}$  lorsque  $n$  est grand. En effet, comme  $\hat{\phi}$  est plus petit que 1 en valeur absolue,  $\hat{\phi}^n$  devient exponentiellement petit et son effet est négligeable. Par exemple,  $F_{10} = 55$  et  $F_{11} = 89$  sont très proches de

$$\frac{\phi^{10}}{\sqrt{5}} \approx 55,00364 \quad \text{et} \quad \frac{\phi^{11}}{\sqrt{5}} \approx 88,99775.$$

Cette observation mène à une nouvelle forme close,

$$F_n = \left\lfloor \frac{\phi^n}{\sqrt{5}} + \frac{1}{2} \right\rfloor = \frac{\phi^n}{\sqrt{5}} \quad \text{arrondi à l'entier le plus proche,} \quad (6.124)$$

du fait que  $|\hat{\phi}^n/\sqrt{5}| < \frac{1}{2}$  pour tout  $n \geq 0$ . Lorsque  $n$  est pair,  $F_n$  est un tout petit peu plus petit que  $\phi^n/\sqrt{5}$  ; dans le cas contraire, il est un tout petit peu plus grand.

L'identité de Cassini (6.103) peut se réécrire

$$\frac{F_{n+1}}{F_n} - \frac{F_n}{F_{n-1}} = \frac{(-1)^n}{F_{n-1} F_n}.$$

Lorsque  $n$  est grand,  $1/F_{n-1} F_n$  est très petit, donc  $F_{n+1}/F_n$  doit être proche de  $F_n/F_{n-1}$ . D'après l'expression (6.124), ce rapport est proche de  $\phi$ . On peut en fait prouver, par induction ou directement en utilisant (6.123), que

$$F_{n+1} = \phi F_n + \hat{\phi}^n. \quad (6.125)$$

Le rapport  $F_{n+1}/F_n$  approche  $\phi$  alternativement par valeur supérieure ou par valeur inférieure.

Par coïncidence,  $\phi$  est aussi très proche du nombre de kilomètres dans un mile (le rapport exact est de 1,609344, car 1 pouce vaut exactement 2,54 centimètres). Cela nous donne un moyen pratique de convertir mentalement les kilomètres en miles et réciproquement, car une distance de  $F_{n+1}$  kilomètres est (avec une bonne approximation) égale à  $F_n$  miles.

Supposons que nous voulions convertir en miles un nombre de kilomètres qui n'est pas un nombre de Fibonacci. Combien y a-t-il de miles dans 30km ? Facile : il suffit de convertir mentalement 30 en sa représentation de Fibonacci  $21 + 8 + 1$  par la méthode gloutonne que nous avons vue tout à l'heure, puis de décaler chaque nombre d'une unité vers le bas, pour obtenir  $13 + 5 + 1$  (le "1" d'origine était  $F_2$  car  $k_r \gg 0$  dans (6.113), donc le nouveau "1" est  $F_1$ ). Cette opération de décalage équivaut à peu près à une division par  $\phi$ . Notre estimation est donc de 19 miles, ce qui est plutôt bon, puisque la bonne réponse est 18,64 miles approximativement. Inversement, pour passer des miles aux kilomètres, on décale d'une unité vers le bas : 30 miles font à peu près  $34 + 13 + 2 = 49$  kilomètres (c'est moins bon car la réponse correcte est à peu près 48,28).

En fait, le décalage vers le bas donne le nombre arrondi de miles correct pour  $n$  kilomètres pour tout  $n \leq 100$  sauf  $n = 4, 12, 54, 62, 75, 83, 91, 96$ , ou 99 ; dans ces cas, l'erreur est tout de même inférieure à  $2/3$  de mile. De son côté, le décalage vers le haut donne soit le bon nombre de kilomètres pour  $n$  miles, soit un kilomètre de trop, pour tout  $n \leq 113$ . Le seul cas vraiment gênant est  $n = 4 = 3 + 1$ , où les deux erreurs d'arrondi vont dans le même sens au lieu de s'annuler.

## 6.7 CONTINUANTS

Les nombres de Fibonacci sont liés à l'arbre de Stern-Brocot que nous avons étudié au chapitre 4. Ils peuvent aussi être généralisés pour donner une suite de polynômes qui fut beaucoup étudiée par Euler. Ces polynômes sont appelés *continuants* car ils s'avèrent d'une importance capitale dans l'étude des fractions continues comme

$$a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4}}}}. \quad (6.126)$$

Le continuant est un polynôme  $K_n(x_1, x_2, \dots, x_n)$  à  $n$  variables défini

*Pour pouvoir vérifier le calcul, il faut aussi savoir qu'un mile vaut 1760 yards, qu'un yard vaut 3 pieds et qu'un pied vaut 12 pouces (N.d.T.).*

*Si les Etats-Unis se mettent au système métrique, la limitation de vitesse passera de 55 mi/h à 89 km/h. A moins qu'ils ne nous laissent généreusement rouler à 90.*

*Le décalage vers le bas transforme  $n$  en  $f(n/\phi)$  et le décalage vers le haut transforme  $n$  en  $f(n\phi)$ , avec  $f(x) = \lfloor x + \phi^{-1} \rfloor$ .*

par la récurrence suivante :

$$\begin{aligned} K_0() &= 1; \\ K_1(x_1) &= x_1; \\ K_n(x_1, \dots, x_n) &= K_{n-1}(x_1, \dots, x_{n-1})x_n + K_{n-2}(x_1, \dots, x_{n-2}). \end{aligned} \quad (6.127)$$

Voici par exemple les trois premières valeurs après  $K_1(x_1)$  :

$$\begin{aligned} K_2(x_1, x_2) &= x_1x_2 + 1; \\ K_3(x_1, x_2, x_3) &= x_1x_2x_3 + x_1 + x_3; \\ K_4(x_1, x_2, x_3, x_4) &= x_1x_2x_3x_4 + x_1x_2 + x_1x_4 + x_3x_4 + 1. \end{aligned}$$

Il est facile de montrer par induction que le nombre de termes est un nombre de Fibonacci :

$$K_n(1, 1, \dots, 1) = F_{n+1}. \quad (6.128)$$

Lorsque le nombre de paramètres est clairement défini par le contexte, on peut écrire simplement "K" à la place de " $K_n$ ", exactement comme nous pouvions omettre les nombre de paramètres lorsque nous utilisions les fonctions hypergéométriques du chapitre 5. Par exemple,  $K(x_1, x_2) = K_2(x_1, x_2) = x_1x_2 + 1$ . En revanche, dans des formules du genre de (6.128), l'indice  $n$  est absolument nécessaire.

Euler remarqua que  $K(x_1, x_2, \dots, x_n)$  peut être construit en partant du produit  $x_1x_2 \dots x_n$  et en éliminant de toutes les façons possibles les variables adjacentes  $x_k x_{k+1}$ . On peut représenter graphiquement cette règle en écrivant des suites du "code de Morse". Ce sont des suites de traits et de points de longueur  $n$ , sachant qu'un trait compte pour deux points. Voici toutes les suites de Morse de longueur 4 :

....      ...-      -.-.      -..      --

Chacune d'entre elles correspond à un facteur de  $K(x_1, x_2, x_3, x_4)$ . Un point représente une variable qui appartient au terme, tandis qu'un trait représente un couple de variables adjacentes qui en sont exclues. Par exemple,  $---$  correspond à  $x_1x_4$ .

Si une suite du code de Morse de longueur  $n$  a  $k$  traits, alors elle contient  $n - 2k$  points et  $n - k$  symboles en tout. Ces symboles peuvent être combinés de  $\binom{n-k}{k}$  façons différentes. Par conséquent, si on remplace chaque point par  $z$  et chaque trait par 1, on obtient

$$K_n(z, z, \dots, z) = \sum_{k=0}^n \binom{n-k}{k} z^{n-2k}. \quad (6.129)$$

## 322 NOMBRES REMARQUABLES

Nous savons aussi que le nombre total de termes d'un continuant est un nombre de Fibonacci. Nous en déduisons l'identité

$$F_{n+1} = \sum_{k=0}^n \binom{n-k}{k}. \quad (6.130)$$

Notons en passant que nous avons trouvé en (5.74) une forme close pour (6.129), qui généralise la formule d'Euler-Binet (6.123).

La relation entre les suites du code de Morse et les continuants fait apparaître la symétrie de ceux-ci :

$$K(x_n, \dots, x_2, x_1) = K(x_1, x_2, \dots, x_n). \quad (6.131)$$

Ainsi, ils satisfont, en plus de la récurrence (6.127), une récurrence qui lui est symétrique en quelque sorte :

$$K_n(x_1, \dots, x_n) = x_1 K_{n-1}(x_2, \dots, x_n) + K_{n-2}(x_3, \dots, x_n). \quad (6.132)$$

Ces deux récurrences sont en fait des cas particuliers d'une règle plus générale :

$$\begin{aligned} K_{m+n}(x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n}) \\ = K_m(x_1, \dots, x_m) K_n(x_{m+1}, \dots, x_{m+n}) \\ + K_{m-1}(x_1, \dots, x_{m-1}) K_{n-1}(x_{m+2}, \dots, x_{m+n}). \end{aligned} \quad (6.133)$$

Celle-ci se comprend bien si on a en tête l'analogie avec le code de Morse. Le premier produit  $K_m K_n$  engendre les termes de  $K_{m+n}$  dans lesquels il n'y a pas de trait en position  $[m, m+1]$ , tandis que le second produit engendre tous les autres. Si on pose  $x_k = 1$  pour tout  $1 \leq k \leq m+n$ , cette identité se réduit à  $F_{m+n+1} = F_{m+1} F_{n+1} + F_m F_n$ , et (6.108) devient alors un cas particulier de (6.133).

Euler [112] découvrit que les continuants obéissent à une règle plus remarquable encore, qui est une généralisation de l'identité de Cassini :

$$\begin{aligned} K_{m+n}(x_1, \dots, x_{m+n}) K_k(x_{m+1}, \dots, x_{m+k}) \\ = K_{m+k}(x_1, \dots, x_{m+k}) K_n(x_{m+1}, \dots, x_{m+n}) \\ + (-1)^k K_{m-1}(x_1, \dots, x_{m-1}) K_{n-k-1}(x_{m+k+2}, \dots, x_{m+n}). \end{aligned} \quad (6.134)$$

Cette règle (démontrée dans l'exercice 29) est vraie pourvu que les indices des  $K$  soient tous positifs ou nuls. Par exemple, pour  $k = 2$ ,  $m = 1$  et  $n = 3$ , on a

$$K(x_1, x_2, x_3, x_4) K(x_2, x_3) = K(x_1, x_2, x_3) K(x_2, x_3, x_4) + 1.$$

Les continuants sont étroitement liés à l'algorithme d'Euclide. Supposons par exemple que le calcul de  $\text{pgcd}(m, n)$  s'effectue en quatre étapes :

$$\begin{aligned}
 \text{pgcd}(m, n) & \\
 &= \text{pgcd}(n_0, n_1) & n_0 &= m, \quad n_1 = n; \\
 &= \text{pgcd}(n_1, n_2) & n_2 &= n_0 \bmod n_1 = n_0 - q_1 n_1; \\
 &= \text{pgcd}(n_2, n_3) & n_3 &= n_1 \bmod n_2 = n_1 - q_2 n_2; \\
 &= \text{pgcd}(n_3, n_4) & n_4 &= n_2 \bmod n_3 = n_2 - q_3 n_3; \\
 &= \text{pgcd}(n_4, 0) = n_4. & 0 &= n_3 \bmod n_4 = n_3 - q_4 n_4.
 \end{aligned}$$

Alors on a

$$\begin{aligned}
 n_4 &= n_4 &= K()n_4; \\
 n_3 &= q_4 n_4 &= K(q_4)n_4; \\
 n_2 &= q_3 n_3 + n_4 = K(q_3, q_4)n_4; \\
 n_1 &= q_2 n_2 + n_3 = K(q_2, q_3, q_4)n_4; \\
 n_0 &= q_1 n_1 + n_2 = K(q_1, q_2, q_3, q_4)n_4.
 \end{aligned}$$

Si l'algorithme d'Euclide trouve le plus grand commun diviseur  $d$  en  $k$  étapes après avoir calculé la suite de quotients  $q_1, \dots, q_k$ , alors les nombres de départ étaient  $K(q_1, q_2, \dots, q_k)d$  et  $K(q_2, \dots, q_k)d$ . Ce fait a été observé au début du dix-huitième siècle par Thomas Fantet de Lagny [232], qui semble être le premier à avoir étudié explicitement les continuants. Il remarqua que les nombres de Fibonacci consécutifs apparaissent comme continuants lorsque les  $q$  sont minimaux ; il en déduisit que ces nombres sont les plus petits qui forcent l'algorithme d'Euclide à prendre un nombre donné d'étapes.

Les continuants sont aussi étroitement liés aux fractions continues, dont ils tirent d'ailleurs leur nom. Par exemple,

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}}} = \frac{K(a_0, a_1, a_2, a_3)}{K(a_1, a_2, a_3)}. \quad (6.135)$$

On peut montrer par induction que ceci est valable pour des fractions de n'importe quelle profondeur. Ainsi, en raison de l'identité

$$\begin{aligned}
 K_n(x_1, \dots, x_{n-1}, x_n + y) \\
 &= K_n(x_1, \dots, x_{n-1}, x_n) + K_{n-1}(x_1, \dots, x_{n-1})y,
 \end{aligned} \quad (6.136)$$

qui est démontrée et généralisée dans l'exercice 30, on a

$$\frac{K(a_0, a_1, a_2, a_3 + 1/a_4)}{K(a_1, a_2, a_3 + 1/a_4)} = \frac{K(a_0, a_1, a_2, a_3, a_4)}{K(a_1, a_2, a_3, a_4)}.$$

Il existe aussi un lien étroit entre les continuants et l'arbre de Stern–Brocot présenté au chapitre 4. Nous savons que tout nœud de cet arbre peut être représenté par une suite de L et de R

$$R^{a_0}L^{a_1}R^{a_2}L^{a_3}\dots R^{a_{n-2}}L^{a_{n-1}}, \quad (6.137)$$

où  $a_0 \geq 0$ ,  $a_1 \geq 1$ ,  $a_2 \geq 1$ ,  $a_3 \geq 1$ , ...,  $a_{n-2} \geq 1$ ,  $a_{n-1} \geq 0$ , et  $n$  est pair. Si on utilise la représentation matricielle de L et R (4.33), il n'est pas difficile de prouver par induction que l'équivalent matriciel de (6.137) est

$$\begin{pmatrix} K_{n-2}(a_1, \dots, a_{n-2}) & K_{n-1}(a_1, \dots, a_{n-2}, a_{n-1}) \\ K_{n-1}(a_0, a_1, \dots, a_{n-2}) & K_n(a_0, a_1, \dots, a_{n-2}, a_{n-1}) \end{pmatrix} \quad (6.138)$$

(cette preuve fait d'ailleurs partie de l'exercice 87). Par exemple,

$$R^aL^bR^cL^d = \begin{pmatrix} bc + 1 & bcd + b + d \\ abc + a + c & abcd + ab + ad + cd + 1 \end{pmatrix}.$$

A partir de là, on peut appliquer (4.34) pour écrire une forme close de la fraction de l'arbre de Stern–Brocot représentée par la suite (6.137) :

$$f(R^{a_0} \dots L^{a_{n-1}}) = \frac{K_{n+1}(a_0, a_1, \dots, a_{n-1}, 1)}{K_n(a_1, \dots, a_{n-1}, 1)}. \quad (6.139)$$

Ce résultat constitue le “théorème de Halphen” [174]. Appliquons-le pour trouver la fraction correspondant à LRRL : dans ce cas,  $a_0 = 0$ ,  $a_1 = 1$ ,  $a_2 = 2$ ,  $a_3 = 1$  et  $n = 4$  ; on déduit de (6.139) que

$$\frac{K(0, 1, 2, 1, 1)}{K(1, 2, 1, 1)} = \frac{K(2, 1, 1)}{K(1, 2, 1, 1)} = \frac{K(2, 2)}{K(3, 2)} = \frac{5}{7}.$$

Dans ce calcul, nous avons utilisé l'identité  $K_n(x_1, \dots, x_{n-1}, x_n + 1) = K_{n+1}(x_1, \dots, x_{n-1}, x_n, 1)$  pour absorber les 1. Cette identité se déduit de (6.136) en posant  $y = 1$ .

Une comparaison de (6.135) et (6.139) montre que la fraction correspondant à un nœud (6.137) de l'arbre de Stern–Brocot admet la représentation en fraction continue

$$f(R^{a_0} \dots L^{a_{n-1}}) = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots + \cfrac{1}{a_{n-1} + \cfrac{1}{1}}}}}. \quad (6.140)$$

Ainsi, on peut directement convertir les nœuds de l'arbre de Stern-Brocot en fractions continues. Par exemple,

$$f(LRRL) = 0 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1}}}}.$$

Nous avons vu au chapitre 4 que tout nombre irrationnel détermine un chemin infini dans l'arbre de Stern-Brocot, et donc qu'il peut être représenté par une suite infinie de L et de R. Si la suite infinie d'un irrationnel donné  $\alpha$  est  $R^{a_0}L^{a_1}R^{a_2}L^{a_3}\dots$ , alors il lui correspond une fraction continue infinie

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cfrac{1}{a_5 + \ddots}}}}} \quad (6.141)$$

On peut aussi obtenir cette fraction continue directement : soit  $\alpha_0 = \alpha$  et, pour  $k \geq 0$ , soient

$$a_k = \lfloor \alpha_k \rfloor; \quad \alpha_k = a_k + \cfrac{1}{\alpha_{k+1}}. \quad (6.142)$$

Les  $a_k$  sont appelés les "quotients partiels" de  $\alpha$ . Si  $\alpha$  est un rationnel  $m/n$ , le processus trouve un à un les mêmes quotients que ceux fournis par l'algorithme d'Euclide avant de s'arrêter (lorsque  $\alpha_{k+1} = \infty$ ).

Ou alors, le secret  
est bien gardé.

La constante d'Euler  $\gamma$  est-elle rationnelle ou irrationnelle ? Personne ne le sait. C'est un célèbre problème non résolu à ce jour. On peut s'en faire une idée en recherchant  $\gamma$  dans l'arbre de Stern-Brocot. Si la constante est rationnelle on doit la trouver, sinon on trouvera toutes les approximations rationnelles les plus proches. La fraction continué correspondant à  $\gamma$  commence par les quotients partiels suivants :

$k$	0	1	2	3	4	5	6	7	8
$a_k$	0	1	1	2	1	2	1	4	3

Voici par conséquent les toutes premières lettres de sa représentation de Stern-Brocot : LRLLRLLRLLLLRRRL... ; on n'y voit pas de motif évident. Selon des calculs effectués par Richard Brent [38], si  $\gamma$  est rationnel, son

dénominateur doit comprendre au moins 10000 chiffres (en base 10). C'est pourquoi personne ne pense que  $\gamma$  est rationnel ; mais personne n'a pu prouver le contraire.

Pour terminer ce chapitre, nous allons démontrer une remarquable identité qui est en rapport avec tout ce que nous venons de voir. Nous avons introduit au chapitre 3 la notion de spectre : le spectre d'une constante donnée  $\alpha$  est le multi-ensemble formé des nombres  $\lfloor n\alpha \rfloor$ . On peut donc dire que la série infinie

$$\sum_{n \geq 1} z^{\lfloor n\phi \rfloor} = z + z^3 + z^4 + z^6 + z^8 + z^9 + \dots$$

*La constante  $\gamma$  doit être irrationnelle, si l'on en croit une réflexion d'Einstein (peu connue) : "Dieu ne s'amuse pas à lancer d'énormes dénominateurs dans l'univers".*

est la fonction génératrice du spectre de  $\phi$ , où  $\phi = (1 + \sqrt{5})/2$  est le nombre d'or. L'identité que nous allons démontrer a été découverte en 1976 par J. L. Davison [73]. Il s'agit d'un lien entre la fonction génératrice que nous venons de définir et la suite de Fibonacci, lien exprimé par une fraction continue infinie :

$$\frac{z^{F_1}}{1 + \frac{z^{F_2}}{1 + \frac{z^{F_3}}{1 + \frac{z^{F_4}}{\ddots}}}} = (1 - z) \sum_{n \geq 1} z^{\lfloor n\phi \rfloor}. \quad (6.143)$$

Les deux membres de (6.143) s'avèrent intéressants. Observons tout d'abord les nombres  $\lfloor n\phi \rfloor$ . Si la représentation de Fibonacci (6.113) de  $n$  est  $F_{k_1} + \dots + F_{k_r}$ , celle de  $n\phi$  doit logiquement être proche de  $F_{k_1+1} + \dots + F_{k_r+1}$  (avec le même décalage que pour convertir des miles en kilomètres). Pour être plus précis, selon (6.125), nous pouvons écrire que

$$n\phi = F_{k_1+1} + \dots + F_{k_r+1} - (\phi^{k_1} + \dots + \phi^{k_r}).$$

Or, comme  $\hat{\phi} = -1/\phi$  et  $k_1 \gg \dots \gg k_r \gg 0$ , on a

$$\begin{aligned} |\phi^{k_1} + \dots + \phi^{k_r}| &< \phi^{-k_r} + \phi^{-k_r-2} + \phi^{-k_r-4} + \dots \\ &= \frac{\phi^{-k_r}}{1 - \phi^{-2}} = \phi^{1-k_r} \leq \phi^{-1} < 1. \end{aligned}$$

De plus,  $\hat{\phi}^{k_1} + \dots + \hat{\phi}^{k_r}$  a le même signe que  $(-1)^{k_r}$ , par un argument similaire. Par conséquent,

$$\lfloor n\phi \rfloor = F_{k_1+1} + \dots + F_{k_r+1} - [k_r(n) \text{ est pair}]. \quad (6.144)$$

Convenons qu'un nombre  $n$  est *Fibonacci impair* (ou tout simplement *F-impair*) si le bit le plus à droite de sa représentation de Fibonacci est 1.

Cela revient à dire que  $k_r(n) = 2$ . Dans le cas contraire, nous dirons que  $n$  est *Fibonacci pair* (ou *F-pair*). Voici par exemple les premiers nombres *F-impairs* : 1, 4, 6, 9, 12, 14, 17, 19. Si  $k_r(n)$  est pair, alors, d'après (6.114),  $n - 1$  est *F-pair*. De même, si  $k_r(n)$  est impair, alors  $n - 1$  est *F-impair*. Par conséquent,

$$k_r(n) \text{ est pair} \iff n - 1 \text{ est } F\text{-pair.}$$

De plus, si  $k_r(n)$  est pair, (6.144) entraîne que  $k_r(\lfloor n\phi \rfloor) = 2$ ; si  $k_r(n)$  est impair, (6.144) implique que  $k_r(\lfloor n\phi \rfloor) = k_r(n) + 1$ . Donc  $k_r(\lfloor n\phi \rfloor)$  est toujours pair, et

$$\lfloor n\phi \rfloor - 1 \text{ est toujours } F\text{-pair.}$$

Inversement, si  $m$  est un nombre *F-pair*, on peut trouver un  $n$  tel que  $m + 1 = \lfloor n\phi \rfloor$ . Pour cela, on additionne 1 à  $m$  en *F*-notation, comme nous l'avons vu précédemment. Si aucune retenue n'est nécessaire ce faisant, alors on trouve  $n$  en décalant ( $m + 2$ ) vers la droite, sinon on le trouve en décalant ( $m + 1$ ) vers la droite. Ainsi, la somme du membre droit de (6.143) peut s'écrire

$$\sum_{n \geq 1} z^{\lfloor n\phi \rfloor} = z \sum_{m \geq 0} z^m [m \text{ est } F\text{-pair}]. \quad (6.145)$$

Passons à la fraction du membre gauche. Réécrivons (6.143) de sorte que tous les numérateurs de la fraction continue soient égaux à 1, comme dans (6.141) :

$$\frac{1}{z^{-F_0} + \frac{1}{z^{-F_1} + \frac{1}{z^{-F_2} + \frac{1}{\ddots}}}} = \frac{1-z}{z} \sum_{n \geq 1} z^{\lfloor n\phi \rfloor}. \quad (6.146)$$

Cette transformation est un peu délicate : il faut que le numérateur et le dénominateur de la fraction ayant  $z^{F_n}$  comme numérateur soient divisés par  $z^{F_{n-1}}$ . Si on coupe la nouvelle fraction continue à  $1/z^{-F_n}$ , sa valeur est un rapport de continuants

$$\frac{K_{n+2}(0, z^{-F_0}, z^{-F_1}, \dots, z^{-F_n})}{K_{n+1}(z^{-F_0}, z^{-F_1}, \dots, z^{-F_n})} = \frac{K_n(z^{-F_1}, \dots, z^{-F_n})}{K_{n+1}(z^{-F_0}, z^{-F_1}, \dots, z^{-F_n})},$$

tout comme en (6.135). Voyons d'abord le dénominateur. Si on pose  $Q_n = K_{n+1}(z^{-F_0}, \dots, z^{-F_n})$ , on trouve  $Q_0 = 1$ ,  $Q_1 = 1 + z^{-1}$ ,  $Q_2 = 1 + z^{-1} + z^{-2}$ ,

$Q_3 = 1 + z^{-1} + z^{-2} + z^{-3} + z^{-4}$  : tout s'arrange bien pour donner la gentille série géométrique

$$Q_n = 1 + z^{-1} + z^{-2} + \cdots + z^{-(F_{n+2}-1)}.$$

Le numérateur correspondant est  $P_n = K_n(z^{-F_1}, \dots, z^{-F_n})$ . Il ressemble à  $Q_n$ , mais il y manque des termes. Ainsi, par exemple,

$$P_5 = z^{-1} + z^{-2} + z^{-4} + z^{-5} + z^{-7} + z^{-9} + z^{-10} + z^{-12},$$

à comparer à  $Q_5 = 1 + z^{-1} + \cdots + z^{-12}$ . En y regardant de plus près, on devine la règle qui régit les termes de  $P_5$  :

$$P_5 = \frac{1+z^2+z^3+z^5+z^7+z^8+z^{10}+z^{11}}{z^{12}} = z^{-12} \sum_{m=0}^{12} z^m [m \text{ est } F\text{-pair}].$$

Nous pouvons la généraliser en prouvant par induction que

$$P_n = z^{1-F_{n+2}} \sum_{m=0}^{F_{n+2}-1} z^m [m \text{ est } F\text{-pair}].$$

Par conséquent,

$$\frac{P_n}{Q_n} = \frac{\sum_{m=0}^{F_{n+2}-1} z^m [m \text{ est } F\text{-pair}]}{\sum_{m=0}^{F_{n+2}-1} z^m}.$$

Il ne reste plus qu'à prendre la limite lorsque  $n \rightarrow \infty$  et qu'à appliquer (6.145) pour aboutir à (6.146).

## Exercices

### Echauffements

- Quelles sont les  $\left[\begin{smallmatrix} 4 \\ 2 \end{smallmatrix}\right] = 11$  permutations de  $\{1, 2, 3, 4\}$  qui contiennent exactement deux cycles ? On désire leur représentation non cyclique, comme 2314, plutôt que leur représentation en cycles donnée en (6.4).
- Le nombre d'applications d'un ensemble à  $n$  éléments dans un ensemble à  $m$  éléments est égal à  $m^n$ . Combien d'entre elles prennent exactement  $k$  valeurs différentes ?
- Lorsqu'on empile des cartes comme dans la section 6.3, l'expérience montre qu'il est sage de laisser un peu de jeu pour que le tas ne bascule pas au moindre souffle. Convenons donc que le centre de gravité des  $k$  cartes du haut doit se trouver à une distance d'au moins  $\epsilon$  du bord de la  $k+1$ ème carte (donc, la carte du haut, par exemple, ne peut

recouvrir la seconde que de  $1 - \epsilon$ ). En supposant qu'on dispose de suffisamment de cartes, est-il possible d'éloigner autant qu'on veut le haut du tas du bord de la table ?

- Si les nombres harmoniques sont les "nombres du ver", les nombres de Fibonacci sont les "nombres du lapin".*
- 4 Exprimez  $1/1 + 1/3 + \dots + 1/(2n+1)$  en fonction de nombres harmoniques.
  - 5 Expliquez comment obtenir la récurrence (6.75) à partir de la définition de  $U_n(x, y)$  en (6.74), puis résolvez la récurrence.
  - 6 Un explorateur a laissé un couple de lapins sur une île. Si on suppose que les bébés lapins deviennent adultes en un mois et que chaque couple de lapins adultes engendre un couple de bébés lapins chaque mois, combien y aura-t-il de couples de lapins après  $n$  mois ? (Après deux mois, il y a deux couples de lapins, dont l'un vient de naître). Trouvez un lien entre ce problème et celui de l'arbre généalogique des abeilles.
  - 7 Montrez que l'identité de Cassini (6.103) est un cas particulier de (6.108), et aussi de (6.134).
  - 8 Utilisez le système de numération de Fibonacci pour convertir 65 miles par heure en un nombre approché de km/h.
  - 9 Combien y a-t-il approximativement de kilomètres carrés dans 8 miles carrés ?
  - 10 Donnez la représentation en fonction continue de  $\phi$ .

### Exercices de base

- 11 Que vaut  $\sum_k (-1)^k \binom{n}{k}$ , la somme alternée d'une ligne du triangle de Stirling de première espèce, lorsque  $n$  est un entier positif ou nul ?
- 12 Montrez que les nombres de Stirling obéissent à une loi d'inversion analogue à (5.48) :

$$g(n) = \sum_k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} (-1)^k f(k) \iff f(n) = \sum_k \left[ \begin{matrix} n \\ k \end{matrix} \right] (-1)^k g(k).$$

- 13 Les opérateurs différentiels  $D = \frac{d}{dz}$  et  $\vartheta = zD$  sont mentionnés dans les chapitres 2 et 5. On a

$$\vartheta^2 = z^2 D^2 + zD,$$

car  $\vartheta^2 f(z) = \vartheta z f'(z) = z \frac{d}{dz} z f'(z) = z^2 f''(z) + z f'(z) = (z^2 D^2 + zD)f(z)$ . De même, on peut prouver que  $\vartheta^3 = z^3 D^3 + 3z^2 D^2 + zD$ . Démontrez les formules générales

$$\vartheta^n = \sum_k \left\{ \begin{matrix} n \\ k \end{matrix} \right\} z^k D^k,$$

$$z^n D^n = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} (-1)^{n-k} \vartheta^k,$$

pour tout  $n \geq 0$ . Ces formules peuvent être utilisées pour effectuer des conversions entre des expressions différentielles de la forme  $\sum_k \alpha_k z^k f^{(k)}(z)$  et  $\sum_k \beta_k \vartheta^k f(z)$ , comme en (5.109).

- 14 Démontrez l'identité (6.37) sur les nombres eulériens.
- 15 Prouvez l'identité (6.39) en considérant la différence mième de (6.37).
- 16 Trouvez la solution de la double récurrence

$$\begin{aligned} A_{n,0} &= a_n \quad [n \geq 0]; \quad A_{0,k} = 0, \quad \text{si } k > 0; \\ A_{n,k} &= kA_{n-1,k} + A_{n-1,k-1}, \quad k, n \text{ entiers}, \end{aligned}$$

lorsque  $k$  et  $n$  parcourent tout l'ensemble des entiers.

- 17 Résolvez les récurrences suivantes en supposant que  $\begin{vmatrix} n \\ k \end{vmatrix}$  est nul lorsque  $n < 0$  ou  $k < 0$ :

$$\begin{aligned} \mathbf{a} \quad \begin{vmatrix} n \\ k \end{vmatrix} &= \begin{vmatrix} n-1 \\ k \end{vmatrix} + n \begin{vmatrix} n-1 \\ k-1 \end{vmatrix} + [n=k=0], \quad \text{pour } n, k \geq 0. \\ \mathbf{b} \quad \begin{vmatrix} n \\ k \end{vmatrix} &= (n-k) \begin{vmatrix} n-1 \\ k \end{vmatrix} + \begin{vmatrix} n-1 \\ k-1 \end{vmatrix} + [n=k=0], \quad \text{pour } n, k \geq 0. \\ \mathbf{c} \quad \begin{vmatrix} n \\ k \end{vmatrix} &= k \begin{vmatrix} n-1 \\ k \end{vmatrix} + k \begin{vmatrix} n-1 \\ k-1 \end{vmatrix} + [n=k=0], \quad \text{pour } n, k \geq 0. \end{aligned}$$

- 18 Montrez que les polynômes de Stirling satisfont

$$(x+1) \sigma_n(x+1) = (x-n) \sigma_n(x) + x \sigma_{n-1}(x).$$

- 19 Montrez que les nombres de Stirling généralisés satisfont

$$\sum_{k=0}^n \left\{ \begin{matrix} x+k \\ x \end{matrix} \right\} \left[ \begin{matrix} x \\ x-n+k \end{matrix} \right] (-1)^k / \binom{x+k}{n+1} = 0, \quad n > 0 \text{ entier.}$$

$$\sum_{k=0}^n \left[ \begin{matrix} x+k \\ x \end{matrix} \right] \left\{ \begin{matrix} x \\ x-n+k \end{matrix} \right\} (-1)^k / \binom{x+k}{n+1} = 0, \quad n > 0 \text{ entier.}$$

- 20 Trouvez une forme close pour  $\sum_{k=1}^n H_k^{(2)}$ .
- 21 Montrez que si  $H_n = a_n/b_n$  où  $a_n$  et  $b_n$  sont des entiers, alors le dénominateur  $b_n$  est un multiple de  $2^{\lfloor \lg n \rfloor}$ . *Suggestion*: considérez le nombre  $2^{\lfloor \lg n \rfloor - 1} H_n - \frac{1}{2}$ .

**22** Prouvez que la somme infinie

$$\sum_{k \geq 1} \left( \frac{1}{k} - \frac{1}{k+z} \right)$$

converge pour tout nombre complexe  $z$  sauf si  $z$  est un entier strictement négatif, et montrez qu'elle est égale à  $H_z$  lorsque  $z$  est un entier positif ou nul (on peut donc utiliser cette formule pour définir les nombres harmoniques  $H_z$  pour  $z$  complexe).

- 23** L'équation (6.81) donne les coefficients du développement en série de  $z/(e^z - 1)$ . Quels sont les coefficients du développement en série de  $z/(e^z + 1)$ ? *Suggestion*: considérez l'identité  $(e^z + 1)(e^z - 1) = e^{2z} - 1$ .
- 24** Montrez que le nombre tangent  $T_{2n+1}$  est un multiple de  $2^n$ . *Suggestion*: montrez que tous les coefficients de  $T_{2n}(x)$  et  $T_{2n+1}(x)$  sont des multiples de  $2^n$ .
- 25** L'équation (6.57) prouve que le ver finira par atteindre le bout de l'élastique au bout d'un certain nombre  $N$  de minutes. Il y a donc forcément un nombre  $n$  tel que, pour la première fois, le ver est plus proche du but après  $n$  minutes qu'après  $n-1$  minutes. Montrez que  $n < \frac{1}{2}N$ .
- 26** Utilisez la sommation par parties pour calculer  $S_n = \sum_{k=1}^n H_k/k$ . *Suggestion*: considérez aussi la somme  $\sum_{k=1}^n H_{k-1}/k$ .
- 27** Prouvez la règle (6.111) concernant le pgcd des nombres de Fibonacci.
- 28** Le *nombre de Lucas*  $L_n$  est, par définition, égal à  $F_{n+1} + F_{n-1}$ . On déduit donc de (6.109) que  $F_{2n} = F_n L_n$ . Voici les premiers nombres de Lucas :

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$L_n$	2	1	3	4	7	11	18	29	47	76	123	199	322	521

- a Avec la méthode du répertoire, montrez qu'on peut exprimer la solution de la récurrence

$$Q_0 = \alpha; \quad Q_1 = \beta; \quad Q_n = Q_{n-1} + Q_{n-2}, \quad n > 1$$

en fonction de  $F_n$  et  $L_n$ .

- b Trouvez une forme close de  $L_n$  en fonction de  $\phi$  et  $\bar{\phi}$ .

- 29** Prouvez l'identité d'Euler (6.134) concernant les continuants.
- 30** Généralisez (6.136) pour trouver une expression du continuant "incrémenté"  $K(x_1, \dots, x_{m-1}, x_m + y, x_{m+1}, \dots, x_n)$ , pour  $1 \leq m \leq n$ .

**Devoirs à la maison**

- 31 Trouvez une forme close pour les coefficients  $\left| \begin{smallmatrix} n \\ k \end{smallmatrix} \right|$  dans cette représentation des puissances montantes par des puissances descendantes :

$$x^n = \sum_k \left| \begin{smallmatrix} n \\ k \end{smallmatrix} \right| x^k, \quad n \geq 0 \text{ entier.}$$

(Par exemple,  $x^4 = x^4 + 12x^3 + 36x^2 + 24x^1$ , donc  $\left| \begin{smallmatrix} 4 \\ 2 \end{smallmatrix} \right| = 36$ ).

- 32 Au chapitre 5, nous avons trouvé les formules

$$\sum_{k \leq m} \binom{n+k}{k} = \binom{n+m+1}{m} \quad \text{et} \quad \sum_{0 \leq k \leq m} \binom{k}{n} = \binom{m+1}{n+1}$$

en développant la récurrence  $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$  de deux manières différentes. Quelles identités obtient-on si on développe la récurrence analogue  $\{ \binom{n}{k} \} = k \{ \binom{n-1}{k} \} + \{ \binom{n-1}{k-1} \}$  ?

- 33 La table 280 donne les valeurs de  $\left[ \begin{smallmatrix} n \\ 2 \end{smallmatrix} \right]$  et  $\left\{ \begin{smallmatrix} n \\ 2 \end{smallmatrix} \right\}$ . Trouvez des formes closes (sans nombres de Stirling) pour les deux cas suivants,  $\left[ \begin{smallmatrix} n \\ 3 \end{smallmatrix} \right]$  et  $\left\{ \begin{smallmatrix} n \\ 3 \end{smallmatrix} \right\}$ .
- 34 Que valent  $\langle \begin{smallmatrix} -1 \\ k \end{smallmatrix} \rangle$  et  $\langle \begin{smallmatrix} -2 \\ k \end{smallmatrix} \rangle$  si on suppose que la récurrence de base (6.35) est valable pour tous entiers  $k$  et  $n$  et que  $\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \rangle = 0$  pour tout  $k < 0$ ?
- 35 Montrez que, pour tout  $\epsilon > 0$ , il existe un entier  $n > 1$  (dépendant de  $\epsilon$ ) tel que  $H_n \bmod 1 < \epsilon$ .
- 36 Est-il possible d'empiler  $n$  briques de façon que la brique du haut ne recouvre aucune partie de la brique du bas et qu'une personne pesant l'équivalent de 100 briques puisse se tenir au milieu de la brique du haut sans faire écrouler la construction ?
- 37 Exprimez  $\sum_{k=1}^{mn} (k \bmod m)/k(k+1)$  en fonction de nombres harmoniques, en supposant que  $m$  et  $n$  sont des entiers strictement positifs. Que vaut la limite lorsque  $n \rightarrow \infty$  ?
- 38 Calculez la somme indéfinie  $\sum \binom{r}{k} (-1)^k H_k \delta k$ .
- 39 Exprimez  $\sum_{k=1}^n H_k^2$  en fonction de  $n$  et  $H_n$ .
- 40 Montrez que 1979 divise le numérateur de  $\sum_{k=1}^{1319} (-1)^{k-1}/k$ , et trouvez un résultat similaire pour 1987. *Suggestion* : utilisez l'astuce de Gauss pour obtenir une somme de fractions dont les numérateurs sont égaux à 1979. Voyez aussi l'exercice 4.
- 41 Calculez la somme

$$\sum_k \binom{\lfloor (n+k)/2 \rfloor}{k}$$

lorsque  $n$  est un entier (éventuellement négatif).

Tiens ! Ces années sont premières.

- 42 Si  $S$  est un ensemble d'entiers, soit  $S + 1$  l'ensemble "incrémenté"  $\{x + 1 \mid x \in S\}$ . Combien de sous-ensembles de  $\{1, 2, \dots, n\}$  possèdent la propriété suivante :  $S \cup (S + 1) = \{1, 2, \dots, n + 1\}$  ?

- 43 Montrez que la somme infinie

$$\begin{aligned} & 0,1 \\ & + 0,01 \\ & + 0,002 \\ & + 0,0003 \\ & + 0,00005 \\ & + 0,000008 \\ & + 0,0000013 \\ & \vdots \end{aligned}$$

converge vers un nombre rationnel.

- 44 Prouvez la réciproque de l'identité de Cassini (6.106) : si  $k$  et  $m$  sont des entiers tels que  $|m^2 - km - k^2| = 1$ , alors il existe un entier  $n$  tel que  $k = \pm F_n$  et  $m = \pm F_{n+1}$ .

- 45 Utilisez la méthode du répertoire pour résoudre la récurrence

$$X_0 = \alpha; \quad X_1 = \beta; \quad X_n = X_{n-1} + X_{n-2} + \gamma n + \delta.$$

- 46 Que valent  $\cos 36^\circ$  et  $\cos 72^\circ$ ?

- 47 Montrez que

$$2^{n-1}F_n = \sum_k \binom{n}{2k+1} 5^k$$

et déduisez de cette identité les valeurs de  $F_p \bmod p$  et de  $F_{p+1} \bmod p$  lorsque  $p$  est premier.

- 48 Montrez qu'on peut supprimer les paramètres nuls d'un polynôme continuant en soudant leurs plus proches voisins :

$$\begin{aligned} K_n(x_1, \dots, x_{m-1}, 0, x_{m+1}, \dots, x_n) \\ = K_{n-2}(x_1, \dots, x_{m-2}, x_{m-1} + x_{m+1}, x_{m+2}, \dots, x_n), \quad 1 < m < n. \end{aligned}$$

- 49 Trouvez la représentation en fraction continue de  $\sum_{n \geq 1} 2^{-\lfloor n\phi \rfloor}$ .

- 50 Soit  $f(n)$  définie pour tout entier strictement positif par la récurrence

$$\begin{aligned} f(1) &= 1; \\ f(2n) &= f(n); \\ f(2n+1) &= f(n) + f(n+1). \end{aligned}$$

- a Dans quels cas  $f(n)$  est-il pair ?  
 b Montrez que  $f(n)$  peut s'écrire avec des continuants.

### Problèmes d'examen

- 51 Soit  $p$  un nombre premier
- Montrez que  $\left\{ \frac{p}{k} \right\} \equiv \left[ \frac{p}{k} \right] \equiv 0 \pmod{p}$ , pour  $1 < k < p$ .
  - Montrez que  $\left[ \frac{p-1}{k} \right] \equiv 1 \pmod{p}$ , pour  $1 \leq k < p$ .
  - Montrez que  $\left\{ \frac{2p-2}{p} \right\} \equiv \left[ \frac{2p-2}{p} \right] \equiv 0 \pmod{p}$  si  $p > 2$ .
  - Montrez que si  $p > 3$ , alors  $\left[ \frac{p}{2} \right] \equiv 0 \pmod{p^2}$ . *Suggestion :* considérez  $p^{\underline{p}}$ .
- 52 Soit  $a_n/b_n$  la fraction réduite égale à  $H_n$ .
- Démontrez que si  $p$  est premier, alors  $p \nmid b_n \iff p \nmid a_{\lfloor n/p \rfloor}$ .
  - Trouvez tous les  $n > 0$  tels que  $a_n$  est divisible par 5.
- 53 Trouvez une forme close pour  $\sum_{k=0}^m \binom{n}{k}^{-1} (-1)^k H_k$ , lorsque  $0 \leq m \leq n$ . *Suggestion :* Il y a la même somme, sans le facteur  $H_k$ , dans l'exercice 5.42.
- 54 Soit  $n > 0$ . Le but de cet exercice est de montrer que le dénominateur de  $B_{2n}$  est le produit de tous les nombres premiers  $p$  tels que  $(p-1) \nmid (2n)$ .
- Montrez que  $S_m(p) + [(p-1) \nmid m]$  est un multiple de  $p$  si  $p$  est premier et  $m > 0$ .
  - Utilisez le résultat de la partie (a) pour montrer que

$$B_{2n} + \sum_{p \text{ prime}} \frac{[(p-1) \nmid (2n)]}{p} = I_{2n} \text{ est un entier.}$$

*Suggestion :* il suffit de prouver que si  $p$  est un nombre premier, alors le dénominateur de la fraction  $B_{2n} + [(p-1) \nmid (2n)]/p$  n'est pas divisible par  $p$ .

- Prouvez que le dénominateur de  $B_{2n}$  est toujours un multiple impair de 6, et qu'il est égal à 6 pour une infinité de  $n$ .
- Montrez que (6.70) est un corollaire d'une identité plus générale, en sommant

$$\sum_{0 \leq k < n} \binom{k}{m} \binom{x+k}{k}$$

et en dérivant par rapport à  $x$ .

- 56 Trouvez une forme close de  $\sum_{k \neq m} \binom{n}{k} (-1)^k k^{n+1}/(k-m)$ , en tant que fonction des entiers  $m$  et  $n$  (la somme porte sur tous les entiers  $k$  sauf la valeur  $k = m$ ).

- 57 On définit les “coefficients binomiaux enroulés” d’ordre 5 par

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{(k-1) \bmod 5}, \quad n > 0,$$

et  $\binom{0}{k} = [k=0]$ . Soit  $Q_n$  la différence entre le plus grand et le plus petit de ces nombres dans la ligne  $n$  :

$$Q_n = \max_{0 \leq k \leq 5} \binom{n}{k} - \min_{0 \leq k \leq 5} \binom{n}{k}.$$

Trouvez et prouvez une relation entre  $Q_n$  et les nombres de Fibonacci.

- 58 Trouvez des formes closes de  $\sum_{n \geq 0} F_n^2 z^n$  et  $\sum_{n \geq 0} F_n^3 z^n$ . Que pouvez-vous en déduire sur  $F_{n+1}^3 - 4F_n^3 - F_{n-1}^3$  ?
- 59 Montrez que si  $m$  et  $n$  sont des entiers strictement positifs, alors il existe un entier  $x$  tel que  $F_x \equiv m \pmod{3^n}$ .
- 60 Trouvez tous les entiers strictement positifs  $n$  tels que  $F_n + 1$  ou  $F_n - 1$  soit un nombre premier.
- 61 Prouvez l’identité

$$\sum_{k=0}^n \frac{1}{F_{2^k}} = 3 - \frac{F_{2^n-1}}{F_{2^n}}, \quad n \geq 1 \text{ entier.}$$

Que vaut  $\sum_{k=0}^n 1/F_{3 \cdot 2^k}$  ?

- 62 Soit  $A_n = \phi^n + \phi^{-n}$  et  $B_n = \phi^n - \phi^{-n}$ .
- Trouvez des constantes  $\alpha$  et  $\beta$  telles que  $A_n = \alpha A_{n-1} + \beta A_{n-2}$  et  $B_n = \alpha B_{n-1} + \beta B_{n-2}$  pour tout  $n \geq 0$ .
  - Exprimez  $A_n$  et  $B_n$  en fonction de  $F_n$  et  $L_n$  (voir l’exercice 28).
  - Prouvez que  $\sum_{k=1}^n 1/(F_{2k+1} + 1) = B_n/A_{n+1}$ .
  - Trouvez une forme close pour  $\sum_{k=1}^n 1/(F_{2k+1} - 1)$ .

#### Questions subsidiaires

- 63 Combien existe-t-il de permutations  $\pi_1 \pi_2 \dots \pi_n$  de  $\{1, 2, \dots, n\}$  qui ont exactement  $k$  indices  $j$  tels que
- $\pi_i < \pi_j$  pour tout  $i < j$  ? (Ces indices  $j$  sont appelés “saillants supérieurs gauches”).
  - $\pi_j > j$  ? (Ces indices  $j$  sont appelés “excédents”).
- 64 Quel est le dénominateur de  $[\frac{1}{1/2-n}]$  lorsque cette fraction est réduite ?
- 65 Prouvez l’identité

$$\int_0^1 \dots \int_0^1 f(\lfloor x_1 + \dots + x_n \rfloor) dx_1 \dots dx_n = \sum_k \binom{n}{k} \frac{f(k)}{n!}.$$

- 66 Que vaut  $\sum_k (-1)^k \binom{n}{k}$ , la somme alternée de la  $n$ ième ligne du triangle d’Euler ?

67 Montrez que

$$\sum_k \binom{n+1}{k+1} \binom{n-k}{m-k} (-1)^{m-k} k! = \binom{n}{m}.$$

68 Montrez que  $\langle\langle \frac{n}{1} \rangle\rangle = 2\langle\langle \frac{n}{1} \rangle\rangle$  et trouvez une forme close pour  $\langle\langle \frac{n}{2} \rangle\rangle$ .

69 Trouvez une forme close pour  $\sum_{k=1}^n k^2 H_{n+k}$ .

70 Montrez que si on développe en série les nombres harmoniques complexes de l'exercice 22, on obtient  $H_z = \sum_{n \geq 2} (-1)^n H_\infty^{(n)} z^{n-1}$ .

71 Montrez que la factorielle généralisée de l'équation (5.83) peut s'écrire

$$\prod_{k \geq 1} \left(1 + \frac{z}{k}\right) e^{-z/k} = \frac{e^{-\gamma z}}{z!},$$

en considérant la limite lorsque  $n \rightarrow \infty$  des  $n$  premiers facteurs de ce produit infini. Montrez que  $\frac{d}{dz}(z!)$  est lié aux nombres harmoniques généraux de l'exercice 22.

72 Montrez que la fonction tangente se développe en série comme indiqué en (6.92). Donnez les développements en séries de  $z/\sin z$  et  $\ln((\tan z)/z)$ .

73 Montrez que  $z \cot z$  est égal à

$$\frac{z}{2^n} \cot \frac{z}{2^n} - \frac{z}{2^n} \tan \frac{z}{2^n} + \sum_{k=1}^{2^{n-1}-1} \frac{z}{2^n} \left( \cot \frac{z+k\pi}{2^n} + \cot \frac{z-k\pi}{2^n} \right),$$

pour tout entier  $n \geq 1$ , et que la limite du  $k$ ième terme général vaut  $2z^2/(z^2 - k^2\pi^2)$ , pour  $k$  fixé lorsque  $n \rightarrow \infty$ .

74 Trouvez une relation entre les nombres  $T_n(1)$  et les coefficients de  $1/\cos z$ .

75 Montrez que les nombres tangents et les coefficients de  $1/\cos z$  apparaissent sur les côtés du triangle infini qui commence comme ceci :

			1			
		0	1			
	1	1	0			
	0	1	2	2		
5	5	4	2	0		
0	5	10	14	16	16	0
61	61	56	46	32	16	0

Chaque ligne contient des sommes partielles de la ligne précédente, allant alternativement de gauche à droite et de droite à gauche. *Suggestion* : considérez les coefficients de la série  $(\sin z + \cos z)/\cos(w+z)$ .

76 Trouvez une forme close pour la somme

$$\sum_k (-1)^k \binom{n}{k} 2^{n-k} k!$$

et montrez qu'elle est nulle lorsque  $n$  est pair.

77 Lorsque  $m$  et  $n$  sont des entiers et  $n \geq 0$ , la valeur de  $\sigma_n(m)$  est donnée par (6.50) si  $m < 0$ , par (6.51) si  $m > n$ , ou par (6.101) si  $m = 0$ . Montrez que, pour tous les autres cas,

$$\sigma_n(m) = \frac{(-1)^{m+n-1}}{m!(n-m)!} \sum_{k=0}^{m-1} \begin{bmatrix} m \\ m-k \end{bmatrix} \frac{B_{n-k}}{n-k}, \quad n \geq m > 0 \text{ entier.}$$

78 Prouvez ce résultat qui relie les nombres de Stirling, les nombres de Bernoulli et les nombres de Catalan :

$$\sum_{k=0}^n \binom{n+k}{k} \binom{2n}{n+k} \frac{(-1)^k}{k+1} = B_n \binom{2n}{n} \frac{1}{n+1}.$$

79 Montrez qu'on peut aussi assembler les quatre parts de l'échiquier du paradoxe  $64 = 65$  pour démontrer que  $64 = 63$ .

80 Soit une suite, définie par la récurrence  $A_1 = x$ ,  $A_2 = y$ , et  $A_n = A_{n-1} + A_{n-2}$ , et telle qu'il existe un entier  $m$  tel que  $A_m = 1000000$ . Quelles sont les valeurs des entiers strictement positifs  $x$  et  $y$  pour lesquelles ce  $m$  est le plus grand possible ?

81 On décrit dans le chapitre une façon de transformer une formule où apparaissent des  $F_{n\pm k}$  en une formule où n'apparaissent que des  $F_n$  et des  $F_{n+1}$ . Il est donc naturel de se demander si deux formules ainsi "réduites" peuvent être égales sans être de forme identique. Soit  $P(x, y)$  un polynôme en  $x$  et  $y$  à coefficients entiers. Trouvez une condition nécessaire et suffisante pour que  $P(F_{n+1}, F_n) = 0$  pour tout  $n \geq 0$ .

82 Expliquez comment additionner deux entiers en travaillant exclusivement dans le système de numération de Fibonacci.

83 Est-il possible qu'une suite  $(A_n)$  qui satisfait la récurrence de Fibonacci  $A_n = A_{n-1} + A_{n-2}$  ne contienne aucun nombre premier, si  $A_0$  et  $A_1$  sont premiers entre eux ?

84 Soient  $m$  et  $n$  deux entiers strictement positifs impairs. Trouvez des formes closes pour

$$S_{m,n}^+ = \sum_{k \geq 0} \frac{1}{F_{2mk+n} + F_m}; \quad S_{m,n}^- = \sum_{k \geq 0} \frac{1}{F_{2mk+n} - F_m}.$$

*Suggestion :* Les sommes de l'exercice 62 sont égales à  $S_{1,3}^+ - S_{1,2n+3}^+$  et  $S_{1,3}^- - S_{1,2n+3}^-$ .

- 85 Caractérissez tous les  $N$  tels que les restes de Fibonacci  $F_n \bmod N$  pour  $n \geq 0$  forment l'ensemble  $\{0, 1, \dots, N - 1\}$  (voir l'exercice 59).
- 86 Soit  $C_1, C_2, \dots$  une suite d'entiers non nuls tels que

$$\text{pgcd}(C_m, C_n) = C_{\text{pgcd}(m, n)}$$

pour tous entiers strictement positifs  $m$  et  $n$ . Prouvez que les coefficients binomiaux généralisés

$$\binom{n}{k}_c = \frac{C_n C_{n-1} \dots C_{n-k+1}}{C_k C_{k-1} \dots C_1}$$

sont tous entiers. (En particulier, d'après (6.111), les "coefficients Fibonomiaux", définis de cette façon à partir des nombres de Fibonacci, sont entiers).

- 87 Montrez que les polynômes continuants apparaissent dans le produit matriciel

$$\begin{pmatrix} 0 & 1 \\ 1 & x_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & x_2 \end{pmatrix} \dots \begin{pmatrix} 0 & 1 \\ 1 & x_n \end{pmatrix}$$

ainsi que dans le déterminant

$$\det \begin{pmatrix} x_1 & 1 & 0 & 0 & \dots & 0 \\ -1 & x_2 & 1 & 0 & & 0 \\ 0 & -1 & x_3 & 1 & & \vdots \\ \vdots & & -1 & & & \ddots \\ & & & & & 1 \\ 0 & 0 & \dots & -1 & x_n & \end{pmatrix}.$$

- 88 En généralisant (6.146), trouvez une fraction continue correspondant à la fonction génératrice  $\sum_{n \geq 1} z^{\lfloor n\alpha \rfloor}$ , où  $\alpha$  est un nombre irrationnel strictement positif quelconque.
- 89 Soit  $\alpha$  un nombre irrationnel de  $(0..1)$  et soient  $a_1, a_2, a_3, \dots$  les quotients partiels de sa représentation en fraction continue. Montrez que  $|D(\alpha, n)| < 2$  si  $n = K(a_1, \dots, a_m)$ , où  $D$  désigne la discrépance définie au chapitre 3.
- 90 Soit  $Q_n$  le plus grand dénominateur du niveau  $n$  de l'arbre de Stern-Brocot (donc  $\langle Q_0, Q_1, Q_2, Q_3, Q_4, \dots \rangle = \langle 1, 2, 3, 5, 8, \dots \rangle$  selon le diagramme du chapitre 4). Montrez que  $Q_n = F_{n+2}$ .

### Problèmes de recherche

- 91 Quelle est la meilleure manière d'étendre la définition de  $\binom{n}{k}$  à toutes les valeurs réelles de  $n$  et  $k$  ?

- 92 Soit  $a_n/b_n$  la représentation de  $H_n$  sous forme de fraction réduite, comme dans l'exercice 52.

- a Existe-t-il une infinité de  $n$  tels que  $p \nmid a_n$ , pour un entier  $p$  donné ?  
 b Existe-t-il une infinité de  $n$  tels que  $b_n = \text{ppcm}(1, 2, \dots, n)$  ( $n = 250$  et  $n = 1000$  sont deux telles valeurs) ?

- 93 Montrez que  $\gamma$  et  $e^\gamma$  sont irrationnels.

- 94 Développez une théorie générale des solutions de la récurrence à deux paramètres

$$\begin{aligned} \binom{n}{k} &= (\alpha n + \beta k + \gamma) \binom{n-1}{k} \\ &\quad + (\alpha' n + \beta' k + \gamma') \binom{n-1}{k-1} + [n=k=0], \quad \text{pour } n, k \geq 0, \end{aligned}$$

en supposant que  $\binom{n}{k} = 0$  lorsque  $n < 0$  ou  $k < 0$ . (Les coefficients binomiaux, les nombres de Stirling, les nombres eulériens et les suites des exercices 17 et 31 obéissent à des cas particuliers de cette récurrence). Quelles sont les valeurs particulières de  $(\alpha, \beta, \gamma, \alpha', \beta', \gamma')$  qui donnent lieu à des "solutions fondamentales" permettant d'exprimer la solution générale ?

- 95 Trouvez un moyen efficace d'étendre l'algorithme de Gosper-Zeilberger aux termes où peuvent apparaître des nombres de Stirling.

# 7

## Fonctions génératrices

Utiliser les fonctions génératrices (ou “fg”) est le moyen le plus puissant que l’on connaisse pour manipuler des suites infinies de nombres. Jusqu’ici, nous avons vu beaucoup de suites et peu de fonctions génératrices. Il est temps maintenant d’étudier ces dernières en profondeur pour découvrir leurs indéniables avantages.

### 7.1 THÉORIE DES DOMINOS ET MONNAIE

Nous allons commencer par une approche en douceur, en jouant un peu pour nous familiariser avec les fonctions génératrices, pour faire leur connaissance plus intuitivement que formellement. Nous allons donc parler d’abord de dominos et de pièces de monnaie.

Quel est le nombre  $T_n$  de façons de recouvrir entièrement un rectangle  $2 \times n$  avec des dominos  $2 \times 1$ ? Nous supposerons que les dominos sont tous identiques (par exemple parce qu’ils sont retournés, ou parce que quelqu’un s’est amusé à tous les peindre en rouge) ; seules comptent leurs orientations, horizontales ou verticales. C’est comme si nous travaillions en fait avec des carreaux ou des pavés en forme de dominos. Par exemple, les trois pavages possibles d’un rectangle  $2 \times 3$  sont  $\square\square\square$ ,  $\square\square\Box$  et  $\Box\square\Box$  ; donc  $T_3 = 3$ .

Pour trouver une forme close de  $T_n$ , commençons, comme d’habitude, par regarder les premiers cas. Si  $n = 1$  il n’y a évidemment qu’un pavage, “Let me count the ways.”

— E. B. Browning

Nous n’avons pas considéré le cas  $n = 0$  : combien y a-t-il de pavages d’un rectangle  $2 \times 0$ ? Le sens de cette question n’est pas évident a priori, mais ce n’est pas la première fois que nous voyons une situation de ce genre : nous savons notamment qu’il existe une permutation de zéro objet, et donc que  $0! = 1$  ; il y a une façon de choisir zéro objet parmi  $n$  (donc de ne rien choisir), ce qui implique que  $\binom{n}{0} = 1$  ; il existe une partition de l’ensemble vide en zéro ensemble non vide, mais par contre il n’existe pas de partition similaire pour un ensemble non vide, c’est pourquoi  $\{\binom{n}{0}\} = [n = 0]$ . Par un

raisonnement similaire, nous pouvons conclure qu'il existe exactement une façon de pavier un rectangle  $2 \times 0$  avec des dominos, qui consiste à n'utiliser aucun domino ; par conséquent,  $T_0 = 1$ . Remarquons que ceci détruit le joli motif qui semblait se dessiner, à savoir que  $T_n = n$  pour  $n = 1, 2$  et  $3$  ; mais c'était inévitable, car la logique veut que  $T_0$  soit égal à  $1$ . La bonne compréhension du cas nul s'avère généralement très utile pour bien résoudre un problème d'énumération.

Regardons encore un petit cas de plus,  $n = 4$ . Il y a deux façons possibles de pavier l'extrême gauche du rectangle : un seul domino vertical ou deux dominos horizontaux. Si on choisit la première solution,  $\square\square$ , il reste un rectangle  $2 \times 3$  qu'on peut pavier de  $T_3$  manières. Dans l'autre cas,  $\square\square$ , la figure peut être complétée de  $T_2$  façons. Par conséquent,  $T_4 = T_3 + T_2 = 5$ , et voici les cinq pavages :  $\square\square\square\square$ ,  $\square\square\square\square$ ,  $\square\square\square\square$ ,  $\square\square\square\square$  et  $\square\square\square\square$ .

Nous connaissons maintenant les cinq premières valeurs de  $T_n$  :

$n$	0	1	2	3	4
$T_n$	1	1	2	3	5

On dirait des nombres de Fibonacci, non ? En effet, et il n'est pas difficile de voir pourquoi : le raisonnement que nous avons fait pour montrer que  $T_4 = T_3 + T_2$  se généralise sans problème à  $T_n = T_{n-1} + T_{n-2}$  pour tout  $n \geq 2$ . La récurrence est donc exactement la même que celle des nombres de Fibonacci ; seules les valeurs initiales,  $T_0 = 1$  et  $T_1 = 1$ , changent ; elles correspondent en fait à  $F_1$  et  $F_2$ . Par conséquent, les  $T_n$  sont des nombres de Fibonacci juste décalés d'un pas :

$$T_n = F_{n+1}, \quad \text{pour } n \geq 0.$$

Nous pouvons considérer que c'est une forme close pour les  $T_n$ , car les nombres de Fibonacci sont suffisamment importants pour qu'on puisse les considérer comme "connus". De toutes façons,  $F_n$  a lui-même une forme close (6.123). Notez en passant que le résultat que nous venons d'établir confirme que l'option  $T_0 = 1$  était la bonne.

Qu'est-ce que tout cela a à voir avec les fonctions génératrices, nous direz-vous ? Patience, elles ne vont pas tarder. Elles interviennent dans une autre façon de calculer  $T_n$ . Cette nouvelle méthode est basée sur une idée audacieuse. Considérons la "somme"  $T$  de tous les pavages  $2 \times n$  possibles, pour tout  $n \geq 0$  :

$$T = 1 + \square + \square\square + \square\square\square + \square\square\square\square + \square\square\square\square\square + \cdots \tag{7.1}$$

(le terme " $1$ " du membre droit représente le pavage nul, celui du rectangle  $2 \times 0$ ).

Allons hardiment  
là où personne n'a  
jamais pavé le pied.

Vous n'êtes pas sans remarquer que les termes de la somme sont des pavages, donc des objets combinatoires. Nous n'allons pas entrer dans les détails pour justifier formellement ce qui est légal ou pas sur ce type de sommes ; sachez simplement que tout cela peut être défini de façon rigoureuse.

Maintenant que nous savons additionner des pavages, voyons comment les multiplier. La multiplication s'effectue par juxtaposition. Par exemple, si on multiplie les pavages  $\square$  et  $\square$ , on obtient  $\square\square$ . Notez bien que cette multiplication n'est pas commutative :  $\square\square$  n'est pas égal à  $\square\square$ . L'élément neutre de la multiplication des pavages est le pavage nul. On a notamment  $\square \times \square = \square = \square \times \square$ , et cela se généralise à tous les pavages.

Nous en connaissons maintenant assez sur l'arithmétique des dominos pour pouvoir l'appliquer à notre somme infinie  $T$  :

$$\begin{aligned} T &= \square + \square\square + \square\square\square + \square\square\square\square + \square\square\square\square\square + \dots \\ &= \square(\square + \square\square + \square\square\square + \dots) + \square\square(\square + \square\square + \square\square\square + \dots) \\ &= \square T + \square\square T. \end{aligned} \tag{7.2}$$

Tout pavage correct apparaît exactement une fois dans chacun des membres droits. Ce que nous venons de faire est donc tout à fait raisonnable, même si nous avons totalement ignoré les avertissements du chapitre 2 concernant la "convergence absolue". La dernière ligne de cette équation indique que tout pavage de  $T$  est soit le pavage nul, soit un domino vertical suivi élément de  $T$ , soit deux dominos horizontaux suivis d'un élément de  $T$ .

Essayons maintenant de résoudre cette équation en  $T$ . En remplaçant le  $T$  de gauche par  $\square T$  et en soustrayant les deux derniers termes de droite, on arrive à

$$(\square - \square - \square)\square T = \square. \tag{7.3}$$

Vérifions que cela est bien cohérent en développant ce dernier résultat :

$$\begin{array}{r} \square + \square\square + \square\square\square + \square\square\square\square + \square\square\square\square\square + \dots \\ - \square - \square\square - \square\square\square - \square\square\square\square - \square\square\square\square\square - \dots \\ - \square - \square\square - \square\square\square - \square\square\square\square - \square\square\square\square\square - \dots \\ \hline \end{array}$$

Chaque terme de la ligne du haut se trouve bien annulé par un terme de la seconde ou de la troisième ligne. Tout est donc correct.

Il n'est pas difficile de saisir la signification combinatoire des équations que nous avons écrites jusqu'ici. Nous allons maintenant faire quelque chose dont le sens peut paraître bien moins évident : diviser les deux membres

*A mon avis, la somme converge à condition que les dominos soient assez petits.*

de (7.3) par  $\boxed{1 - \square - \blacksquare}$ . Nous obtenons ainsi

$$T = \frac{1}{\boxed{1 - \square - \blacksquare}}. \quad (7.4)$$

En fait, nous trichons un peu car nous ne distinguons pas la division à droite et la division à gauche, alors que la multiplication n'est pas commutative. Toutefois, cela n'a aucune importance dans le cas présent car  $\boxed{1}$  commute avec tous les autres pavages. L'étape suivante consiste à appliquer la règle

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + \dots$$

pour développer notre fraction en série. Le pavage nul, qui est l'élément neutre de notre multiplication combinatoire, joue naturellement le rôle du  $1$  ; et  $\square + \blacksquare$  remplace  $z$ . Nous en déduisons donc que

$$\begin{aligned} \frac{1}{\boxed{1 - \square - \blacksquare}} &= \boxed{1} + (\square + \blacksquare) + (\square + \blacksquare)^2 + (\square + \blacksquare)^3 + \dots \\ &= \boxed{1} + (\square + \blacksquare) + (\square\square + \square\blacksquare + \blacksquare\square + \blacksquare\blacksquare) \\ &\quad + (\square\square\square + \square\square\blacksquare + \square\blacksquare\square + \square\blacksquare\blacksquare + \blacksquare\square\square + \blacksquare\square\blacksquare + \blacksquare\blacksquare\square + \blacksquare\blacksquare\blacksquare) + \dots \end{aligned}$$

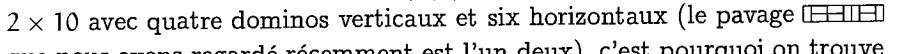
Nous retrouvons exactement  $T$ , bien que l'ordre des pavages soit différent de celui de tout à l'heure. Chaque pavage apparaît exactement une fois dans cette somme. Par exemple,  $\square\square\square\square$  se trouve dans le développement de  $(\square + \blacksquare)^7$ . Pour obtenir plus d'informations, on peut "compresser" la somme afin de n'en conserver que ce qui est important. Imaginons par exemple que les dominos puissent se séparer les uns des autres et commuter dans un pavage. Alors un terme comme  $\square\square\square\square\square\square$ , qui contient quatre dominos verticaux et six horizontaux, devient  $\square^4 \square^6$ . On obtient ainsi la série

$$T = \boxed{1} + \square + \square^2 + \square^3 + \square^4 + 2\square\square^2 + \square^4 + 3\square^2\square^2 + \square^4 + \dots$$

le terme  $2\square\square^2$  représente les deux termes de l'ancienne somme qui contenaient un domino vertical et deux horizontaux :  $\square\square$  et  $\square\square$ . De même,  $3\square^2\square^2$  représente les trois termes  $\square\square\square$ ,  $\square\square\blacksquare$  et  $\blacksquare\square\square$ . En fait, nous considérons  $\square$  et  $\blacksquare$  exactement comme des variables (commutatives) ordinaires.

Il suffit d'appliquer la formule du binôme pour trouver une forme close des coefficients de cette série commutative  $T$  :

$$\begin{aligned} \frac{1}{\boxed{1 - (\square + \blacksquare)^2}} &= \boxed{1} + (\square + \blacksquare^2) + (\square + \blacksquare^2)^2 + (\square + \blacksquare^2)^3 + \dots \\ &= \sum_{k \geq 0} (\square + \blacksquare^2)^k = \sum_{j, k \geq 0} \binom{k}{j} \square^j \blacksquare^{2k-2j} \\ &= \sum_{j, m \geq 0} \binom{j+m}{j} \square^j \blacksquare^{2m}. \end{aligned} \quad (7.5)$$

(Nous pouvons remplacer  $k - j$  par  $m$  dans la dernière ligne car  $\binom{k}{j} = 0$  pour  $0 \leq k < j$ ). Nous en concluons que  $\binom{j+m}{j}$  compte le nombre de façons de pavier un rectangle  $2 \times (j+2m)$  avec  $j$  dominos verticaux et  $2m$  dominos horizontaux. Il y a par exemple  $\binom{4+3}{4} = 35$  façons de pavier un rectangle  $2 \times 10$  avec quatre dominos verticaux et six horizontaux (le pavage  que nous avons regardé récemment est l'un deux), c'est pourquoi on trouve dans la version commutative de  $T$  le terme  $35 \square^4 \square^6$ .

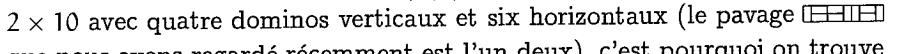
On peut "compresser" encore la série en ignorant l'orientation des dominos. Supposons que la différence entre dominos horizontaux et dominos verticaux ne nous intéresse pas ; nous voulons seulement connaître le nombre total de pavages  $2 \times n$  (c'est d'ailleurs ce problème-ci que nous nous sommes posés au début). Pour cela, il nous suffit de remplacer  $\square$  et  $\square$  par une seule variable  $z$ . Tant que nous y sommes, remplaçons aussi  $|$  par  $1$ , pour obtenir

*Je suis désorienté.*

$$T = \frac{1}{1 - z - z^2}. \quad (7.6)$$

C'est la fonction génératrice (6.117) des nombres de Fibonacci, à laquelle il manque juste le facteur  $z$  au numérateur. Nous en concluons donc que le coefficient de  $z^n$  dans  $T$  est égal à  $F_{n+1}$ .

Les expressions  $|/(|-\square-\square|)$ ,  $|/(|-\square-\square^2|)$  et  $1/(1-z-z^2)$  que nous avons trouvées pour  $T$  sont appelées des *fonctions génératrices*, car c'est avec elles qu'on peut "engendrer" les coefficients qui nous intéressent.

Remarquons qu'on peut déduire de notre calcul que le nombre de pavages  $2 \times n$  comprenant exactement  $m$  paires de dominos horizontaux est égal à  $\binom{n-m}{m}$ . En effet, comme il y a  $j = n - 2m$  dominos verticaux, il y a  $\binom{j+m}{j} = \binom{j+m}{m} = \binom{n-m}{m}$  pavages possibles. Nous avons vu au chapitre 6 que  $\binom{n-m}{m}$  est aussi le nombre de suites du code de Morse de longueur  $n$  contenant  $m$  traits. La correspondance entre les pavages  $2 \times n$  par des dominos et ces suites est facile à trouver : par exemple, le pavage  est associé à "`.-.-.-.-.`". Ainsi, les pavages par des dominos sont étroitement liés aux polynômes continuants que nous avons étudiés au chapitre 6. Comme le monde est petit !

Nous avons résolu le problème de  $T_n$  de deux façons différentes. La première méthode a consisté à conjecturer la réponse et à la prouver par induction. Dans la seconde, nous avons construit des sommes infinies de pavages puis extrait de ces sommes les coefficients intéressants. La première s'est avérée plus facile que la seconde, qui était en revanche bien plus originale. Croyez-vous que nous avons introduit cette dernière juste pour le plaisir de jouer aux dominos comme s'ils étaient des variables algébriques ? Bien sûr que non : la véritable raison, c'est que cette approche est énormément plus puissante que la première. Elle peut s'appliquer à une variété beaucoup plus étendue de problèmes, notamment parce qu'elle permet de

trouver la réponse sans qu'il nous soit nécessaire de la deviner.

Généralisons notre problème de façon qu'il nous soit raisonnablement impossible de conjecturer la réponse : quel est le nombre  $U_n$  de façons de pavier un rectangle  $3 \times n$  avec des dominos ?

Regardons les premiers cas. Comme précédemment, nous avons  $U_0 = 1$ . Si  $n = 1$ , aucun pavage n'est possible car un rectangle  $3 \times 1$  est trop grand pour un seul domino et trop petit pour deux. Le cas suivant,  $n = 2$ , ne pose pas de problème : il y a trois pavages,  $\square$ ,  $\square\square$  et  $\square\square\square$ , par conséquent  $U_2 = 3$ . En fait nous le savions déjà car  $T_3 = 3$  et le nombre de pavages d'un rectangle  $3 \times 2$  est bien évidemment égal à celui d'un rectangle  $2 \times 3$ . Pour  $n = 3$ , comme pour  $n = 1$ , il n'y a pas de pavage possible. Pour s'en rendre compte, on peut faire une recherche exhaustive des candidats possibles ; on peut aussi voir les choses avec un peu de recul : comme un rectangle  $3 \times 3$  est d'aire impaire, il ne peut certainement pas être pavé par des dominos d'aire paire (le même argument s'applique bien entendu pour tout  $n$  impair). Si  $n = 4$ , il semble y avoir une douzaine de pavages, mais, pour être sûr du nombre exact, il faut passer un très long moment à vérifier si on n'en a pas oublié.

Essayons donc d'utiliser l'approche qui nous a si bien réussi la dernière fois, à savoir les sommes infinies :

$$U = | + \square + \square\square + \square\square\square + \square\square\square\square + \square\square\square\square\square + \square\square\square\square\square\square + \cdots. \quad (7.7)$$

Tout pavage non nul commence soit par  $\square$ , soit par  $\square\square$ , soit par  $\square\square\square$ . Hélas, les deux premières possibilités ne se factorisent pas pour donner des termes en  $U$ , comme les termes en  $T$  que nous avions tout à l'heure. Toutefois, la somme de tous les termes en  $U$  qui commencent par  $\square$  peut s'écrire  $\square V$ , où

$$V = \square + \square\square + \square\square\square + \square\square\square\square + \cdots$$

désigne la somme de tous les pavages d'un rectangle  $3 \times n$  auquel il manque le coin inférieur gauche. De même, les termes de  $U$  qui commencent par  $\square\square$  peuvent s'écrire  $\square\square \Lambda$ , où

$$\Lambda = \square + \square\square + \square\square\square + \square\square\square\square + \cdots$$

représente la somme de tous les rectangles auxquels manque le coin supérieur gauche. La série  $\Lambda$  est l'image-miroir de  $V$ . Ces factorisations nous permettent d'écrire

$$U = | + \square V + \square\square \Lambda + \square\square\square U.$$

Nous pouvons aussi factoriser  $V$  et  $\Lambda$ , car leurs pavages ne peuvent commencer que de deux façons :

$$V = \square U + \square\square V, \quad \Lambda = \square U + \square\square \Lambda.$$

Nous disposons maintenant d'un système de trois équations à trois inconnues  $U$ ,  $V$  et  $\Lambda$ . Pour le résoudre, commençons par exprimer  $V$  et  $\Lambda$  en fonction de  $U$ , puis injectons ces résultats dans l'équation de  $U$  :

$$V = (| - \square) \cdot U, \quad \Lambda = (| - \square^2) \cdot U;$$

$$U = | + \square(| - \square) \cdot U + \square^2(| - \square^2) \cdot U + \square \cdot U.$$

Il ne reste qu'à résoudre cette dernière équation pour trouver

$$U = \frac{1}{| - \square(| - \square) \cdot U - \square^2(| - \square^2) \cdot U - \square \cdot U}. \quad (7.8)$$

Cette expression définit la somme infinie  $U$ , exactement comme (7.4) définit  $T$ .

Détachons tous les dominos et autorisons la commutativité pour ne faire apparaître que des puissances de  $\square$  et  $\square^2$ . Voyez comme tout se simplifie admirablement :

$$\begin{aligned} U &= \frac{1}{1 - \square^2 \cdot \square (1 - \square^3)^{-1} - \square^2 \cdot \square (1 - \square^3)^{-1} - \square^3} \\ &= \frac{1 - \square^3}{(1 - \square^3)^2 - 2\square^2 \cdot \square} \\ &= \frac{(1 - \square^3)^{-1}}{1 - 2\square^2 \cdot \square (1 - \square^3)^{-2}} \\ &= \frac{1}{1 - \square^3} + \frac{2\square^2 \cdot \square}{(1 - \square^3)^3} + \frac{4\square^4 \cdot \square^2}{(1 - \square^3)^5} + \frac{8\square^6 \cdot \square^3}{(1 - \square^3)^7} + \dots \\ &= \sum_{k \geq 0} \frac{2^k \cdot \square^{2k} \cdot \square^k}{(1 - \square^3)^{2k+1}} \\ &= \sum_{k, m \geq 0} \binom{m+2k}{m} 2^k \cdot \square^{2k} \cdot \square^{k+3m}. \end{aligned}$$

N'hésitez pas à étudier attentivement ce calcul. Notez en particulier que la dernière étape s'effectue en appliquant l'identité (5.56),  $(1 - w)^{-2k-1} = \sum_m \binom{m+2k}{m} w^m$ . Voyons maintenant ce que nous dit la dernière ligne. D'abord, elle nous dit que tout pavage  $3 \times n$  contient un nombre pair de dominos verticaux. De plus, s'il y a  $2k$  dominos verticaux, il doit y en avoir au moins  $k$  horizontaux, et il existe un entier  $m \geq 0$  tel que le nombre de ces dominos horizontaux est égal à  $k + 3m$ . Enfin, le nombre de pavages avec  $2k$  dominos verticaux et  $k + 3m$  dominos horizontaux est exactement  $\binom{m+2k}{m} 2^k$ .

*Dans un autre cours, j'ai appris à manipuler des "expressions rationnelles". Si je ne me trompe, on peut écrire*

$$U = (\square \cdot \square^*) \cdot (\square^2 \cdot \square^*)$$

*dans le langage des expressions rationnelles. Il y a donc certainement un lien entre les expressions rationnelles et les fonctions génératrices.*

Nous pouvons maintenant essayer de compter les pavages  $3 \times 4$ , avec bien plus de chances de succès que tout à l'heure. Lorsque  $n = 4$ , l'aire du rectangle est égale à 12, donc il nous faut six dominos. Parmi ceux-ci, il doit y en avoir  $2k$  verticaux et  $k+3m$  horizontaux ; donc  $2k+k+3m=6$ , ce qui donne  $k+m=2$ . Si on n'utilise pas de domino vertical, alors  $k=0$ ,  $m=2$ , et le nombre correspondant de pavages est  $\binom{2+0}{2}2^0 = 1$  (c'est le pavage ). Si on en utilise deux verticaux, alors  $k=1$ ,  $m=1$ , et on obtient  $\binom{1+2}{1}2^1 = 6$  pavages. Enfin, pour quatre dominos verticaux, on a  $k=2$ ,  $m=0$ , et donc  $\binom{0+4}{0}2^2 = 4$  pavages possibles. Tout cela nous donne  $U_4 = 11$ . De façon générale, si  $n$  est pair, on montre avec le même raisonnement que  $k+m=\frac{1}{2}n$ , donc  $\binom{m+2k}{m} = \binom{n/2+k}{n/2-k}$ , et le nombre total de pavages  $3 \times n$  est

$$U_n = \sum_k \binom{n/2+k}{n/2-k} 2^k = \sum_m \binom{n-m}{m} 2^{n/2-m}. \quad (7.9)$$

Comme auparavant, nous pouvons remplacer  $\square$  et  $\square$  par  $z$  pour obtenir une fonction génératrice qui ne différencie pas les orientations des dominos. Voici ce que cela donne :

$$U = \frac{1}{1-z^3(1-z^3)^{-1}-z^3(1-z^3)^{-1}-z^3} = \frac{1-z^3}{1-4z^3+z^6}. \quad (7.10)$$

En développant en série, on trouve

$$U = 1 + U_2 z^3 + U_4 z^6 + U_6 z^9 + U_8 z^{12} + \dots$$

Le curieux décalage entre les indices et les exposants de cette formule s'explique facilement. La coefficient de  $z^9$ , par exemple, est  $U_6$ , qui compte le nombre de pavages  $3 \times 6$ . C'est bien ce que nous voulons, car chacun de ces pavages contient neuf dominos.

Nous pourrions étudier (7.10) et en tirer une forme close pour ses coefficients. Laissons cependant cela pour plus tard, lorsque nous serons plus expérimentés. Quittons donc les dominos pour le moment et passons au second problème promis, "la monnaie".

De combien de façons différentes peut-on payer une somme de 50 cents américains ? Nous supposerons disposer de pièces de 1 cent  $\textcircled{1}$ , 5 cents  $\textcircled{5}$ , 10 cents  $\textcircled{10}$ , 25 cents  $\textcircled{25}$  et 50 cents  $\textcircled{50}$ . C'est George Pólya [298] qui a popularisé ce problème en montrant qu'il pouvait être résolu de façon très instructive avec les fonctions génératrices.

Nous allons considérer des sommes infinies qui représentent toutes les façons possibles de payer avec ces pièces. Pour débuter doucement, considérons tout d'abord que nous n'avons que des pièces de un cent. La somme de

toutes les façons possibles de payer avec ces pièces uniquement peut s'écrire

$$\begin{aligned} P &= \cancel{\$} + \textcircled{1} + \textcircled{1}\textcircled{1} + \textcircled{1}\textcircled{1}\textcircled{1} + \textcircled{1}\textcircled{1}\textcircled{1}\textcircled{1} + \cdots \\ &= \cancel{\$} + \textcircled{1} + \textcircled{1}^2 + \textcircled{1}^3 + \textcircled{1}^4 + \cdots . \end{aligned}$$

Le premier terme désigne la façon de payer zéro cent, le deuxième celle de payer un cent, puis deux, trois et ainsi de suite. Supposons maintenant que nous disposons, en plus des pièces de 1 cent, de pièces de 5 cents. On obtient ainsi la somme

$$\begin{aligned} N &= P + \textcircled{5}P + \textcircled{5}\textcircled{5}P + \textcircled{5}\textcircled{5}\textcircled{5}P + \textcircled{5}\textcircled{5}\textcircled{5}\textcircled{5}P + \cdots \\ &= (\cancel{\$} + \textcircled{5} + \textcircled{5}^2 + \textcircled{5}^3 + \textcircled{5}^4 + \cdots)P, \end{aligned}$$

du fait que tout paiement s'effectue avec un certain nombre de pièces de 5 cents choisies dans le premier facteur et un certain nombre de pièces de 1 cent choisies dans P. Notez bien que N n'est pas la somme  $\cancel{\$} + \textcircled{1} + \textcircled{5} + (\textcircled{1} + \textcircled{5})^2 + (\textcircled{1} + \textcircled{5})^3 + \cdots$ , car cette somme-ci peut contenir plusieurs fois un même mode de paiement. Par exemple, le terme  $(\textcircled{1} + \textcircled{5})^2 = \textcircled{1}\textcircled{1} + \textcircled{1}\textcircled{5} + \textcircled{5}\textcircled{1} + \textcircled{5}\textcircled{5}$  considère  $\textcircled{1}\textcircled{5}$  et  $\textcircled{5}\textcircled{1}$  comme différents, alors que l'ordre ne compte pas dans notre problème.

Procérons de manière similaire pour introduire les pièces de 10 cents. Nous obtenons la somme infinie

$$D = (\cancel{\$} + \textcircled{10} + \textcircled{10}^2 + \textcircled{10}^3 + \textcircled{10}^4 + \cdots)N,$$

qui contient des termes comme  $\textcircled{10}^3\textcircled{5}^3\textcircled{1}^5 = \textcircled{10}\textcircled{10}\textcircled{10}\textcircled{5}\textcircled{5}\textcircled{5}\textcircled{1}\textcircled{1}\textcircled{1}\textcircled{1}\textcircled{1}\textcircled{1}$ . Chacun de ces termes représente une façon différente de payer une certaine somme. Si on ajoute à cela les pièces de 25 et 50 cents, on trouve

$$\begin{aligned} Q &= (\cancel{\$} + \textcircled{25} + \textcircled{25}^2 + \textcircled{25}^3 + \textcircled{25}^4 + \cdots)D; \\ C &= (\cancel{\$} + \textcircled{50} + \textcircled{50}^2 + \textcircled{50}^3 + \textcircled{50}^4 + \cdots)Q. \end{aligned}$$

Il nous faut maintenant trouver le nombre de termes de C qui valent exactement 50 cents.

Une simple astuce va nous permettre de résoudre élégamment ce problème : remplaçons  $\textcircled{1}$  par  $z$ ,  $\textcircled{5}$  par  $z^5$ ,  $\textcircled{10}$  par  $z^{10}$ ,  $\textcircled{25}$  par  $z^{25}$  et  $\textcircled{50}$  par  $z^{50}$ . Chaque terme de valeur  $n$  dans la somme d'origine se trouve alors remplacé par  $z^n$ . Par exemple, le terme  $\textcircled{50}\textcircled{10}\textcircled{5}\textcircled{5}\textcircled{1}$  devient  $z^{50+10+5+5+1} = z^{71}$ . Chacune des quatre façons possibles de payer 13 cents, à savoir  $\textcircled{10}\textcircled{1}^3$ ,  $\textcircled{5}\textcircled{1}^8$ ,  $\textcircled{5}^2\textcircled{1}^3$  et  $\textcircled{1}^{13}$ , se transforme en  $z^{13}$ ; le coefficient de  $z^{13}$  sera donc égal à 4 après ces substitutions.

Soit  $P_n$ , (respectivement  $N_n$ ,  $D_n$ ,  $Q_n$  et  $C_n$ ) le nombre de façons de payer  $n$  cents avec des pièces d'au plus 1 (respectivement 5, 10, 25 et

*Les noms P, N, D, Q des séries génératrices sont les initiales des noms que les Américains donnent à leurs pièces de monnaie : penny (1 cent), nickel (5 cents), dime (10 cents), quarter (25 cents) (N.d.T.).*

50) cent(s). D'après notre analyse, ces nombres sont les coefficients de  $z^n$  dans les séries

$$\begin{aligned} P &= 1 + z + z^2 + z^3 + z^4 + \dots, \\ N &= (1 + z^5 + z^{10} + z^{15} + z^{20} + \dots)P, \\ D &= (1 + z^{10} + z^{20} + z^{30} + z^{40} + \dots)N, \\ Q &= (1 + z^{25} + z^{50} + z^{75} + z^{100} + \dots)D, \\ C &= (1 + z^{50} + z^{100} + z^{150} + z^{200} + \dots)Q. \end{aligned}$$

*Combien existe-t-il réellement de pièces de 1 cent ?*

*Je parie que si  $n$  est plus grand que, disons  $10^{10}$ , alors  $P_n = 0$  dans le "monde réel".*

Il est évident que  $P_n = 1$  pour tout  $n \geq 0$ . De même, il suffit de réfléchir un petit peu pour prouver que  $N_n = \lfloor n/5 \rfloor + 1$  : pour faire  $n$  cents avec des pièces de 1 et 5 cents, il faut prendre 0 ou 1 ou ... ou  $\lfloor n/5 \rfloor$  pièces de 5 cents, après quoi on n'a pas d'autre choix que de compléter la somme avec des pièces de 1 cent. Ainsi,  $P_n$  et  $N_n$  sont simples. Cependant, les valeurs de  $D_n$ ,  $Q_n$  et  $C_n$  sont bien plus compliquées.

Voici ce que nous pouvons faire. Souvenons-nous que  $1 + z^m + z^{2m} + \dots$  est tout simplement égal à  $1/(1 - z^m)$ , et écrivons

$$\begin{aligned} P &= 1/(1 - z), \\ N &= P/(1 - z^5), \\ D &= N/(1 - z^{10}), \\ Q &= D/(1 - z^{25}), \\ C &= Q/(1 - z^{50}). \end{aligned}$$

Multiplions ces équations par leurs dénominateurs :

$$\begin{aligned} (1 - z)P &= 1, \\ (1 - z^5)N &= P, \\ (1 - z^{10})D &= N, \\ (1 - z^{25})Q &= D, \\ (1 - z^{50})C &= Q. \end{aligned}$$

Nous pouvons maintenant mettre en équations les coefficients de  $z^n$  de chacune de ces égalités. Nous obtenons ainsi des relations de récurrence grâce auxquelles nous pourrons rapidement calculer les coefficients désirés :

$$\begin{aligned} P_n &= P_{n-1} + [n=0], \\ N_n &= N_{n-5} + P_n, \\ D_n &= D_{n-10} + N_n, \\ Q_n &= Q_{n-25} + D_n, \\ C_n &= C_{n-50} + Q_n. \end{aligned}$$

Par exemple, c'est parce que le coefficient de  $z^n$  dans  $D = (1 - z^{25})Q$  est égal à  $Q_n - Q_{n-25}$  qu'on peut écrire que  $Q_n - Q_{n-25} = D_n$ .

Nous pourrions développer ces récurrences pour trouver des formules du genre de  $Q_n = D_n + D_{n-25} + D_{n-50} + D_{n-75} + \dots$ , qui s'arrête lorsque les indices deviennent négatifs. Toutefois, la forme non itérée est bien pratique car une addition suffit pour calculer chaque coefficient, comme dans le triangle de Pascal.

Appliquons ces récurrences pour trouver  $C_{50}$ . Pour commencer,  $C_{50} = C_0 + Q_{50}$ ; il nous faut donc  $Q_{50}$ . Or,  $Q_{50} = Q_{25} + D_{50}$ , et  $Q_{25} = Q_0 + D_{25}$ ; il nous faut donc aussi  $D_{50}$  et  $D_{25}$ . Ces coefficients  $D_n$  dépendent à leur tour de  $D_{40}, D_{30}, D_{20}, D_{15}, D_{10}, D_5$  et de  $N_{50}, N_{45}, \dots, N_5$ . Ce simple calcul suffit pour déterminer tous les coefficients nécessaires :

$n$	0	5	10	15	20	25	30	35	40	45	50
$P_n$	1	1	1	1	1	1	1	1	1	1	1
$N_n$	1	2	3	4	5	6	7	8	9	10	11
$D_n$	1	2	4	6	9	12	16		25		36
$Q_n$	1				13					49	
$C_n$	1										50

La dernière valeur de la table nous donne la réponse,  $C_{50}$ : il y a exactement 50 façons de laisser un pourboire de 50 cents.

Pouvons-nous espérer une forme close pour  $C_n$ ? En multipliant toutes les équations ensemble, on trouve une expression compacte :

$$C = \frac{1}{1-z} \frac{1}{1-z^5} \frac{1}{1-z^{10}} \frac{1}{1-z^{25}} \frac{1}{1-z^{50}}. \quad (7.11)$$

Il n'est pas évident à première vue d'en tirer le coefficient de  $z^n$ . C'est néanmoins possible; nous verrons comment plus tard dans ce même chapitre.

Si on habite dans un pays où circulent des pièces de toutes les valeurs entières possibles ( $\textcircled{1}$ ,  $\textcircled{2}$ ,  $\textcircled{3}$ , ... ) on a la chance de pouvoir obtenir des formules plus élégantes. La fonction génératrice correspondante est un produit infini de fractions :

$$\frac{1}{(1-z)(1-z^2)(1-z^3)\dots}.$$

Le coefficient de  $z^n$  dans le développement en série de cette expression s'appelle le nombre de *partitions* de  $n$ ; on le note  $p(n)$ . Une partition de  $n$  est une représentation de  $n$  par une somme d'entiers strictement positifs, sans considération d'ordre. Il y a par exemple sept partitions distinctes de 5, qui sont

$$5 = 4+1 = 3+2 = 3+1+1 = 2+2+1 = 2+1+1+1 = 1+1+1+1+1;$$

(Sans compter la possibilité de le payer par carte de crédit).

donc  $p(5) = 7$ . Comme  $p(2) = 2$ ,  $p(3) = 3$ ,  $p(4) = 5$  et  $p(6) = 11$ , on pourrait croire a priori que  $p(n)$  est toujours premier. Ce n'est pas le cas car  $p(7) = 15$ . Il n'y a pas de forme close pour  $p(n)$ . Sachez toutefois que la théorie des partitions est une branche fascinante des mathématiques, qui a donné lieu à beaucoup de découvertes remarquables. Par exemple, Ramanujan, en effectuant d'ingénieuses transformations de séries génératrices, a démontré que  $p(5n + 4) \equiv 0 \pmod{5}$ ,  $p(7n + 5) \equiv 0 \pmod{7}$  et  $p(11n + 6) \equiv 0 \pmod{11}$  (voir Andrews [11, chapitre 10]).

## 7.2 MANOEUVRES DE BASE

Regardons maintenant de plus près quelques techniques grâce auxquelles les séries génératrices s'avèrent si puissantes.

Voici tout d'abord quelques mots concernant la terminologie et les notations. Notre fonction génératrice générique, sur laquelle nous allons travailler, est de la forme

$$G(z) = g_0 + g_1 z + g_2 z^2 + \cdots = \sum_{n \geq 0} g_n z^n. \quad (7.12)$$

On dit que  $G(z)$ , ou  $G$  tout simplement, est la fonction génératrice de la suite  $\langle g_0, g_1, g_2, \dots \rangle$ , qu'on note aussi  $\langle g_n \rangle$ . Le coefficient  $g_n$  de  $z^n$  dans  $G(z)$  est souvent désigné par  $[z^n] G(z)$ , comme dans la section 5.4.

La somme de (7.12) porte sur tous les  $n \geq 0$ , mais il est souvent plus pratique de l'étendre à tous les entiers  $n$ . Pour cela, il suffit de considérer que  $g_{-1} = g_{-2} = \dots = 0$ . Dans ce cas, on continue de parler de la suite  $\langle g_0, g_1, g_2, \dots \rangle$ , comme si les  $g_n$  n'existaient pas lorsque  $n$  est strictement négatif.

Lorsqu'on travaille avec les fonctions génératrices, on est confronté à deux types de "formes closes". On peut avoir une forme close pour  $G(z)$ , qui s'exprime en fonction de  $z$ ; ou bien une forme close pour  $g_n$ , qui s'exprime en fonction de  $n$ . Par exemple, la fonction génératrice des nombres de Fibonacci admet la forme close  $z/(1 - z - z^2)$ , tandis que les nombres de Fibonacci eux-mêmes admettent la forme close  $(\phi^n - \bar{\phi}^n)/\sqrt{5}$ . C'est le contexte qui indique généralement le type de forme close que l'on considère.

Selon le point de vue d'où on se place, la fonction génératrice  $G(z)$  peut prendre deux aspects très différents. Elle peut être une fonction de la variable complexe  $z$  et statisfaire toutes les propriétés standard décrites dans le livre d'analyse. Elle peut aussi être simplement une série formelle en  $z$ . C'est cette dernière interprétation qui a prévalu, par exemple, dans la section précédente : en plusieurs occasions, nous avons introduit  $z$  pour remplacer des constituants d'objets combinatoires dans des "sommes" de ces objets. Le coefficient de  $z^n$  représentait alors le nombre d'objets comportant

$n$  constituants.

Quand on voit  $G(z)$  comme une fonction d'une variable complexe, on se pose le problème de sa convergence. Nous avons vu au chapitre 2 que la série infinie  $\sum_{n \geq 0} g_n z^n$  converge (absolument) si et seulement s'il existe une constante  $A$  telle que, pour tout  $N$ , la somme finie  $\sum_{0 \leq n \leq N} |g_n z^n|$  ne dépasse jamais  $A$ . On en déduit immédiatement que si  $\sum_{n \geq 0} g_n z^n$  converge pour une certaine valeur  $z = z_0$ , elle converge aussi pour tout  $z$  tel que  $|z| < |z_0|$ . De plus, dans ce cas,  $\lim_{n \rightarrow \infty} |g_n z_0^n| = 0$ ; donc, si la série converge en  $z_0$ , alors on peut écrire, avec les notations du chapitre 9,  $g_n = O(|1/z_0|^n)$ . Réciproquement, si  $g_n = O(M^n)$ , alors la série  $\sum_{n \geq 0} g_n z^n$  converge pour tout  $|z| < 1/M$ . Ces faits constituent le B.A.-Ba de la convergence des séries.

En ce qui nous concerne, les problèmes de convergence ne nous intéresseront que lorsque nous étudierons le comportement asymptotique des coefficients de séries génératrices. Ceci mis à part, nous n'utiliserons pratiquement que des opérations qui sont rigoureusement justifiées lorsqu'elles s'appliquent aux séries formelles, sans que la convergence entre en ligne de compte (on trouvera tous les éléments théoriques dans Bell [23], Niven [282], ou Henrici [182, chapitre 1] par exemple).

De toutes façons, même si on manipule des séries génératrices sans s'assurer de la légalité des opérations, le résultat qu'on trouve ainsi peut généralement être prouvé par induction. Par exemple, la fonction génératrice des nombres de Fibonacci ne converge que si  $|z| < 1/\phi \approx 0,618$ , mais nous n'avons pas eu besoin de le savoir pour démontrer que  $F_n = (\phi^n - \bar{\phi}^n)/\sqrt{5}$ . Une fois qu'on a établi cette formule, on peut la vérifier d'une autre manière si on ne fait pas confiance à la théorie des séries formelles. C'est pourquoi nous ignorerons les problèmes de convergence dans ce chapitre ; dans le cadre qui nous intéresse ici, les questions de convergence sont plus gênantes qu'utiles.

Passons maintenant à l'examen de nos principaux outils de manipulation de fonctions génératrices : addition, décalage, changement de variables, dérivation, intégration et multiplication. Dans ce qui suit, nous supposons, sauf indication contraire, que  $F(z)$  et  $G(z)$  sont les fonctions génératrices des suites  $\langle f_n \rangle$  et  $\langle g_n \rangle$ , et que les  $f_n$  et les  $g_n$  sont nuls pour tout  $n$  strictement négatif.

Le résultat d'une combinaison linéaire de  $F$  et  $G$  est tout à fait évident :

$$\begin{aligned}\alpha F(z) + \beta G(z) &= \alpha \sum_n f_n z^n + \beta \sum_n g_n z^n \\ &= \sum_n (\alpha f_n + \beta g_n) z^n.\end{aligned}\tag{7.13}$$

On obtient ainsi la fonction génératrice de la suite  $\langle \alpha f_n + \beta g_n \rangle$ .

Il n'est pas beaucoup plus difficile de décaler une fonction génératrice.

Pour décaler  $G(z)$  de  $m$  places vers la droite, c'est-à-dire pour construire la fonction génératrice de la suite  $\langle 0, \dots, 0, g_0, g_1, \dots \rangle = \langle g_{n-m} \rangle$  qui commence par  $m$  coefficients nuls, il suffit de multiplier par  $z^m$  :

$$z^m G(z) = \sum_n g_n z^{n+m} = \sum_n g_{n-m} z^n, \quad m \geq 0 \text{ entier.} \quad (7.14)$$

C'est exactement l'opération que nous avons appliquée (deux fois), associée à l'addition, pour trouver l'équation  $(1 - z - z^2)F(z) = z$  au chapitre 6, lorsque nous cherchions une forme close pour les nombres de Fibonacci.

A l'inverse, pour décaler  $G(z)$  de  $m$  places vers la gauche, c'est-à-dire pour former la fonction génératrice de la suite  $\langle g_m, g_{m+1}, g_{m+2}, \dots \rangle = \langle g_{n+m} \rangle$ , il faut soustraire les  $m$  premiers termes puis diviser par  $z^m$  :

$$\frac{G(z) - g_0 - g_1 z - \dots - g_{m-1} z^{m-1}}{z^m} = \sum_{n \geq m} g_n z^{n-m} = \sum_{n \geq 0} g_{n+m} z^n. \quad (7.15)$$

Notez que cette dernière somme ne peut pas être étendue à tout  $n$  entier, sauf si  $g_0 = \dots = g_{m-1} = 0$ .

On peut aussi remplacer la variable  $z$  par un des ses multiples :

$$G(cz) = \sum_n g_n (cz)^n = \sum_n c^n g_n z^n. \quad (7.16)$$

On obtient ainsi la fonction génératrice de la suite  $\langle c^n g_n \rangle$ . Le cas particulier  $c = -1$  s'avère particulièrement utile.

On a souvent besoin de multiplier chaque coefficient par un facteur  $n$ . La dérivation est là pour ça :

$$G'(z) = g_1 + 2g_2 z + 3g_3 z^2 + \dots = \sum_n (n+1) g_{n+1} z^n. \quad (7.17)$$

En décalant le tout d'une place vers la droite, on obtient quelque chose qui est parfois plus utile :

$$zG'(z) = \sum_n n g_n z^n. \quad (7.18)$$

C'est la fonction génératrice de la suite  $\langle n g_n \rangle$ . En appliquant plusieurs fois ce type d'opérations, on peut multiplier  $g_n$  par n'importe quel polynôme en  $n$ .

Inversement, l'intégration revient à diviser les termes par  $n$  :

$$\int_0^z G(t) dt = g_0 z + \frac{1}{2} g_1 z^2 + \frac{1}{3} g_2 z^3 + \dots = \sum_{n \geq 1} \frac{1}{n} g_{n-1} z^n. \quad (7.19)$$

Notez que le terme constant est nul. Si on veut la fonction génératrice de  $\langle g_n/n \rangle$  à la place de celle de  $\langle g_{n-1}/n \rangle$ , il faut d'abord effectuer un

décalage d'un cran vers la gauche, pour remplacer  $G(t)$  par  $(G(t) - g_0)/t$  dans l'intégrale.

Pour finir, voici comment multiplier deux fonctions génératrices :

$$\begin{aligned} F(z)G(z) &= (f_0 + f_1z + f_2z^2 + \dots)(g_0 + g_1z + g_2z^2 + \dots) \\ &= (f_0g_0) + (f_0g_1 + f_1g_0)z + (f_0g_2 + f_1g_1 + f_2g_0)z^2 + \dots \\ &= \sum_n \left( \sum_k f_k g_{n-k} \right) z^n. \end{aligned} \quad (7.20)$$

Comme nous l'avons vu au chapitre 5, on obtient ainsi la fonction génératrice d'une suite  $\langle h_n \rangle$  qui est la *convolution* de  $\langle f_n \rangle$  et  $\langle g_n \rangle$ . La somme  $h_n = \sum_k f_k g_{n-k}$  peut aussi s'écrire  $h_n = \sum_{k=0}^n f_k g_{n-k}$ , car  $f_k = 0$  lorsque  $k < 0$  et  $g_{n-k} = 0$  lorsque  $k > n$ . Cette opération de multiplication/convolution est plus compliquée que les précédentes. Elle est toutefois extrêmement utile, à tel point que la section 7.5 lui sera entièrement consacrée.

La multiplication admet plusieurs cas particuliers qui peuvent être considérés comme des opérations à part entière. Nous en avons déjà vu un : si  $F(z) = z^m$ , on se ramène à l'opération de décalage (7.14). Dans ce cas, la somme  $h_n$  est réduite à l'unique terme  $g_{n-m}$ , car tous les  $f_k$  sont nuls sauf  $f_m$  qui vaut 1.

Un autre cas particulier intéressant se présente si  $F(z)$  est la fonction bien connue  $1/(1-z) = 1 + z + z^2 + \dots$ . Tous les  $f_k$  (pour  $k \geq 0$ ) sont égaux à 1 et on aboutit à cette formule importante :

$$\frac{1}{1-z}G(z) = \sum_n \left( \sum_{k \geq 0} g_{n-k} \right) z^n = \sum_n \left( \sum_{k \leq n} g_k \right) z^n. \quad (7.21)$$

En multipliant une fonction génératrice par  $1/(1-z)$ , on obtient la fonction génératrice des sommes partielles de la suite d'origine.

La table 355 récapitule les opérations que nous avons vues jusqu'à présent. Pour bien les appliquer, il est utile de disposer aussi d'un bon répertoire de fonctions génératrices. La table 356 présente les plus simples d'entre elles. Chacune d'entre elles est assez importante pour qu'on se donne la peine de la mémoriser. Beaucoup sont des cas particuliers d'autres fonctions de la table ; beaucoup peuvent aussi se déduire très rapidement des autres. Par conséquent, il n'est pas très difficile de les retenir toutes.

Considérons par exemple la suite  $\langle 1, 2, 3, 4, \dots \rangle$ , dont la fonction génératrice  $1/(1-z)^2$  s'avère souvent utile. Cette fonction, qui est vers le milieu de la table 356, constitue le cas particulier  $m = 1$  de  $\langle 1, \binom{m+1}{m}, \binom{m+2}{m}, \binom{m+3}{m}, \dots \rangle$  qui apparaît plus bas. Elle est aussi le cas particulier  $c = 2$  de la suite  $\langle 1, c, \binom{c+1}{2}, \binom{c+2}{3}, \dots \rangle$ . On peut la trouver en prenant les sommes partielles de la fonction génératrice de  $\langle 1, 1, 1, 1, \dots \rangle$  comme

*Voici un truc : si une suite est composée de coefficients binomiaux, alors sa fonction génératrice contient généralement un binôme  $1 \pm z$ .*

**Table 355** Manipulations de fonctions génératrices.

$\alpha F(z) + \beta G(z) = \sum_n (\alpha f_n + \beta g_n) z^n$	
$z^m G(z) = \sum_n g_{n-m} z^n, \quad m \geq 0 \text{ entier}$	
$\frac{G(z) - g_0 - g_1 z - \cdots - g_{m-1} z^{m-1}}{z^m} = \sum_{n \geq 0} g_{n+m} z^n, \quad m \geq 0 \text{ entier}$	
$G(cz) = \sum_n c^n g_n z^n$	
$G'(z) = \sum_n (n+1) g_{n+1} z^n$	
$z G'(z) = \sum_n n g_n z^n$	
$\int_0^z G(t) dt = \sum_{n \geq 1} \frac{1}{n} g_{n-1} z^n$	
$F(z) G(z) = \sum_n \left( \sum_k f_k g_{n-k} \right) z^n$	
$\frac{1}{1-z} G(z) = \sum_n \left( \sum_{k \leq n} g_k \right) z^n$	

Ça va, ça va, vous  
m'avez convaincu.

en (7.21), c'est-à-dire en divisant  $1/(1-z)$  par  $(1-z)$ . On peut aussi la déduire de  $\langle 1, 1, 1, 1, \dots \rangle$  par dérivation, en appliquant (7.17).

La fonction génératrice de la suite  $\langle 1, 0, 1, 0, \dots \rangle$  peut aussi être obtenue de bien des façons. On peut bien sûr trouver la formule  $\sum_n z^{2n} = 1/(1-z^2)$  en remplaçant  $z$  par  $z^2$  dans l'identité  $\sum_n z^n = 1/(1-z)$ . On peut aussi prendre les sommes partielles de la suite  $\langle 1, -1, 1, -1, \dots \rangle$ , de fonction génératrice  $1/(1+z)$ , pour obtenir  $1/(1+z)(1-z) = 1/(1-z^2)$ . Il existe aussi une troisième façon de faire, basée sur une méthode générale destinée à extraire les termes d'indice pair ( $g_0, 0, g_2, 0, g_4, 0, \dots$ ) de *n'importe quelle* suite donnée. Si on additionne  $G(-z)$  et  $G(+z)$ , on obtient

$$\begin{aligned} G(z) + G(-z) &= \sum_n g_n (1 + (-1)^n) z^n \\ &= 2 \sum_n g_n [n \text{ pair}] z^n. \end{aligned}$$

## 356 FONCTIONS GÉNÉRATRICES

Table 356 Quelques suites simples et leurs fonctions génératrices.

suite	fonction génératrice	forme close
$\langle 1, 0, 0, 0, 0, 0, \dots \rangle$	$\sum_{n \geq 0} [n=0] z^n$	1
$\langle 0, \dots, 0, 1, 0, 0, \dots \rangle$	$\sum_{n \geq 0} [n=m] z^n$	$z^m$
$\langle 1, 1, 1, 1, 1, 1, \dots \rangle$	$\sum_{n \geq 0} z^n$	$\frac{1}{1-z}$
$\langle 1, -1, 1, -1, 1, -1, \dots \rangle$	$\sum_{n \geq 0} (-1)^n z^n$	$\frac{1}{1+z}$
$\langle 1, 0, 1, 0, 1, 0, \dots \rangle$	$\sum_{n \geq 0} [2 n] z^n$	$\frac{1}{1-z^2}$
$\langle 1, 0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots \rangle$	$\sum_{n \geq 0} [m \setminus n] z^n$	$\frac{1}{1-z^m}$
$\langle 1, 2, 3, 4, 5, 6, \dots \rangle$	$\sum_{n \geq 0} (n+1) z^n$	$\frac{1}{(1-z)^2}$
$\langle 1, 2, 4, 8, 16, 32, \dots \rangle$	$\sum_{n \geq 0} 2^n z^n$	$\frac{1}{1-2z}$
$\langle 1, 4, 6, 4, 1, 0, 0, \dots \rangle$	$\sum_{n \geq 0} \binom{4}{n} z^n$	$(1+z)^4$
$\langle 1, c, \binom{c}{2}, \binom{c}{3}, \dots \rangle$	$\sum_{n \geq 0} \binom{c}{n} z^n$	$(1+z)^c$
$\langle 1, c, \binom{c+1}{2}, \binom{c+2}{3}, \dots \rangle$	$\sum_{n \geq 0} \binom{c+n-1}{n} z^n$	$\frac{1}{(1-z)^c}$
$\langle 1, c, c^2, c^3, \dots \rangle$	$\sum_{n \geq 0} c^n z^n$	$\frac{1}{1-cz}$
$\langle 1, \binom{m+1}{m}, \binom{m+2}{m}, \binom{m+3}{m}, \dots \rangle$	$\sum_{n \geq 0} \binom{m+n}{m} z^n$	$\frac{1}{(1-z)^{m+1}}$
$\langle 0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots \rangle$	$\sum_{n \geq 1} \frac{1}{n} z^n$	$\ln \frac{1}{1-z}$
$\langle 0, 1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \dots \rangle$	$\sum_{n \geq 1} \frac{(-1)^{n+1}}{n} z^n$	$\ln(1+z)$
$\langle 1, 1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \frac{1}{120}, \dots \rangle$	$\sum_{n \geq 0} \frac{1}{n!} z^n$	$e^z$

Par conséquent,

$$\frac{G(z) + G(-z)}{2} = \sum_n g_{2n} z^{2n}. \quad (7.22)$$

Il est possible aussi d'extraire les termes d'indice impair. On procède ainsi :

$$\frac{G(z) - G(-z)}{2} = \sum_n g_{2n+1} z^{2n+1}. \quad (7.23)$$

Dans le cas particulier où  $g_n = 1$  et  $G(z) = 1/(1-z)$ , la fonction génératrice de  $\langle 1, 0, 1, 0, \dots \rangle$  est  $\frac{1}{2}(G(z) + G(-z)) = \frac{1}{2}\left(\frac{1}{1-z} + \frac{1}{1+z}\right) = \frac{1}{1-z^2}$ .

Essayons ce nouveau truc sur la série génératrice des nombres de Fibonacci. Nous savons que  $\sum_n F_n z^n = z/(1-z-z^2)$ ; donc

$$\begin{aligned} \sum_n F_{2n} z^{2n} &= \frac{1}{2} \left( \frac{z}{1-z-z^2} + \frac{-z}{1+z-z^2} \right) \\ &= \frac{1}{2} \left( \frac{z+z^2-z^3-z+z^2+z^3}{(1-z^2)^2-z^2} \right) = \frac{z^2}{1-3z^2+z^4}. \end{aligned}$$

C'est la série génératrice de  $\langle F_0, 0, F_2, 0, F_4, \dots \rangle$ . Par conséquent, la suite  $\langle F_0, F_2, F_4, F_6, \dots \rangle = \langle 0, 1, 3, 8, \dots \rangle$  admet une fonction génératrice simple :

$$\sum_n F_{2n} z^n = \frac{z}{1-3z+z^2}. \quad (7.24)$$

### 7.3 RÉSOLUTION DE RÉCURRENCES

Nous allons maintenant concentrer toute notre attention sur l'une des plus importantes applications des fonctions génératrices : la résolution de relations de récurrence.

Etant donnée une suite  $\langle g_n \rangle$  qui satisfait une certaine récurrence, nous cherchons une forme close pour  $g_n$  en fonction de  $n$ . Pour résoudre ce problème à l'aide des fonctions génératrices, il faut procéder en quatre étapes, qui pourrait presque être exécutées automatiquement par un ordinateur :

- 1 Ecrire une unique équation qui exprime  $g_n$  en fonction d'autres éléments de la suite. Cette équation doit être vraie pour tout entier  $n$ , en supposant que  $g_{-1} = g_{-2} = \dots = 0$ .
- 2 Multiplier les deux membres de l'équation par  $z^n$  et sommer sur tout  $n$ . Ceci donne, dans le membre gauche, la somme  $\sum_n g_n z^n$ , donc la fonction génératrice  $G(z)$ . Le membre droit doit être réorganisé de façon à devenir une expression en fonction de  $G(z)$ .
- 3 Résoudre la nouvelle équation pour obtenir une forme close de  $G(z)$ .
- 4 Développer  $G(z)$  en série et prendre le coefficient de  $z^n$ . C'est la forme close de  $g_n$  que nous cherchons.

Ce qui fait fonctionner cette méthode, c'est le fait que  $G(z)$  représente la suite  $\langle g_n \rangle$  dans son entier, sous une forme qui rend les manipulations aisées.

**Exemple 1 : Fibonacci, le retour.**

Nous avons déjà trouvé au chapitre 6, une forme close pour les nombres de Fibonacci. A ce moment là, nous marchions à tâtons car nous apprenions une nouvelle méthode. Maintenant, nous avons les moyens d'opérer systématiquement. La récurrence de départ est :

$$\begin{aligned} g_0 &= 0; \quad g_1 = 1; \\ g_n &= g_{n-1} + g_{n-2}, \quad \text{pour } n \geq 2. \end{aligned}$$

Nous allons trouver une forme close pour  $g_n$  en suivant les quatre étapes décrites ci-dessus.

Dans la première étape, nous devons poser une "unique équation" pour  $g_n$ . Nous pourrions écrire

$$g_n = \begin{cases} 0, & \text{si } n \leq 0; \\ 1, & \text{si } n = 1; \\ g_{n-1} + g_{n-2}, & \text{si } n > 1; \end{cases}$$

mais ce serait de la triche. Il faut trouver une formule sans construction cas par cas. L'équation

$$g_n = g_{n-1} + g_{n-2}$$

marche pour  $n \geq 2$ , et aussi pour  $n \leq 0$  ( $g_n = 0$  pour tout  $n \leq 0$ ). Par contre, si  $n = 1$ , le membre gauche vaut 1 tandis que le membre droit vaut 0. Heureusement, on peut facilement régler ce problème en ajoutant  $[n=1]$  dans le membre droit ; ainsi on lui ajoute 1 si  $n = 1$ , et rien du tout dans les autres cas. Voici donc l'équation demandée à la première étape :

$$g_n = g_{n-1} + g_{n-2} + [n=1].$$

La deuxième étape consiste à transformer l'équation de  $\langle g_n \rangle$  en une équation pour  $G(z) = \sum_n g_n z^n$ . La tâche est aisée :

$$\begin{aligned} G(z) &= \sum_n g_n z^n = \sum_n g_{n-1} z^n + \sum_n g_{n-2} z^n + \sum_n [n=1] z^n \\ &= \sum_n g_n z^{n+1} + \sum_n g_n z^{n+2} + z \\ &= zG(z) + z^2 G(z) + z. \end{aligned}$$

La troisième n'est pas plus compliquée : nous trouvons

$$G(z) = \frac{z}{1 - z - z^2},$$

ce qui n'est pas une surprise.

Nous voici déjà à la quatrième est dernière étape. Au chapitre 6, nous nous en sommes tirés grâce à une inspiration soudaine. Maintenant, nous allons procéder plus doucement. Cela nous permettra de faire cette quatrième étape dans de bonnes conditions quand nous serons confrontés à des problèmes plus difficiles. Que vaut

$$[z^n] \frac{z}{1-z-z^2},$$

le coefficient de  $z^n$  dans le développement en série de  $z/(1-z-z^2)$ ? Posons le problème de façon plus générale : étant donnée une fonction rationnelle, c'est-à-dire une fonction

$$R(z) = \frac{P(z)}{Q(z)},$$

où  $P$  et  $Q$  sont des polynômes, quel est la valeur de  $[z^n] R(z)$ ?

Il existe un type de fonctions rationnelles dont les coefficients sont particulièrement sympathiques :

$$\frac{a}{(1-\rho z)^{m+1}} = \sum_{n \geq 0} \binom{m+n}{m} a \rho^n z^n. \quad (7.25)$$

On trouve le cas  $\rho = 1$  dans la table 356, et la formule plus générale ci-dessus s'obtient en remplaçant  $z$  par  $\rho z$ . Si on considère une somme finie de fonctions du genre de (7.25),

$$S(z) = \frac{a_1}{(1-\rho_1 z)^{m_1+1}} + \frac{a_2}{(1-\rho_2 z)^{m_2+1}} + \cdots + \frac{a_l}{(1-\rho_l z)^{m_l+1}}, \quad (7.26)$$

on trouve aussi des coefficients sympathiques :

$$\begin{aligned} [z^n] S(z) &= a_1 \binom{m_1+n}{m_1} \rho_1^n + a_2 \binom{m_2+n}{m_2} \rho_2^n \\ &\quad + \cdots + a_l \binom{m_l+n}{m_l} \rho_l^n. \end{aligned} \quad (7.27)$$

Nous allons montrer que toute fonction rationnelle  $R(z)$  telle que  $R(0) \neq \infty$  peut s'écrire

$$R(z) = S(z) + T(z), \quad (7.28)$$

où  $S(z)$  est de la forme (7.26) et  $T(z)$  est un polynôme. Par conséquent, il existe une forme close pour les coefficients  $[z^n] R(z)$ . Trouver  $S(z)$  et  $T(z)$  équivaut à trouver le "développement en éléments simples" de  $R(z)$ .

Remarquez que  $S(z) = \infty$  lorsque  $z$  prend les valeurs  $1/\rho_1, \dots, 1/\rho_l$ . Les nombres  $\rho_k$  que nous devons trouver pour écrire  $R(z)$  sous la forme

$S(z) + T(z)$  sont donc les inverses des nombres  $\alpha_k$  tels que  $Q(\alpha_k) = 0$  (souvenez-vous que  $R(z) = P(z)/Q(z)$ , où  $P$  et  $Q$  sont des polynômes, et  $R(z)$  ne peut être égal à  $\infty$  que si  $Q(z) = 0$ ).

Supposons que  $Q(z)$  est de la forme

$$Q(z) = q_0 + q_1 z + \cdots + q_m z^m, \quad \text{où } q_0 \neq 0 \text{ et } q_m \neq 0.$$

On appelle "polynôme réciproque" de  $Q$  le polynôme suivant :

$$Q^R(z) = q_0 z^m + q_1 z^{m-1} + \cdots + q_m.$$

Il est lié à  $Q$  aussi par la relation suivante :

$$\begin{aligned} Q^R(z) &= q_0(z - \rho_1) \dots (z - \rho_m) \\ \iff Q(z) &= q_0(1 - \rho_1 z) \dots (1 - \rho_m z). \end{aligned}$$

Ainsi, les racines de  $Q^R$  sont les inverses des racines de  $Q$ , et vice versa. Pour trouver les nombres  $\rho_k$ , on peut donc factoriser le polynôme réciproque  $Q^R(z)$ .

Par exemple, pour Fibonacci, nous avons

$$Q(z) = 1 - z - z^2; \quad Q^R(z) = z^2 - z - 1.$$

On trouve les racines de  $Q^R$  en posant  $(a, b, c) = (1, -1, -1)$  dans la formule classique  $(-b \pm \sqrt{b^2 - 4ac})/2a$ . Nous trouvons ainsi

$$\phi = \frac{1 + \sqrt{5}}{2} \quad \text{et} \quad \hat{\phi} = \frac{1 - \sqrt{5}}{2},$$

donc  $Q^R(z) = (z - \phi)(z - \hat{\phi})$  et  $Q(z) = (1 - \phi z)(1 - \hat{\phi} z)$ .

Une fois que nous avons trouvé les nombres  $\rho$ , nous pouvons passer à la recherche du développement en éléments simples. Regardons d'abord le cas le plus facile, celui où toutes les racines sont distinctes.

**Théorème : développement rationnel pour des racines distinctes.**

Si  $R(z) = P(z)/Q(z)$ , où  $Q(z) = q_0(1 - \rho_1 z) \dots (1 - \rho_l z)$  et où les nombres  $(\rho_1, \dots, \rho_l)$  sont distincts, et si  $P(z)$  est un polynôme de degré strictement inférieur à  $l$ , alors

$$[z^n] R(z) = a_1 \rho_1^n + \cdots + a_l \rho_l^n, \quad \text{où } a_k = \frac{-\rho_k P(1/\rho_k)}{Q'(1/\rho_k)}. \quad (7.29)$$

Preuve : soient  $a_1, \dots, a_l$  les constantes du théorème. La formule (7.29) est vraie si  $R(z) = P(z)/Q(z)$  est égal à

$$S(z) = \frac{a_1}{1 - \rho_1 z} + \cdots + \frac{a_l}{1 - \rho_l z}.$$

*Pour impressionner parents et amis, laissez donc le livre ouvert à cette page.*

Pour prouver que  $R(z) = S(z)$ , nous allons démontrer que la fonction  $T(z) = R(z) - S(z)$  a une valeur finie lorsque  $z \rightarrow 1/\rho_k$ . En effet, cela entraînera que la fonction rationnelle  $T(z)$  ne prend jamais de valeur infinie, donc qu'elle est polynomiale. Comme, de plus,  $T(z) \rightarrow 0$  lorsque  $z \rightarrow \infty$ ,  $T(z)$  est forcément nul.

Soit  $\alpha_k = 1/\rho_k$ . Pour montrer que  $\lim_{z \rightarrow \alpha_k} T(z) \neq \infty$ , il suffit de montrer que  $\lim_{z \rightarrow \alpha_k} (z - \alpha_k)T(z) = 0$ , car  $T(z)$  est une fonction rationnelle de  $z$ . Il nous faut donc prouver que

$$\lim_{z \rightarrow \alpha_k} (z - \alpha_k)R(z) = \lim_{z \rightarrow \alpha_k} (z - \alpha_k)S(z).$$

Le membre droit est égal à  $\lim_{z \rightarrow \alpha_k} a_k(z - \alpha_k)/(1 - \rho_k z) = -a_k/\rho_k$ , car  $(1 - \rho_k z) = -\rho_k(z - \alpha_k)$  et  $(z - \alpha_k)/(1 - \rho_j z) \rightarrow 0$  pour  $j \neq k$ . Le membre gauche vaut

$$\lim_{z \rightarrow \alpha_k} (z - \alpha_k) \frac{P(z)}{Q(z)} = P(\alpha_k) \lim_{z \rightarrow \alpha_k} \frac{z - \alpha_k}{Q(z)} = \frac{P(\alpha_k)}{Q'(\alpha_k)},$$

d'après la règle de L'Hospital. Le théorème est donc démontré.

Voyons cela sur Fibonacci. On a  $P(z) = z$  et  $Q(z) = 1 - z - z^2 = (1 - \phi z)(1 - \bar{\phi}z)$ , donc  $Q'(z) = -1 - 2z$  et

$$\frac{-\rho P(1/\rho)}{Q'(1/\rho)} = \frac{-1}{-1 - 2/\rho} = \frac{\rho}{\rho + 2}.$$

Par conséquent, d'après (7.29), le coefficient de  $\phi^n$  dans  $[z^n] R(z)$  est  $\phi/(\phi + 2) = 1/\sqrt{5}$ ; le coefficient de  $\bar{\phi}^n$  est  $\bar{\phi}/(\bar{\phi} + 2) = -1/\sqrt{5}$ . Le théorème nous mène bien à  $F_n = (\phi^n - \bar{\phi}^n)/\sqrt{5}$ , comme dans (6.123).

Lorsque  $Q(z)$  a des racines multiples, les calculs sont plus compliqués, mais le principe reste le même et on peut prouver le résultat plus général suivant :

**Théorème : développement rationnel pour des racines multiples.**

Si  $R(z) = P(z)/Q(z)$ , où  $Q(z) = q_0(1 - \rho_1 z)^{d_1} \dots (1 - \rho_l z)^{d_l}$  et où les nombres  $(\rho_1, \dots, \rho_l)$  sont tous distincts, et si  $P(z)$  est un polynôme de degré strictement inférieur à  $d_1 + \dots + d_l$ , alors

$$[z^n] R(z) = f_1(n)\rho_1^n + \dots + f_l(n)\rho_l^n \quad \text{pour tout } n \geq 0, \quad (7.30)$$

où chacun des  $f_k(n)$  est un polynôme de degré  $d_k - 1$  de coefficient directeur

$$\begin{aligned} a_k &= \frac{(-\rho_k)^{d_k} P(1/\rho_k) d_k}{Q^{(d_k)}(1/\rho_k)} \\ &= \frac{P(1/\rho_k)}{(d_k - 1)! q_0 \prod_{j \neq k} (1 - \rho_j/\rho_k)^{d_j}}. \end{aligned} \quad (7.31)$$

Ce résultat peut se démontrer par induction sur  $\max(d_1, \dots, d_l)$ , en utilisant le fait que

$$R(z) - \frac{a_1(d_1 - 1)!}{(1 - \rho_1 z)^{d_1}} - \dots - \frac{a_l(d_l - 1)!}{(1 - \rho_l z)^{d_l}}$$

est une fonction rationnelle dont le dénominateur n'est pas divisible par  $(1 - \rho_k z)^{d_k}$ , et cela pour tout  $k$ .

**Exemple 2 : une récurrence plus ou moins aléatoire.**

Maintenant que nous avons passé en revue quelques méthodes générales, nous sommes fin prêts pour nous attaquer à de nouveaux problèmes. Essayons de trouver une forme close pour la récurrence

$$\begin{aligned} g_0 &= g_1 = 1; \\ g_n &= g_{n-1} + 2g_{n-2} + (-1)^n, \quad \text{pour } n \geq 2. \end{aligned} \tag{7.32}$$

Ne perdons pas nos bonnes habitudes et commençons par faire une table des premières valeurs. Avec la récurrence, c'est immédiat :

n	0	1	2	3	4	5	6	7
$(-1)^n$	1	-1	1	-1	1	-1	1	-1
$g_n$	1	1	4	5	14	23	52	97

Aucune forme close ne nous saute aux yeux. Il va donc nous falloir encore une fois parcourir les quatre étapes du calcul.

La première étape ne pose aucun problème : il y a juste deux choses à ajouter pour que tout fonctionne bien lorsque  $n < 2$  : l'équation

$$g_n = g_{n-1} + 2g_{n-2} + (-1)^n [n \geq 0] + [n = 1]$$

est valable pour tout entier  $n$ . Passons à la deuxième étape :

$$\begin{aligned} G(z) &= \sum_n g_n z^n \\ &= \sum_n g_{n-1} z^n + 2 \sum_n g_{n-2} z^n + \sum_{n \geq 0} (-1)^n z^n + \sum_{n=1} z^n \\ &= zG(z) + 2z^2G(z) + \frac{1}{1+z} + z. \end{aligned}$$

N.B.: la somme  $\sum_{n=1} z^n$  n'a pas d'indice du haut, mais ce n'est pas un oubli !

Notons que nous aurions pu prendre  $\binom{-1}{n}$  au lieu de  $(-1)^n [n \geq 0]$ , pour obtenir  $\sum_n \binom{-1}{n} z^n = (1+z)^{-1}$  avec la formule du binôme. Réglons la troisième étape avec un peu d'algèbre élémentaire :

$$G(z) = \frac{1+z(1+z)}{(1+z)(1-z-2z^2)} = \frac{1+z+z^2}{(1-2z)(1+z)^2}.$$

Il ne reste plus que la quatrième étape.

Le facteur de degré deux dans le dénominateur ne nous arrange pas, car nous savons que les racines multiples sont plus difficiles à gérer que les racines simples. Il faudra cependant faire avec. Nous avons deux racines,  $\rho_1 = 2$  et  $\rho_2 = -1$ . Le théorème (7.30) nous dit qu'il existe une constante  $c$  telle que

$$g_n = a_1 2^n + (a_2 n + c)(-1)^n,$$

avec

$$a_1 = \frac{1 + 1/2 + 1/4}{(1 + 1/2)^2} = \frac{7}{9}; \quad a_2 = \frac{1 - 1 + 1}{1 - 2/(-1)} = \frac{1}{3}.$$

Parmi les deux expressions de  $a_k$  données en (7.31), nous avons préféré ici la seconde. Elle est en effet plus facile à utiliser que la première lorsque le dénominateur a des facteurs simples : il suffit de faire la substitution  $z = 1/\rho_k$  partout dans  $R(z)$ , sauf dans le facteur où cela donne zéro, puis de diviser par  $(d_k - 1)!$ ; on obtient ainsi le coefficient de  $n^{d_k-1} \rho_k^n$ .

Il ne reste plus qu'à poser  $n = 0$  pour trouver que la constante  $c$  doit être égale à  $\frac{2}{9}$ . La réponse à notre problème est donc

$$g_n = \frac{7}{9} 2^n + \left(\frac{1}{3}n + \frac{2}{9}\right)(-1)^n. \quad (7.33)$$

Il n'est pas interdit de vérifier les cas  $n = 1$  et  $2$ , juste pour être sûrs de ne pas nous être trompés dans les calculs ; et aussi peut-être le cas  $n = 3$ , car cette formule est franchement inhabituelle. Tout est correct cependant.

Aurions nous pu conjecturer la formule (7.33) ? Après avoir calculé quelques valeurs de plus, peut-être aurions nous remarqué que  $g_{n+1} \approx 2g_n$  lorsque  $n$  est grand. Avec de la patience et de la chance, nous aurions pu aussi deviner le facteur  $\frac{7}{9}$ . Ce qui est sûr en tout cas, c'est que la méthode des séries génératrices est bien plus sérieuse et efficace.

### *Exemple 3 : Suites mutuellement récursives.*

On est parfois confronté à deux récurrences, ou même plus, qui dépendent les unes des autres. Notre méthode des quatre étapes peut être adaptée à ce genre de problèmes.

Revenons par exemple au problème des pavages  $3 \times n$  que nous avons étudié il y a quelque temps. Si on veut simplement connaître le nombre  $U_n$  de façons de pavier un rectangle  $3 \times n$ , sans différentier les dominos verticaux des dominos horizontaux, on n'a pas besoin de détailler autant que nous l'avons fait alors. Posons simplement les récurrences

$$\begin{aligned} U_0 &= 1, & U_1 &= 0; & V_0 &= 0, & V_1 &= 1; \\ U_n &= 2V_{n-1} + U_{n-2}, & V_n &= U_{n-1} + V_{n-2}, & \text{pour } n \geq 2. \end{aligned}$$

Ici,  $V_n$  est le nombre de façons de pavé un rectangle  $3 \times n$  auquel il manque un coin avec  $(3n - 1)/2$  dominos. Ces récurrences sont faciles à trouver si on considère, comme auparavant, les configurations possibles à l'extrême gauche du rectangle. Voici les premières valeurs de  $U_n$  et  $V_n$  :

$n$	0	1	2	3	4	5	6	7	
$U_n$	1	0	3	0	11	0	41	0	(7.34)
$V_n$	0	1	0	4	0	15	0	56	

Appliquons nos quatre étapes pour en trouver des formes closes. Première étape :

$$U_n = 2V_{n-1} + U_{n-2} + [n=0], \quad V_n = U_{n-1} + V_{n-2},$$

pour tout  $n$ . Deuxième étape :

$$U(z) = 2zV(z) + z^2U(z) + 1, \quad V(z) = zU(z) + z^2V(z).$$

Pour la troisième étape, nous avons deux équations à deux inconnues à résoudre. Cela ne présente aucune difficulté car la seconde équation donne  $V(z) = zU(z)/(1 - z^2)$ . Nous trouvons donc

$$U(z) = \frac{1 - z^2}{1 - 4z^2 + z^4}; \quad V(z) = \frac{z}{1 - 4z^2 + z^4}. \quad (7.35)$$

Nous avions la même formule pour  $U(z)$  en (7.10), mais avec un  $z^3$  à la place du  $z^2$ , car  $n$  représentait alors le nombre de dominos, alors que maintenant c'est la largeur du rectangle.

Le dénominateur  $1 - 4z^2 + z^4$  est une fonction de  $z^2$ ; c'est grâce à cela que  $U_{2n+1} = 0$  et  $V_{2n} = 0$ , exactement comme il faut. Nous pouvons profiter de ce  $z^2$  pour factoriser le dénominateur : pas besoin en effet de factoriser  $1 - 4z^2 + z^4$  en quatre facteurs  $(1 - \rho_k z)$ , puisque deux facteurs de la forme  $(1 - \rho_k z^2)$  suffiront à nous donner les coefficients qu'il nous faut. En d'autres termes, si nous considérons la fonction génératrice

$$W(z) = \frac{1}{1 - 4z + z^2} = W_0 + W_1 z + W_2 z^2 + \dots, \quad (7.36)$$

nous aurons  $V(z) = zW(z^2)$  et  $U(z) = (1 - z^2)W(z^2)$ , donc  $V_{2n+1} = W_n$  et  $U_{2n} = W_n - W_{n-1}$ . Utiliser la fonction  $W(z)$  nous fera donc économiser du temps et de l'énergie.

Les facteurs de  $1 - 4z + z^2$  sont  $(z - 2 - \sqrt{3})$  et  $(z - 2 + \sqrt{3})$ , qui peuvent aussi s'écrire  $(1 - (2 + \sqrt{3})z)$  et  $(1 - (2 - \sqrt{3})z)$ , du fait que ce polynôme est son propre polynôme réciproque. Nous obtenons donc

$$V_{2n+1} = W_n = \frac{3+2\sqrt{3}}{6}(2 + \sqrt{3})^n + \frac{3-2\sqrt{3}}{6}(2 - \sqrt{3})^n;$$

$$\begin{aligned} U_{2n} = W_n - W_{n-1} &= \frac{3+\sqrt{3}}{6}(2+\sqrt{3})^n + \frac{3-\sqrt{3}}{6}(2-\sqrt{3})^n \\ &= \frac{(2+\sqrt{3})^n}{3-\sqrt{3}} + \frac{(2-\sqrt{3})^n}{3+\sqrt{3}}. \end{aligned} \quad (7.37)$$

C'est la forme close que nous cherchions pour le nombre de pavages de rectangles  $3 \times n$  par des dominos.

On peut encore simplifier la formule de  $U_{2n}$  en remarquant que le second terme est toujours entre 0 et 1. Comme  $U_{2n}$  est un nombre entier, on a

$$U_{2n} = \left\lceil \frac{(2+\sqrt{3})^n}{3-\sqrt{3}} \right\rceil, \quad \text{pour } n \geq 0. \quad (7.38)$$

En fait, l'autre terme,  $(2-\sqrt{3})^n/(3+\sqrt{3})$ , est extrêmement petit lorsque  $n$  est grand, car  $2-\sqrt{3} \approx 0,268$ . Ceci doit être pris en compte si on utilise la formule (7.38) pour des calculs numériques. Par exemple, une calculatrice de poche peut donner le résultat 413403,0005 lorsqu'on lui demande de calculer  $(2+\sqrt{3})^{10}/(3-\sqrt{3})$ . Les neuf premiers chiffres sont corrects, mais la vraie valeur est légèrement *plus petite* que 413403, et non plus grande. Ce serait donc une erreur de prendre la partie entière supérieure de 413403,0005. La réponse correcte,  $U_{20} = 413403$ , s'obtient en arrondissant à l'entier le plus proche. La partie entière inférieure est parfois traîtresse.

*La partie entière supérieure n'a rien à lui envier.*

#### **Exemple 4 : une forme close pour la monnaie.**

Nous n'avons pas terminé notre problème de monnaie : nous nous sommes contentés de calculer le nombre de façons de payer 50 cents. Essayons maintenant de trouver le nombre de façons de payer un dollar, ou un million de dollars, toujours avec des pièces de 1, 5, 10, 25 et 50 cents.

Nous connaissons déjà la fonction génératrice,

$$C(z) = \frac{1}{1-z} \frac{1}{1-z^5} \frac{1}{1-z^{10}} \frac{1}{1-z^{25}} \frac{1}{1-z^{50}},$$

qui est rationnelle et dont le dénominateur est de degré 91. Nous pourrions donc décomposer ce dénominateur en 91 facteurs et obtenir ainsi une "forme close" à 91 termes pour  $C_n$ . Ce serait plutôt cauchemardesque, non ? Essayons donc, pour ce cas particulier, de faire mieux que ce que la méthode générale nous propose.

Voici une lueur d'espoir : le dénominateur est presque une fonction de  $z^5$ . Si nous remplaçons  $1/(1-z)$  par  $(1+z+z^2+z^3+z^4)/(1-z^5)$ , nous allons donc pouvoir simplifier les calculs, comme nous l'avons fait il n'y a pas si longtemps lorsque nous avons remarqué que  $1-4z^2+z^4$  était une

fonction de  $z^2$  :

$$\begin{aligned} C(z) &= \frac{1+z+z^2+z^3+z^4}{1-z^5} \frac{1}{1-z^5} \frac{1}{1-z^{10}} \frac{1}{1-z^{25}} \frac{1}{1-z^{50}} \\ &= (1+z+z^2+z^3+z^4)\check{C}(z^5), \\ \check{C}(z) &= \frac{1}{1-z} \frac{1}{1-z} \frac{1}{1-z^2} \frac{1}{1-z^5} \frac{1}{1-z^{10}}. \end{aligned}$$

La fonction  $\check{C}(z)$  a un dénominateur de degré 19 seulement ; c'est un grand progrès. De plus, cette nouvelle expression de  $C(z)$  nous montre que  $C_{5n} = C_{5n+1} = C_{5n+2} = C_{5n+3} = C_{5n+4}$ . A posteriori, ce n'est pas étonnant du tout : il y a autant de façons de payer 53 cents que d'en payer 50, car le nombre de pièces de 1 cent est prédéterminé modulo 5.

Cependant, la forme close que nous pouvons trouver pour  $\check{C}(z)$ , basée sur les racines de son dénominateur, n'est toujours pas vraiment simple. Voyons quand même ce que cela donne. Remarquons que tous les facteurs de ce dénominateur sont des diviseurs de  $1 - z^{10}$ . Il est donc probable que la meilleure façon de calculer les coefficients de  $\check{C}(z)$  soit celle qui consiste à écrire

$$\check{C}(z) = \frac{A(z)}{(1-z^{10})^5}, \quad \text{où } A(z) = A_0 + A_1 z + \cdots + A_{31} z^{31}. \quad (7.39)$$

Voici, pour les curieux, la valeur de  $A(z)$  :

$$\begin{aligned} &(1+z+\cdots+z^9)^2(1+z^2+\cdots+z^8)(1+z^5) \\ &= 1 + 2z + 4z^2 + 6z^3 + 9z^4 + 13z^5 + 18z^6 + 24z^7 \\ &\quad + 31z^8 + 39z^9 + 45z^{10} + 52z^{11} + 57z^{12} + 63z^{13} + 67z^{14} + 69z^{15} \\ &\quad + 69z^{16} + 67z^{17} + 63z^{18} + 57z^{19} + 52z^{20} + 45z^{21} + 39z^{22} + 31z^{23} \\ &\quad + 24z^{24} + 18z^{25} + 13z^{26} + 9z^{27} + 6z^{28} + 4z^{29} + 2z^{30} + z^{31}. \end{aligned}$$

Pour finir, comme  $1/(1-z^{10})^5 = \sum_{k \geq 0} \binom{k+4}{4} z^{10k}$ , nous pouvons calculer les coefficients  $\check{C}_n = [z^n] \check{C}(z)$  comme suit, lorsque  $n = 10q+r$  et  $0 \leq r < 10$  :

$$\begin{aligned} \check{C}_{10q+r} &= \sum_{j,k} A_j \binom{k+4}{4} [10q+r=10k+j] \\ &= A_r \binom{q+4}{4} + A_{r+10} \binom{q+3}{4} + A_{r+20} \binom{q+2}{4} + A_{r+30} \binom{q+1}{4}. \quad (7.40) \end{aligned}$$

Nous avons donc dix cas, un pour chaque valeur de  $r$ . C'est cependant une forme close tout à fait convenable par rapport à d'autres qui font apparaître des puissances de nombres complexes.

*Attention, raisonnement rapide.*

Appliquons par exemple cette expression pour calculer  $C_{50q} = \check{C}_{10q}$ . Dans ce cas,  $r = 0$  et

$$C_{50q} = \binom{q+4}{4} + 45\binom{q+3}{4} + 52\binom{q+2}{4} + 2\binom{q+1}{4}.$$

On peut payer 50 cents de  $\binom{5}{4} + 45\binom{4}{4} = 50$  façons différentes ; on peut payer 1 dollar de  $\binom{6}{4} + 45\binom{5}{4} + 52\binom{4}{4} = 292$  façons différentes ; on peut payer un million de dollars de

$$\binom{2000004}{4} + 45\binom{2000003}{4} + 52\binom{2000002}{4} + 2\binom{2000001}{4} \\ = 6666679333412666685000001$$

façons différentes (toujours avec des pièces de 1, 5, 10, 25 et 50 cents).

#### *Exemple 5 : une série divergente.*

Cherchons une forme close des nombres  $g_n$  définis par

$$g_0 = 1; \\ g_n = ng_{n-1}, \quad \text{pour } n > 0.$$

*A l'heure actuelle,  
on parle plutôt de  
femtosecondes.*

Quelques nanosecondes suffisent pour réaliser que  $g_n$  vaut exactement  $n!$ . La méthode des facteurs de sommation du chapitre 2 fournit immédiatement cette réponse. Essayons quand même, juste pour voir, de résoudre ce problème avec les fonctions génératrices. Si cette technique est aussi puissante que nous le prétendons, elle doit pouvoir résoudre les récurrences faciles aussi bien que les autres !

L'équation

$$g_n = ng_{n-1} + [n=0],$$

qui est vraie pour tout  $n$ , entraîne que

$$G(z) = \sum_n g_n z^n = \sum_n ng_{n-1} z^n + \sum_{n=0} z^n.$$

Pour terminer la deuxième étape, il nous faut exprimer  $\sum_n ng_{n-1} z^n$  en fonction de  $G(z)$ . Les manœuvres de base de la table 355 tendent à suggérer que la dérivée  $G'(z) = \sum_n ng_n z^{n-1}$  pourrait y jouer un rôle. Dirigeons donc nos efforts pour aboutir à quelque chose de ce genre :

$$G(z) = 1 + \sum_n (n+1)g_n z^{n+1} = 1 + \sum_n ng_n z^{n+1} + \sum_n g_n z^{n+1} \\ = 1 + z^2 G'(z) + zG(z).$$

Vérifions cette équation pour les petites valeurs de  $n$ . Comme

$$\begin{aligned} G &= 1 + z + 2z^2 + 6z^3 + 24z^4 + \dots, \\ G' &= 1 + 4z + 18z^2 + 96z^3 + \dots, \end{aligned}$$

nous avons

$$\begin{aligned} z^2 G' &= \dots, z^2 + 4z^3 + 18z^4 + 96z^5 + \dots, \\ zG &= z + z^2 + 2z^3 + 6z^4 + 24z^5 + \dots, \\ 1 &= 1. \end{aligned}$$

La somme de ces trois dernières lignes donne bien  $G$ . Remarquez que nous avons écrit “ $G$ ” pour “ $G(z)$ ”; cela nous arrivera souvent, car le  $(z)$  ne sert qu’à encombrer les formules, sauf dans les cas où on modifie le paramètre  $z$ .

Pour la troisième étape, les calculs vont être différents de ceux que nous avons faits jusqu’à présent. Nous avons en effet une équation différentielle à résoudre. Par bonheur, nous pouvons en venir à bout avec les techniques à base de séries hypergéométriques que nous avons vues en section 5.6 (que les lecteurs non familiers avec les séries hypergéométriques ne s’inquiètent pas, ce sera court).

Commençons par nous débarrasser de la constante 1 en dérivant les deux membres de l’équation :

$$\begin{aligned} G' &= (z^2 G' + zG + 1)' = (2zG' + z^2 G'') + (G + zG') \\ &= z^2 G'' + 3zG' + G. \end{aligned}$$

Selon la théorie développée au chapitre 5, nous devons maintenant utiliser l’opérateur  $\vartheta$ . L’exercice 6.13 nous dit que

$$\vartheta G = zG', \quad \vartheta^2 G = z^2 G'' + zG'.$$

Nous en déduisons une forme adéquate pour notre équation différentielle :

$$\vartheta G = z\vartheta^2 G + 2z\vartheta G + zG = z(\vartheta + 1)^2 G.$$

Selon (5.109), la solution telle que  $g_0 = 1$  est la série hypergéométrique  $F(1, 1; ; z)$ .

Maintenant que nous connaissons la fonction  $G$ , la quatrième étape est particulièrement facile. D’après la définition des séries hypergéométriques (5.76), nous avons :

$$G(z) = F\left(\begin{matrix} 1, 1 \\ \end{matrix} \middle| z\right) = \sum_{n \geq 0} \frac{1^n 1^n z^n}{n!} = \sum_{n \geq 0} n! z^n.$$

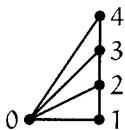
*“Ce sera court”.  
C'est ce que le  
médecin a dit juste  
avant de me piquer  
avec sa seringue. A  
la réflexion, “hy-  
pergéométrique”  
sonne un peu comme  
“hypodermique”,  
non ?*

Cela confirme bien la forme close que nous connaissons depuis longtemps,  $g_n = n!$ .

Notez bien que la technique que nous venons d'employer a donné la bonne réponse, alors que  $G(z)$  diverge pour tout  $z$  non nul. La suite  $n!$  croît tellement vite que, sauf si  $z = 0$ , les termes  $|n! z^n|$  tendent vers  $\infty$  lorsque  $n \rightarrow \infty$ . Ceci illustre bien le fait qu'on peut effectuer des manipulations algébriques sur les séries sans se préoccuper un seul instant de convergence.

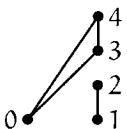
#### **Exemple 6 : une récurrence non bornée.**

Terminons cette section en appliquant les fonctions génératrices à un problème de théorie des graphes. Un éventail d'ordre  $n$  est un graphe à  $n + 1$  sommets  $\{0, 1, \dots, n\}$  et  $2n - 1$  arêtes définies comme suit : le sommet 0 est relié par une arête à chacun des  $n$  autres sommets, et, pour tout  $1 \leq k < n$ , le sommet  $k$  est relié au sommet  $k + 1$ . Voici par exemple l'éventail d'ordre 4, qui a cinq sommets et sept arêtes :



Voici le problème qui nous intéresse : combien existe-t-il d'arbres couvrants dans un tel graphe ? Un arbre couvrant est un sous-graphe connexe et sans cycle qui contient tous les sommets du graphe. On trouvera dans tout livre de théorie des graphes la propriété suivante : un arbre couvrant d'un graphe à  $n + 1$  sommets a exactement  $n$  arêtes ; s'il en avait moins il ne serait pas connexe, et s'il en avait plus il contiendrait au moins un cycle.

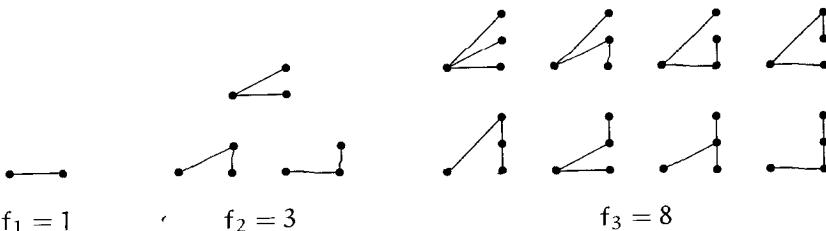
Il y a  $\binom{2n-1}{n}$  façons de choisir  $n$  arêtes parmi les  $2n - 1$  qui sont présentes dans un éventail d'ordre  $n$ , mais tous les configurations ainsi obtenues ne sont pas forcément des arbres couvrants. Par exemple, le sous-graphe



a quatre arêtes mais n'est pas un arbre couvrant : il contient le cycle  $0, 4, 3, 0$ , et il n'y a aucune arête entre  $\{1, 2\}$  et les autres sommets. Nous voulons compter le nombre de choix, parmi les  $\binom{2n-1}{n}$  possibles, qui donnent vraiment lieu à un arbre couvrant.

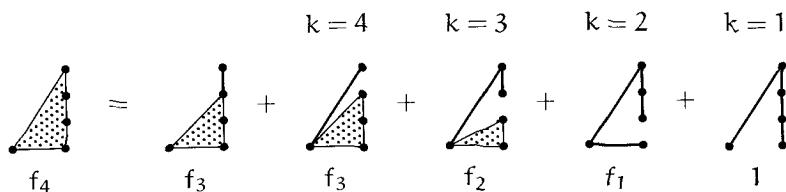
Voyons quelques petits cas. Compter les arbres couvrants pour  $n = 1, 2$  et 3 ne présente aucune difficulté. Les voici donc, avec la convention que les sommets sont toujours numérotés dans l'ordre donné dans les exemples

précédents, ce qui évite d'écrire les numéros à chaque fois :



Que dire du cas  $n = 0$ ? *A priori*, il semble raisonnable de poser  $f_0 = 1$ ; nous prendrons néanmoins  $f_0 = 0$ , car l'existence d'un éventail d'ordre 0 est assez douteuse (il aurait  $2n - 1 = -1$  arête).

Pour suivre notre processus en quatre étapes, nous devons d'abord trouver une récurrence pour  $f_n$  valable pour tout  $n$ . Pour cela, examinons comment le sommet du haut (le numéro  $n$ ) est connecté au reste de l'arbre couvrant. S'il n'est pas relié au sommet 0, il doit l'être au sommet  $n - 1$  pour que le sous-graphe soit connexe. Dans ce cas, chacun des  $f_{n-1}$  arbres couvrants du reste de l'éventail (sur les sommets 0 à  $n - 1$ ) donne lieu à un arbre couvrant du graphe entier. Dans l'autre cas, le sommet  $n$  est relié à 0, et il existe un nombre  $k \leq n$  tel que les sommets  $n, n - 1, \dots, k$  sont directement connectés mais qu'il n'y a pas d'arête entre  $k$  et  $k - 1$ . Alors il ne peut pas y avoir d'arête entre 0 et  $\{n - 1, \dots, k\}$ , sinon le sous-graphe contiendrait un cycle. Si  $k = 1$ , l'arbre couvrant est complètement déterminé. Si  $k > 1$ , chacune des  $f_{k-1}$  façons de produire un arbre couvrant de  $\{0, 1, \dots, k - 1\}$  donne lieu à un arbre couvrant du graphe entier. Voici par exemple le résultat de ce raisonnement dans le cas  $n = 4$ :



L'équation générale, valable pour tout  $n \geq 1$ , est la suivante :

$$f_n = f_{n-1} + f_{n-1} + f_{n-2} + f_{n-3} + \cdots + f_1 + 1.$$

Le "1" tout à la fin semble nous dire que nous aurions mieux fait de choisir  $f_0 = 1$ ; nous resterons cependant obstinément fidèles à notre choix initial. Pour que l'équation soit valide pour tout  $n$  entier, une petite modification suffit :

$$f_n = f_{n-1} + \sum_{k < n} f_k + [n > 0]. \quad (7.41)$$

Cette récurrence est d'un genre différent de celles que nous avons étudiées jusqu'ici dans ce chapitre : la valeur de  $f_n$  dépend des valeurs de tous les  $f_k$  précédemment calculés. Nous disons qu'elle est "non bornée" car elle dépend de paramètres en nombre non borné. Au chapitre 2, nous avons utilisé une méthode particulière pour venir à bout d'une récurrence similaire, celle du tri rapide (2.12) : nous avons effectué une soustraction de deux de ses instances,  $f_{n+1}$  et  $f_n$ . Cette astuce marcherait aussi bien dans le cas présent, mais nous nous en passerons car les fonctions génératrices nous permettent de manipuler directement la récurrence non bornée. C'est d'ailleurs heureux, car nous verrons d'ici peu des récurrences beaucoup plus difficiles que celle-ci.

La première étape est terminée. Pour la deuxième, nous allons faire quelque chose d'inédit :

$$\begin{aligned} F(z) &= \sum_n f_n z^n = \sum_n f_{n-1} z^n + \sum_{k,n} f_k z^n [k < n] + \sum_n [n > 0] z^n \\ &= zF(z) + \sum_k f_k z^k \sum_n [n > k] z^{n-k} + \frac{z}{1-z} \\ &= zF(z) + F(z) \sum_{m>0} z^m + \frac{z}{1-z} \\ &= zF(z) + F(z) \frac{z}{1-z} + \frac{z}{1-z}. \end{aligned}$$

Le point principal de ce calcul réside dans le fait d'avoir réécrit  $z^n$  en  $z^k z^{n-k}$ . C'est ce qui nous a permis d'exprimer la somme double en fonction de  $F(z)$ , comme nous l'impose la deuxième étape.

La troisième étape se résume à de l'algèbre élémentaire, et nous trouvons

$$F(z) = \frac{z}{1 - 3z + z^2}.$$

Ceux d'entre nous qui ont un peu de mémoire ont tout de suite reconnu la fonction génératrice (7.24) des nombres de Fibonacci d'indice pair. Inutile donc d'aller plus loin ; nous avons trouvé une réponse plutôt surprenante au problème du recouvrement d'éventails :

$$f_n = F_{2n}, \quad \text{pour } n \geq 0. \tag{7.42}$$

## 7.4 FG REMARQUABLES

La quatrième étape de notre procédure devient bien plus facile si on connaît les coefficients d'un grand nombre de séries entières. Les développements de séries de la table 356 sont déjà bien utiles, mais il existe bien

d'autres types de formes closes. C'est pourquoi nous y ajoutons une table qui donne la liste des séries correspondant aux "nombres remarquables" du chapitre 6.

La table 373 est la base de données qu'il nous faut ; elle est conçue pour servir de référence lorsqu'on rencontre un problème nouveau. Nous ne nous étendrons pas sur les identités qu'on y trouve, car elles sont faciles à démontrer. Toutefois, la première formule, (7.44), a une preuve qui vaut la peine d'être vue : partons de l'identité

$$\frac{1}{(1-z)^{x+1}} = \sum_n \binom{x+n}{n} z^n$$

et dérivons la par rapport à  $x$ . Dans le membre gauche,  $(1-z)^{-x-1}$  est égal à  $e^{(x+1)\ln(1/(1-z))}$ , ce qui donne le facteur  $\ln(1/(1-z))$  de la dérivée. Dans le membre droit, le numérateur de  $\binom{x+n}{n}$  est  $(x+n) \dots (x+1)$  et, en dérivant, on obtient  $n$  termes dont la somme peut s'exprimer en multipliant  $\binom{x+n}{n}$  par

$$\frac{1}{x+n} + \dots + \frac{1}{x+1} = H_{x+n} - H_x.$$

En remplaçant  $x$  par  $m$ , on trouve (7.44). Notez que  $H_{x+n} - H_x$  a un sens même si  $x$  n'est pas un entier.

Cette méthode — dériver un produit compliqué et laisser le résultat sous forme de produit — est généralement meilleure que celle consistant à exprimer la dérivée comme une somme. Par exemple, le membre droit de

$$\frac{d}{dx} ((x+n)^n \dots (x+1)^1) = (x+n)^n \dots (x+1)^1 \left( \frac{n}{x+n} + \dots + \frac{1}{x+1} \right)$$

serait bien plus compliqué si on l'écrivait comme une somme.

Les identités de la table 373 comportent un bon nombre de cas particuliers importants. Par exemple, si on pose  $m = 0$ , (7.44) donne la fonction génératrice de  $H_n$  :

$$\frac{1}{1-z} \ln \frac{1}{1-z} = \sum_n H_n z^n. \quad (7.43)$$

Il y a aussi d'autres manières d'obtenir cette équation. Par exemple, on peut prendre le développement en série de  $\ln(1/(1-z))$  et le diviser par  $1-z$  pour obtenir une somme de sommes partielles.

Les identités (7.52) et (7.53) font apparaître respectivement les rapports  $\{\frac{m}{m-n}\}/(\frac{m-1}{n})$  et  $[\frac{m}{m-n}]/(\frac{m-1}{n})$ , qui donnent tous deux l'expression indéfinie  $0/0$  si  $n \geq m$ . Il y a toutefois un moyen de leur donner un sens en

**Table 373** Fonctions génératrices de nombres remarquables.

$$\frac{1}{(1-z)^{m+1}} \ln \frac{1}{1-z} = \sum_{n \geq 0} (H_{m+n} - H_m) \binom{m+n}{n} z^n \quad (7.44)$$

$$\frac{z}{e^z - 1} = \sum_{n \geq 0} B_n \frac{z^n}{n!} \quad . \quad (7.45)$$

$$\frac{F_m z}{1 - (F_{m-1} + F_{m+1})z + (-1)^m z^2} = \sum_{n \geq 0} F_{mn} z^n \quad (7.46)$$

$$\sum_k \left\{ \begin{matrix} m \\ k \end{matrix} \right\} \frac{k! z^k}{(1-z)^{k+1}} = \sum_{n \geq 0} n^m z^n \quad (7.47)$$

$$(z^{-1})^{\overline{-m}} = \frac{z^m}{(1-z)(1-2z)\dots(1-mz)} = \sum_{n \geq 0} \left\{ \begin{matrix} n \\ m \end{matrix} \right\} z^n \quad (7.48)$$

$$z^{\overline{m}} = z(z+1)\dots(z+m-1) = \sum_{n \geq 0} \left[ \begin{matrix} m \\ n \end{matrix} \right] z^n \quad (7.49)$$

$$(e^z - 1)^m = m! \sum_{n \geq 0} \left\{ \begin{matrix} n \\ m \end{matrix} \right\} \frac{z^n}{n!} \quad (7.50)$$

$$\left( \ln \frac{1}{1-z} \right)^m = m! \sum_{n \geq 0} \left[ \begin{matrix} n \\ m \end{matrix} \right] \frac{z^n}{n!} \quad (7.51)$$

$$\left( \frac{z}{\ln(1+z)} \right)^m = \sum_{n \geq 0} \frac{z^n}{n!} \left\{ \begin{matrix} m \\ m-n \end{matrix} \right\} \Big/ \binom{m-1}{n} \quad (7.52)$$

$$\left( \frac{z}{1-e^{-z}} \right)^m = \sum_{n \geq 0} \frac{z^n}{n!} \left[ \begin{matrix} m \\ m-n \end{matrix} \right] \Big/ \binom{m-1}{n} \quad (7.53)$$

$$e^{z+wz} = \sum_{m,n \geq 0} \left( \begin{matrix} n \\ m \end{matrix} \right) w^m \frac{z^n}{n!} \quad (7.54)$$

$$e^{w(e^z-1)} = \sum_{m,n \geq 0} \left\{ \begin{matrix} n \\ m \end{matrix} \right\} w^m \frac{z^n}{n!} \quad (7.55)$$

$$\frac{1}{(1-z)^w} = \sum_{m,n \geq 0} \left[ \begin{matrix} n \\ m \end{matrix} \right] w^m \frac{z^n}{n!} \quad (7.56)$$

$$\frac{1-w}{e^{(w-1)z}-w} = \sum_{m,n \geq 0} \left\langle \begin{matrix} n \\ m \end{matrix} \right\rangle w^m \frac{z^n}{n!} \quad (7.57)$$

utilisant les polynômes de Stirling de (6.45). On a en effet

$$\left\{ \begin{matrix} m \\ m-n \end{matrix} \right\} / \binom{m-1}{n} = (-1)^{n+1} n! m \sigma_n(m-n); \quad (7.58)$$

$$\left[ \begin{matrix} m \\ m-n \end{matrix} \right] / \binom{m-1}{n} = n! m \sigma_n(m). \quad (7.59)$$

Ainsi, par exemple, le cas  $m = 1$  de (7.52) ne doit pas être vu comme la série  $\sum_{n \geq 0} (z^n/n!) \left\{ \begin{matrix} 1 \\ 1-n \end{matrix} \right\} / \binom{0}{n}$ , mais comme

$$\begin{aligned} \frac{z}{\ln(1+z)} &= - \sum_{n \geq 0} (-z)^n \sigma_n(n-1) \\ &= 1 + \frac{1}{2}z - \frac{1}{12}z^2 + \dots \end{aligned}$$

Les fonctions des identités (7.54), (7.55), (7.56) et (7.57) sont appelées des "fonctions génératrices doubles" car elles peuvent s'écrire sous la forme  $G(w, z) = \sum_{m,n} g_{m,n} w^m z^n$ . Le coefficient de  $w^m$  est une fonction génératrice en  $z$  et le coefficient de  $z^n$  est une fonction génératrice en  $w$ . L'équation (7.57) peut s'écrire de façon plus symétrique :

$$\frac{e^w - e^z}{we^z - ze^w} = \sum_{m,n \geq 0} \left\langle \begin{matrix} m+n+1 \\ m \end{matrix} \right\rangle \frac{w^m z^n}{(m+n+1)!}. \quad (7.60)$$

## 7.5 CONVOLUTIONS

La *convolution* de deux suites données  $\langle f_0, f_1, \dots \rangle = \langle f_n \rangle$  et  $\langle g_0, g_1, \dots \rangle = \langle g_n \rangle$  est la suite  $\langle f_0 g_0, f_0 g_1 + f_1 g_0, \dots \rangle = \langle \sum_k f_k g_{n-k} \rangle$ . Nous avons observé dans les sections 5.4 et 7.2 qu'une convolution de deux suites équivaut à la multiplication de leurs fonctions génératrices. Grâce à cela, on peut calculer beaucoup de sommes qui auraient été difficiles à manipuler autrement.

**Exemple 1 : une convolution de Fibonacci.**

Essayons par exemple de trouver une forme close pour  $\sum_{k=0}^n F_k F_{n-k}$ . Comme c'est la convolution de la suite  $\langle F_n \rangle$  avec elle-même, la somme représente le coefficient de  $z^n$  dans  $F(z)^2$ , où  $F(z)$  est la fonction génératrice de  $\langle F_n \rangle$ . Tout ce que nous avons à faire est de calculer la valeur de ce coefficient.

La fonction génératrice de  $F(z)$  est  $z/(1-z-z^2)$ . D'après le théorème de développement des fonctions génératrices rationnelles (7.30), le problème peut être résolu par une représentation en éléments simples. Il y a aussi

J'ai toujours cru qu'une convolution était ce à quoi ressemble mon cerveau quand j'essaie de faire une preuve.

une autre façon de faire, qui consiste à écrire

$$\begin{aligned} F(z)^2 &= \left( \frac{1}{\sqrt{5}} \left( \frac{1}{1-\phi z} - \frac{1}{1-\bar{\phi}z} \right) \right)^2 \\ &= \frac{1}{5} \left( \frac{1}{(1-\phi z)^2} - \frac{2}{(1-\phi z)(1-\bar{\phi}z)} + \frac{1}{(1-\bar{\phi}z)^2} \right) \\ &= \frac{1}{5} \sum_{n \geq 0} (n+1)\phi^n z^n - \frac{2}{5} \sum_{n \geq 0} F_{n+1} z^n + \frac{1}{5} \sum_{n \geq 0} (n+1)\bar{\phi}^n z^n. \end{aligned}$$

Essayons d'exprimer la réponse en termes de nombres de Fibonacci plutôt qu'en fonction de  $\phi$  et  $\bar{\phi}$ . Comme  $\phi + \bar{\phi} = 1$ , on a

$$\begin{aligned} \phi^n + \bar{\phi}^n &= [z^n] \left( \frac{1}{1-\phi z} + \frac{1}{1-\bar{\phi}z} \right) \\ &= [z^n] \frac{2 - (\phi + \bar{\phi})z}{(1-\phi z)(1-\bar{\phi}z)} = [z^n] \frac{2-z}{1-z-z^2} = 2F_{n+1} - F_n. \end{aligned}$$

Par conséquent,

$$F(z)^2 = \frac{1}{5} \sum_{n \geq 0} (n+1)(2F_{n+1} - F_n) z^n - \frac{2}{5} \sum_{n \geq 0} F_{n+1} z^n,$$

et voici ce que nous cherchons :

$$\sum_{k=0}^n F_k F_{n-k} = \frac{2nF_{n+1} - (n+1)F_n}{5}. \quad (7.61)$$

Par exemple, pour  $n = 3$ , cette formule donne  $F_0 F_3 + F_1 F_2 + F_2 F_1 + F_3 F_0 = 0 + 1 + 1 + 0 = 2$  dans le membre gauche et  $(6F_4 - 4F_3)/5 = (18 - 8)/5 = 2$  dans le membre droit.

### **Exemple 2 : convolutions harmoniques.**

Il existe une méthode de tri par ordinateur dont l'efficacité dépend de la valeur de la somme

$$T_{m,n} = \sum_{0 \leq k \leq n} \binom{k}{m} \frac{1}{n-k}, \quad m, n \geq 0 \text{ entiers.}$$

Dans l'exercice 5.58, on l'obtient au moyen d'une double induction quelque peu compliquée, avec des facteurs de sommation. Cela devient bien plus facile si on réalise que  $T_{m,n}$  n'est rien d'autre que le  $n$ ième terme de la convolution des suites  $\langle \binom{0}{m}, \binom{1}{m}, \binom{2}{m}, \dots \rangle$  et  $\langle 0, \frac{1}{1}, \frac{1}{2}, \dots \rangle$ . Elles ont chacune une

A l'attention des informaticiens : cette méthode, que les Américains appellent "samplesort", est une variante du tri rapide, dans laquelle le pivot est choisi par échantillonnage (N.d.T.).

fonction génératrice simple, que l'on peut trouver dans la table 356 :

$$\sum_{n \geq 0} \binom{n}{m} z^n = \frac{z^m}{(1-z)^{m+1}};$$

$$\sum_{n > 0} \frac{z^n}{n} = \ln \frac{1}{1-z}.$$

Par conséquent, d'après (7.44),

$$\begin{aligned} T_{m,n} &= [z^n] \frac{z^m}{(1-z)^{m+1}} \ln \frac{1}{1-z} \\ &= [z^{n-m}] \frac{1}{(1-z)^{m+1}} \ln \frac{1}{1-z} \\ &= (H_n - H_m) \binom{n}{n-m}. \end{aligned}$$

En fait, beaucoup d'autres sommes donnent lieu au même genre de convolution, car

$$\frac{1}{(1-z)^{r+1}} \ln \frac{1}{1-z} \cdot \frac{1}{(1-z)^{s+1}} = \frac{1}{(1-z)^{r+s+2}} \ln \frac{1}{1-z}$$

pour tous  $r$  et  $s$ . En mettant en équation les coefficients de  $z^n$  dans les deux membres, on trouve l'identité générale

$$\begin{aligned} \sum_k \binom{r+k}{k} \binom{s+n-k}{n-k} (H_{r+k} - H_r) \\ = \binom{r+s+n+1}{n} (H_{r+s+n+1} - H_{r+s+1}). \end{aligned} \quad (7.62)$$

C'est presque trop beau pour être vrai ; mais ça marche, au moins pour  $n = 2$  : *C'est tellement harmonieux !*

$$\begin{aligned} \binom{r+1}{1} \binom{s+1}{1} \frac{1}{r+1} + \binom{r+2}{2} \binom{s+0}{0} \left( \frac{1}{r+2} + \frac{1}{r+1} \right) \\ = \binom{r+s+3}{2} \left( \frac{1}{r+s+3} + \frac{1}{r+s+2} \right). \end{aligned}$$

Certains cas particuliers, comme  $s = 0$ , sont aussi remarquables que le cas général.

Il y a mieux. Comme  $H_r$  est indépendant de  $k$ , on peut utiliser l'identité

$$\sum_k \binom{r+k}{k} \binom{s+n-k}{n-k} = \binom{r+s+n+1}{n}$$

pour faire passer  $H_r$  du membre gauche au membre droit :

$$\begin{aligned} \sum_k \binom{r+k}{k} \binom{s+n-k}{n-k} H_{r+k} \\ = \binom{r+s+n+1}{n} (H_{r+s+n+1} - H_{r+s+1} + H_r). \end{aligned} \quad (7.63)$$

Il y a même encore mieux. Si  $r$  et  $s$  sont des entiers positifs ou nuls  $l$  et  $m$ , on peut remplacer  $\binom{r+k}{k}$  par  $\binom{l+k}{l}$  et  $\binom{s+n-k}{n-k}$  par  $\binom{m+n-k}{m}$ , puis  $k$  par  $k-l$  et  $n$  par  $n-m-l$ , pour obtenir

$$\sum_{k=0}^n \binom{k}{l} \binom{n-k}{m} H_k = \binom{n+1}{l+m+1} (H_{n+1} - H_{l+m+1} + H_l), \quad l, m, n \geq 0 \text{ entiers.} \quad (7.64)$$

Souvenez-vous, au chapitre 2 nous avions des difficultés à manipuler le cas particulier  $l = m = 0$  de cette identité (voir (2.36).) ! Nous avons parcouru un bon bout de chemin depuis.

### **Exemple 3 : convolutions de convolutions.**

Si on fait une convolution de deux suites  $\langle f_n \rangle$  et  $\langle g_n \rangle$ , puis de nouveau une convolution du résultat et d'une troisième suite  $\langle h_n \rangle$ , on obtient une suite dont le  $n$ ième terme vaut

$$\sum_{j+k+l=n} f_j g_k h_l.$$

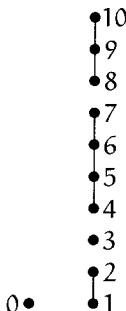
La fonction génératrice de cette double convolution est bien évidemment le double produit  $F(z)G(z)H(z)$ . De manière similaire, le  $n$ ième terme d'une convolution de  $m$  mêmes suites  $\langle g_n \rangle$  est égal à

$$\sum_{k_1+k_2+\dots+k_m=n} g_{k_1} g_{k_2} \dots g_{k_m}$$

et sa fonction génératrice est  $G(z)^m$ .

Ces observations peuvent s'appliquer au problème des arbres couvrants de l'éventail que nous avons récemment résolu (exemple 6, section 7.3). Rappelons qu'il s'agissait de calculer  $f_n$ , le nombre d'arbres couvrants d'un éventail d'ordre  $n$ . Nous allons le faire d'une autre manière, en regardant les configurations d'arêtes possibles entre les sommets  $\{1, 2, \dots, n\}$  : l'arête entre les sommet  $k$  et  $k+1$  peut appartenir ou non à l'arbre couvrant ; chacune des configurations obtenues forme un certain nombre d'ensembles connexes de sommets adjacents. Par exemple, si  $n = 10$ , on peut obtenir la

configuration suivante, formée des blocs  $\{1, 2\}$ ,  $\{3\}$ ,  $\{4, 5, 6, 7\}$  et  $\{8, 9, 10\}$  :



Combien d'arbres couvrants peut-on construire en ajoutant des arêtes vers le sommet 0 ? Il faut relier 0 à chacun des quatre blocs, et il y a deux façons possibles de le relier au bloc  $\{1, 2\}$ , une seule pour le bloc  $\{3\}$ , quatre pour le bloc  $\{4, 5, 6, 7\}$ , et trois pour le bloc  $\{8, 9, 10\}$ , ce qui donne  $2 \cdot 1 \cdot 4 \cdot 3 = 24$  configurations. En sommant sur tous les découpages possibles en blocs, on obtient l'expression suivante, qui compte le nombre total d'arbres couvrants :

$$f_n = \sum_{m > 0} \sum_{\substack{k_1 + k_2 + \dots + k_m = n \\ k_1, k_2, \dots, k_m > 0}} k_1 k_2 \dots k_m. \quad (7.65)$$

Par exemple,  $f_4 = 4 + 3 \cdot 1 + 2 \cdot 2 + 1 \cdot 3 + 2 \cdot 1 \cdot 1 + 1 \cdot 2 \cdot 1 + 1 \cdot 1 \cdot 2 + 1 \cdot 1 \cdot 1 \cdot 1 = 21$ .

C'est la somme des convolutions des  $m$  mêmes suites  $\langle 0, 1, 2, 3, \dots \rangle$ , pour  $m = 1, 2, 3, \dots$ . La fonction génératrice de  $\langle f_n \rangle$  est donc

$$F(z) = G(z) + G(z)^2 + G(z)^3 + \dots = \frac{G(z)}{1 - G(z)}$$

où  $G(z)$  est la fonction génératrice de  $\langle 0, 1, 2, 3, \dots \rangle$ , c'est-à-dire  $z/(1-z)^2$ .

Par conséquent, nous retrouvons

$$F(z) = \frac{z}{(1-z)^2 - z} = \frac{z}{1 - 3z + z^2}.$$

Cette approche est à la fois plus symétrique et plus attrayante que la précédente, qui faisait appel à une récurrence compliquée.

#### **Exemple 4 : une récurrence à base de convolution.**

L'exemple qui suit est extrêmement important. C'est en fait l'exemple qui illustre par excellence l'utilité des fonctions génératrices dans la résolution de récurrences.

Supposons que nous devions calculer le produit de  $n+1$  variables  $x_0, x_1, \dots, x_n$ . De combien de façons peut-on insérer des parenthèses dans ce produit de sorte que l'ordre des multiplications soit totalement déterminé ?

Par exemple, si  $n = 2$  il y a deux façons de faire le produit,  $x_0 \cdot (x_1 \cdot x_2)$  et  $(x_0 \cdot x_1) \cdot x_2$ . Si  $n = 3$ , il y en a cinq,

$$\begin{aligned} &x_0 \cdot (x_1 \cdot (x_2 \cdot x_3)), \quad x_0 \cdot ((x_1 \cdot x_2) \cdot x_3), \quad (x_0 \cdot x_1) \cdot (x_2 \cdot x_3), \\ &\quad (x_0 \cdot (x_1 \cdot x_2)) \cdot x_3, \quad ((x_0 \cdot x_1) \cdot x_2) \cdot x_3. \end{aligned}$$

Ainsi,  $C_2 = 2$  et  $C_3 = 5$  ; on a aussi  $C_1 = 1$  et  $C_0 = 1$ .

Appliquons la procédure en quatre étapes de la section 7.3, et cherchons donc une récurrence pour  $C_n$ . Voici la clé du problème : lorsque  $n > 0$ , il y a toujours exactement un signe “.” à l’extérieur de toutes les parenthèses ; c’est la multiplication finale, celle qui lie tout ensemble. Si ce “.” est situé entre  $x_k$  et  $x_{k+1}$ , il y a  $C_k$  façons de parentheser  $x_0 \cdots x_k$ , et  $C_{n-k-1}$  façons de parentheser  $x_{k+1} \cdots x_n$ . Par conséquent,

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-1} C_0, \quad \text{si } n > 0.$$

Nous reconnaissons immédiatement une convolution, et c’est maintenant routine que d’arranger la formule pour qu’elle soit valable pour tout entier  $n$  :

$$C_n = \sum_k C_k C_{n-1-k} + [n=0]. \quad (7.66)$$

La première étape est terminée. Suivant l’étape 2, multiplions par  $z^n$  et sommes :

$$\begin{aligned} C(z) &= \sum_n C_n z^n \\ &= \sum_{k,n} C_k C_{n-1-k} z^n + \sum_{n=0} z^n \\ &= \sum_k C_k z^k \sum_n C_{n-1-k} z^{n-k} + 1 \\ &= C(z) \cdot zC(z) + 1. \end{aligned}$$

Ça alors ! La convolution, en passant dans le monde des séries génératrices, est devenu un produit. La vie est décidément pleine de surprises.

La troisième étape est facile aussi. Résolvons cette équation du second degré en  $C(z)$ , et nous obtenons

$$C(z) = \frac{1 \pm \sqrt{1 - 4z}}{2z}.$$

Il y a deux solutions. Laquelle choisir, celle qui a le signe + ou celle qui a le signe - ? Les deux fonctions satisfont  $C(z) = zC(z)^2 + 1$ , mais une seule

peut répondre à notre problème. Si on choisit le signe +, on trouve que  $C(0) = \infty$ , alors qu'on doit avoir  $C(0) = C_0 = 1$ . Nous en concluons donc que

$$C(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

Nous voici à la quatrième étape. Que vaut  $[z^n] C(z)$ ? La formule du binôme nous indique que

$$\sqrt{1 - 4z} = \sum_{k \geq 0} \binom{1/2}{k} (-4z)^k = 1 + \sum_{k \geq 1} \frac{1}{2k} \binom{-1/2}{k-1} (-4z)^k;$$

donc, à l'aide de l'équation (5.37), nous trouvons que

$$\begin{aligned} \frac{1 - \sqrt{1 - 4z}}{2z} &= \sum_{k \geq 1} \frac{1}{k} \binom{-1/2}{k-1} (-4z)^{k-1} \\ &= \sum_{n \geq 0} \binom{-1/2}{n} \frac{(-4z)^n}{n+1} = \sum_{n \geq 0} \binom{2n}{n} \frac{z^n}{n+1}. \end{aligned}$$

Le nombre  $C_n$  de façons de parentheser est égal à  $\binom{2n}{n} \frac{1}{n+1}$ .

Nous avons anticipé sur ce résultat au chapitre 5, lorsque nous avons introduit la suite des *nombres de Catalan*  $\langle 1, 1, 2, 5, 14, \dots \rangle = \langle C_n \rangle$ . Cette suite apparaît dans des dizaines de problèmes qui semblent à première vue n'avoir aucun rapport entre eux [46]. En fait, on trouve dans beaucoup de situations une structure récursive qui correspond à la récurrence (7.66).

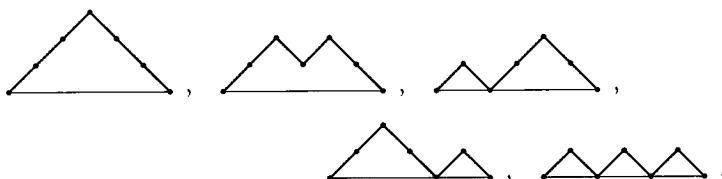
Considérons par exemple le problème suivant : combien existe-t-il de suites  $\langle a_1, a_2, \dots, a_{2n} \rangle$  de nombres +1 et -1 telles que

$$a_1 + a_2 + \cdots + a_{2n} = 0$$

et que toutes les sommes partielles

$$a_1, \quad a_1 + a_2, \quad \dots, \quad a_1 + a_2 + \cdots + a_{2n}$$

soient positives ou nulles ? Il y a évidemment  $n$  occurrences de +1 et  $n$  occurrences de -1. On peut représenter ce problème en considérant la suite des sommes partielles  $s_n = \sum_{k=1}^n a_k$  comme une fonction de  $n$  et en traçant son graphe. Voici les cinq solutions pour  $n = 3$  :



Ainsi, la récurrence de convolutions nous mène à une convolution récurrente.

Ce sont des “chaînes de montagne” de longueur  $2n$ , tracées uniquement avec des segments  $\diagup$  et  $\diagdown$ . Il se trouve qu'il y a en a exactement  $C_n$ . Voici comment mettre ces suites en relation avec le problème des parenthèses : ajoutez une paire de parenthèses qui contient la formule entière, de sorte qu'il y ait  $n$  paires de parenthèses correspondant aux  $n$  multiplications ; puis remplacez chaque “.” par  $+1$  et chaque “)” par  $-1$  et effacez tout le reste. Par exemple, la formule  $x_0 \cdot ((x_1 \cdot x_2) \cdot (x_3 \cdot x_4))$  correspond à la suite  $\langle +1, +1, -1, +1, +1, -1, -1, -1 \rangle$ . Les cinq façons de parentheser  $x_0 \cdot x_1 \cdot x_2 \cdot x_3$  correspondent aux cinq chaînes de montagnes ci-dessus.

De plus, il suffit de modifier très légèrement notre problème de dénombrement de suites pour en obtenir une solution combinatoire étonnamment simple, qui évite les fonction génératrices. Combien existe-t-il de suites  $\langle a_0, a_1, a_2, \dots, a_{2n} \rangle$  de  $+1$  et de  $-1$  telles que

$$a_0 + a_1 + a_2 + \dots + a_{2n} = 1,$$

et que toutes les sommes partielles

$$a_0, \quad a_0 + a_1, \quad a_0 + a_1 + a_2, \quad \dots, \quad a_0 + a_1 + \dots + a_{2n}$$

soient *strictement positives* ? Ce sont clairement les mêmes suites que celles du problème précédent, auquel on ajoute l’élément  $a_0 = +1$  tout au début. Cependant, ces nouvelles suites peuvent être dénombrées en utilisant un simple argument, basé sur un fait remarquable découvert par George Raney [302] en 1959 : *soit une suite d’entiers  $\langle x_1, x_2, \dots, x_m \rangle$  dont la somme est égale à  $+1$ . Parmi les permutations circulaires*

$$\langle x_1, x_2, \dots, x_m \rangle, \quad \langle x_2, \dots, x_m, x_1 \rangle, \quad \dots, \quad \langle x_m, x_1, \dots, x_{m-1} \rangle$$

*il en existe une seule dont toutes les sommes partielles sont strictement positives.* Par exemple, considérons la suite  $\langle 3, -5, 2, -2, 3, 0 \rangle$ . Ses permutations circulaires sont

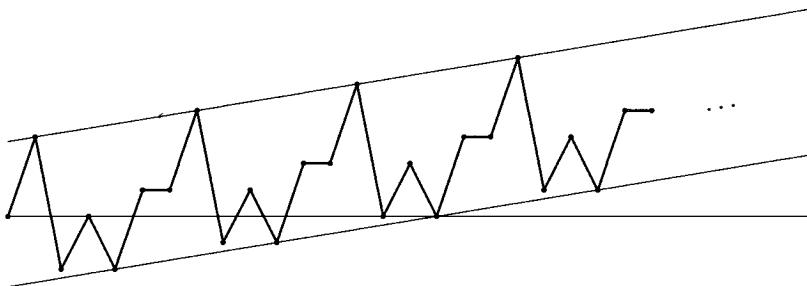
$\langle 3, -5, 2, -2, 3, 0 \rangle$	$\langle -2, 3, 0, 3, -5, 2 \rangle$
$\langle -5, 2, -2, 3, 0, 3 \rangle$	$\langle 3, 0, 3, -5, 2, -2 \rangle$ ✓
$\langle 2, -2, 3, 0, 3, -5 \rangle$	$\langle 0, 3, -5, 2, -2, 3 \rangle$

et un seul d’entre eux, celui qui est marqué, ne donne lieu qu’à des sommes partielles strictement positives.

Le lemme de Raney se démontre à l’aide d’un argument géométrique tout simple. Supposons qu’on répète indéfiniment la suite  $\langle x_1, x_2, \dots, x_m \rangle$ . On obtient ainsi la suite infinie périodique

$$\langle x_1, x_2, \dots, x_m, x_1, x_2, \dots, x_m, x_1, x_2, \dots \rangle,$$

telle que  $x_{m+k} = x_k$  pour tout  $k \geq 0$ . Si on dessine le graphe des sommes partielles  $s_n = x_1 + \dots + x_n$  en fonction de  $n$ , la courbe de  $s_n$  a une "pente moyenne" de  $1/m$ , du fait que  $s_{m+n} = s_n + 1$ . Par exemple, le graphe correspondant à notre suite  $(3, -5, 2, -2, 3, 0, 3, -5, 2, \dots)$  commence ainsi :



Comme on le voit, tout le graphe peut être contenu entre deux droites de pente  $1/m$  ( $m = 6$  dans notre cas). Chacune de ces droites touche la courbe exactement une fois par période de  $m$  points, car toute droite de pente  $1/m$  ne peut toucher un point à coordonnées entières qu'une fois toutes les  $m$  unités. Voici l'argument qui prouve le lemme de Raney : dans une période donnée, le seul endroit à partir duquel toutes les sommes partielles sont positives est le point d'intersection avec la droite du bas. En effet, tout autre point de la courbe est situé à moins de  $m$  unités à gauche d'une intersection de la courbe et de la droite.

*Ah, si mes actions pouvaient monter comme ça...*

*(A l'attention des informaticiens : dans ce problème, les sommes partielles représentent la taille de la pile en fonction du temps lorsqu'on calcule un produit de  $n$  facteurs ; chaque empilement augmente la taille de 1 et chaque multiplication la diminue de 1).*

Avec le lemme de Raney, nous pouvons facilement dénombrer les suites  $\langle a_0, \dots, a_{2n} \rangle$  de  $+1$  et  $-1$  dont les sommes partielles sont strictement positives et dont la somme totale vaut  $+1$ . Il existe  $\binom{2n+1}{n}$  suites contenant  $n$  occurrences de  $-1$  et  $n+1$  de  $+1$  et, d'après le lemme de Raney, il y a exactement une suite sur  $2n+1$  dont toutes les sommes partielles sont strictement positives. Pour voir cela, listez les  $N = \binom{2n+1}{n}$  suites et leurs  $2n+1$  décalages circulaires dans un tableau  $N \times (2n+1)$ . Chaque ligne contient exactement une solution, et chaque solution apparaît exactement une fois dans chaque colonne. Il y a donc  $N/(2n+1)$  solutions distinctes dans le tableau, chacune apparaissant  $(2n+1)$  fois. Voici par conséquent le nombre total de suites à sommes partielles strictement positives :

$$\binom{2n+1}{n} \frac{1}{2n+1} = \binom{2n}{n} \frac{1}{n+1} = c_n.$$

#### **Exemple 5 : une récurrence avec une convolution de $m$ suites**

On peut généraliser le problème précédent en considérant des suites  $\langle a_0, \dots, a_{mn} \rangle$  de  $+1$  et de  $(1-m)$  dont les sommes partielles sont toutes strictement positives et dont la somme totale vaut 1. Nous les appellerons des *suites de Raney d'ordre  $m$* . S'il y a  $k$  occurrences de  $(1-m)$  et

$mn + 1 - k$  occurrences de  $+1$ , alors

$$k(1-m) + (mn + 1 - k) = 1,$$

(A l'attention des informaticiens : ces suites peuvent aussi représenter des piles, mais avec des opérations d'arité  $m$  au lieu d'opérations binaires).

par conséquent  $k = n$ . Il existe  $\binom{mn+1}{n}$  suites contenant  $n$  occurrences de  $(1-m)$  et  $mn + 1 - n$  occurrences de  $+1$ . D'après le lemme de Raney, le nombre de telles suites ayant toutes leurs sommes partielles strictement positives est égal à

$$\binom{mn+1}{n} \frac{1}{mn+1} = \binom{mn}{n} \frac{1}{(m-1)n+1}. \quad (7.67)$$

C'est le nombre de suites de Raney d'ordre  $m$ . Nous le noterons  $C_n^{(m)}$  et l'appellerons "nombre de Fuss-Catalan", car c'est N.I. Fuss [135] qui a, en 1791, étudié le premier la suite  $\langle C_n^{(m)} \rangle$ , bien avant Catalan. Les nombres de Catalan ordinaires s'écrivent alors  $C_n = C_n^{(2)}$ .

Maintenant que nous connaissons la réponse, (7.67), jouons au "Jeopardy" et devinons la question qui lui correspond. Dans le cas  $m = 2$ , la question était la suivante : "Quelle suite de nombres  $C_n$  satisfait la récurrence  $C_n = \sum_k C_k C_{n-1-k} + [n=0]$ " ? Nous allons essayer de trouver une question similaire (c'est-à-dire une récurrence similaire) pour le cas général.

La suite triviale  $\langle +1 \rangle$ , de longueur 1, est bien sûr une suite de Raney d'ordre  $m$ . Si on ajoute le nombre  $(1-m)$  à droite d'une concaténation de  $m$  suites de Raney d'ordre  $m$  quelconques, on obtient une nouvelle suite de Raney d'ordre  $m$ . Réciproquement, on peut montrer que, si  $n > 0$ , toute suite de Raney d'ordre  $m$   $\langle a_0, \dots, a_{mn} \rangle$  peut se décomposer de cette manière. En effet, le dernier terme  $a_{mn}$  est forcément égal à  $(1-m)$ . De plus, les sommes partielles  $s_j = a_0 + \dots + a_{j-1}$  sont strictement positives pour tout  $1 \leq j \leq mn$ , et  $s_{mn} = m$  car  $s_{mn} + a_{mn} = 1$ . Soit  $k_1$  le plus grand indice  $\leq mn$  tel que  $s_{k_1} = 1$ ,  $k_2$  le plus grand tel que  $s_{k_2} = 2$  et ainsi de suite. Alors  $s_{k_j} = j$  et  $s_k > j$ , pour tous  $j$  et  $k$  tels que  $k_j < k \leq mn$  et  $1 \leq j \leq m$ . Il s'ensuit que  $k_m = mn$ , et on peut vérifier sans difficulté que chacune des sous-suites  $\langle a_0, \dots, a_{k_1-1} \rangle, \langle a_{k_1}, \dots, a_{k_2-1} \rangle, \dots, \langle a_{k_{m-1}}, \dots, a_{k_m-1} \rangle$  est une suite de Raney d'ordre  $m$ . Il existe donc des entiers positifs ou nuls  $n_1, n_2, \dots, n_m$  tels que  $k_1 = mn_1 + 1, k_2 - k_1 = mn_2 + 1, \dots, k_m - k_{m-1} = mn_m + 1$ .

Nous en déduisons que  $\binom{mn+1}{n} \frac{1}{mn+1}$  est la réponse aux deux intéressantes questions qui suivent : "Quelle suite de nombres  $C_n^{(m)}$  est définie par la récurrence

$$C_n^{(m)} = \left( \sum_{n_1+n_2+\dots+n_m=n-1} C_{n_1}^{(m)} C_{n_2}^{(m)} \dots C_{n_m}^{(m)} \right) + [n=0] \quad (7.68)$$

pour tout  $n$  entier" ? "Si  $G(z)$  est une série qui satisfait

$$G(z) = z G(z)^m + 1, \quad (7.69)$$

que vaut  $[z^n] G(z)$ " ?

Notez bien que ce ne sont pas des questions faciles. Pour résoudre (7.69) dans le cas des nombres de Catalan ordinaires, il nous a suffi de résoudre une équation du second degré et d'appliquer la formule du binôme. En revanche, pour  $m = 3$ , aucune technique standard ne peut nous aider à résoudre l'équation du troisième degré  $G = zG^3 + 1$ . C'est pourquoi il valait mieux répondre à la question avant de l'avoir posée.

Toutefois, nous en savons maintenant assez pour pouvoir nous poser des questions plus difficiles encore et y répondre. En voici une : "Que vaut  $[z^n] G(z)^l$  si  $l$  est un entier strictement positif et  $G(z)$  est la série définie par (7.69)" ? Avec les mêmes arguments que précédemment, nous pouvons prouver que  $[z^n] G(z)^l$  est le nombre de suites de longueur  $mn + l$  qui satisfont les trois propriétés suivantes :

- Chaque élément est soit  $+1$ , soit  $(1 - m)$ .
- Toutes les sommes partielles sont strictement positives.
- La somme totale vaut  $l$ .

En effet, chacune des suites ainsi définies s'obtient de façon unique en concatenant  $l$  suites de Raney d'ordre  $m$ . Le nombre de façons de faire cela est

$$\sum_{n_1+n_2+\dots+n_l=n} C_{n_1}^{(m)} C_{n_2}^{(m)} \dots C_{n_l}^{(m)} = [z^n] G(z)^l.$$

Raney a donné une généralisation de son lemme, qui permet de compter ces suites : *Si  $\langle x_1, x_2, \dots, x_m \rangle$  est une suite d'entiers tels que  $x_j \leq 1$  pour tout  $j$  et que  $x_1 + x_2 + \dots + x_m = l > 0$ , alors, parmi les permutations circulaires*

$$\langle x_1, x_2, \dots, x_m \rangle, \langle x_2, \dots, x_m, x_1 \rangle, \dots, \langle x_m, x_1, \dots, x_{m-1} \rangle,$$

*il en existe exactement  $l$  dont les sommes partielles sont toutes strictement positives.*

Vérifions cela sur la suite  $\langle -2, 1, -1, 0, 1, 1, -1, 1, 1, 1 \rangle$ . Voici les permutations circulaires correspondantes :

$\langle -2, 1, -1, 0, 1, 1, -1, 1, 1, 1 \rangle$	$\langle 1, -1, 1, 1, 1, -2, 1, -1, 0, 1 \rangle$
$\langle 1, -1, 0, 1, 1, -1, 1, 1, 1, -2 \rangle$	$\langle -1, 1, 1, 1, -2, 1, -1, 0, 1, 1 \rangle$
$\langle -1, 0, 1, 1, -1, 1, 1, 1, -2, 1 \rangle$	$\langle 1, 1, 1, -2, 1, -1, 0, 1, 1, -1 \rangle \checkmark$
$\langle 0, 1, 1, -1, 1, 1, 1, -2, 1, -1 \rangle$	$\langle 1, 1, -2, 1, -1, 0, 1, 1, -1, 1 \rangle$
$\langle 1, 1, -1, 1, 1, 1, -2, 1, -1, 0 \rangle \checkmark$	$\langle 1, -2, 1, -1, 0, 1, 1, -1, 1, 1 \rangle$

Seuls les deux décalages indiqués par “✓” ont toutes leurs sommes partielles strictement positives. On prouve ce lemme généralisé dans l'exercice 13.

Toute suite de +1 et de  $(1-m)$  de longueur  $mn+l$  et de somme totale l contient exactement n occurrences de  $(1-m)$ . Le lemme généralisé nous dit que l suites sur  $(mn+l)$  parmi ces  $\binom{mn+l}{n}$  suites ont toutes leurs sommes partielles strictement positives. La réponse à notre question difficile est donc étonnamment simple :

$$[z^n] G(z)^l = \binom{mn+l}{n} \frac{l}{mn+l}, \quad (7.70)$$

pour tout entier  $l > 0$ .

Ceux qui n'ont pas oublié le chapitre 5 ont probablement une impression de déjà-vu : n'avons-nous pas rencontré cette formule quelque part ? Eh bien si. Voici l'équation de Lambert (5.60) :

$$[z^n] B_t(z)^r = \binom{tn+r}{n} \frac{r}{tn+r}.$$

Par conséquent, la fonction génératrice  $G(z)$  de (7.69) et la série binomiale généralisée  $B_m(z)$  ne font qu'une. L'équation (5.59) indique que

$$B_m(z)^{1-m} - B_m(z)^{-m} = z,$$

ce qui revient à

$$B_m(z) - 1 = z B_m(z)^m.$$

Maintenant que nous travaillons en connaissance de cause avec les binomiaux généralisés, utilisons donc les notations du chapitre 5. Quelques identités de ce chapitre 5 avaient été énoncées sans preuve. Nous venons de combler en partie cette lacune en montrant que la série  $B_t(z)$  définie par

$$B_t(z) = \sum_n \binom{tn+1}{n} \frac{z^n}{tn+1}$$

possède la remarquable propriété suivante :

$$B_t(z)^r = \sum_n \binom{tn+r}{n} \frac{rz^n}{tn+r},$$

pour tous entiers strictement positifs t et r.

Pouvons-nous généraliser ce résultat à des valeurs *quelconques* de t et r ? Oui, car les coefficients  $\binom{tn+r}{n} \frac{r}{tn+r}$  sont des polynômes en t et r. Les coefficients de la puissance rième générale définie par

$$B_t(z)^r = e^{r \ln B_t(z)} = \sum_{n \geq 0} \frac{(r \ln B_t(z))^n}{n!} = \sum_{n \geq 0} \frac{r^n}{n!} \left( - \sum_{m \geq 1} \frac{(1-B_t(z))^m}{m} \right)^n$$

sont des polynômes en  $t$  et  $r$ , et ces polynômes sont égaux à  $\binom{tn+r}{n} \frac{r^r}{(tn+r)^{tn+r}}$  pour une infinité de valeurs de  $t$  et  $r$ . Par conséquent, les deux suites de polynômes sont égales.

On mentionne aussi dans le chapitre 5 la série exponentielle généralisée

$$\mathcal{E}_t(z) = \sum_{n \geq 0} \frac{(tn+1)^{n-1}}{n!} z^n,$$

qui, d'après (5.60), satisferait elle aussi une propriété remarquable :

$$[z^n] \mathcal{E}_t(z)^r = \frac{r(tn+r)^{n-1}}{n!}. \quad (7.71)$$

Nous pouvons maintenant prouver ce résultat, car il n'est pas difficile de montrer que

$$\mathcal{E}_t(z)^r = \lim_{x \rightarrow \infty} \mathcal{B}_{xt}(z/x)^{xr}.$$

## 7.6 FG EXPONENTIELLES

Il peut arriver qu'une suite  $\langle g_n \rangle$  ait une fonction génératrice plutôt compliquée et que la suite  $\langle g_n/n! \rangle$  en ait une tout à fait simple. Dans ce cas, il vaut mieux bien sûr travailler avec  $\langle g_n/n! \rangle$ , quitte à multiplier par  $n!$  pour finir. La série

$$\widehat{G}(z) = \sum_{n \geq 0} g_n \frac{z^n}{n!} \quad (7.72)$$

est appelée la *fonction génératrice exponentielle*, ou plus brièvement “fge”, de la suite  $\langle g_0, g_1, g_2, \dots \rangle$ . Cette appellation vient du fait que la fonction exponentielle  $e^z$  est la fge de  $\langle 1, 1, 1, \dots \rangle$ .

Beaucoup de fonctions de la table 373 sont en réalité des fge. Par exemple, d'après l'équation (7.51), l'expression  $(\ln \frac{1}{1-z})^m/m!$  est la fge de la suite  $\langle [0], [1], [2], \dots \rangle$ . La série génératrice ordinaire de cette suite est bien plus compliquée (de plus, elle diverge).

Il existe des manœuvres de base pour manipuler les fge, analogues à celles que nous avons vues en section 7.2 pour les fonctions génératrices ordinaires. Par exemple, si on multiplie la fge de  $\langle g_n \rangle$  par  $z$ , on obtient

$$\sum_{n \geq 0} g_n \frac{z^{n+1}}{n!} = \sum_{n \geq 1} g_{n-1} \frac{z^n}{(n-1)!} = \sum_{n \geq 0} n g_{n-1} \frac{z^n}{n!},$$

la fge de  $\langle 0, g_0, 2g_1, \dots \rangle = \langle ng_{n-1} \rangle$ .

En dérivant la fge de  $\langle g_0, g_1, g_2, \dots \rangle$  par rapport à  $z$ , on aboutit à

$$\sum_{n \geq 0} n g_n \frac{z^{n-1}}{n!} = \sum_{n \geq 1} g_n \frac{z^{n-1}}{(n-1)!} = \sum_{n \geq 0} g_{n+1} \frac{z^n}{n!}, \quad (7.73)$$

la fge de  $\langle g_1, g_2, \dots \rangle$ . Ainsi, la dérivation des fge correspond à l'opération de décalage vers la gauche  $(G(z) - g_0)/z$  des fonctions génératrices ordinaires (nous avons déjà appliqué cette propriété de décalage des fge en (5.106), lorsque nous étudions les séries hypergéométriques). Si on intègre une fge, on trouve

$$\int_0^z \sum_{n \geq 0} g_n \frac{t^n}{n!} dt = \sum_{n \geq 0} g_n \frac{z^{n+1}}{(n+1)!} = \sum_{n \geq 1} g_{n-1} \frac{z^n}{n!}, \quad (7.74)$$

ce qui donne, par un décalage vers la droite, la fge de  $\langle 0, g_0, g_1, \dots \rangle$ .

L'opération la plus intéressante sur les fge, comme sur les fonctions génératrices ordinaires, est la multiplication. Si on note  $\widehat{F}(z)$  et  $\widehat{G}(z)$  les fge respectives de  $\langle f_n \rangle$  et  $\langle g_n \rangle$ , alors  $\widehat{F}(z)\widehat{G}(z) = \widehat{H}(z)$  désigne la fge d'une suite  $\langle h_n \rangle$  que l'on appelle la *convolution binomiale* de  $\langle f_n \rangle$  et  $\langle g_n \rangle$  :

$$h_n = \sum_k \binom{n}{k} f_k g_{n-k}. \quad (7.75)$$

C'est parce que  $\binom{n}{k} = n!/k!(n-k)!$ , donc que

$$\frac{h_n}{n!} = \sum_{k=0}^n \frac{f_k}{k!} \frac{g_{n-k}}{(n-k)!}$$

qu'on voit des coefficients binomiaux dans cette expression. En d'autres termes, on peut dire que  $\langle h_n/n! \rangle$  est la convolution ordinaire de  $\langle f_n/n! \rangle$  et  $\langle g_n/n! \rangle$ .

Les convolutions binomiales ne sont pas rares. Par exemple, nous avons défini les nombres de Bernoulli, en (6.79), par la récurrence implicite

$$\sum_{j=0}^m \binom{m+1}{j} B_j = [m=0], \quad \text{pour tout } m \geq 0.$$

Elle peut aussi s'écrire comme une convolution binomiale, en posant  $n = m + 1$  et en ajoutant le terme  $B_n$  aux deux membres :

$$\sum_k \binom{n}{k} B_k = B_n + [n=1], \quad \text{pour tout } n \geq 0. \quad (7.76)$$

Nous pouvons maintenant exprimer cette récurrence en termes de séries génératrices, comme promis au chapitre 6. Soit  $\widehat{B}(z) = \sum_{n \geq 0} B_n z^n/n!$  la

fge des nombres de Bernoulli. Le membre gauche de (7.76) est exactement la convolution binomiale de  $\langle B_n \rangle$  et de la suite constante  $\langle 1, 1, 1, \dots \rangle$ ; la fge correspondante est donc  $\widehat{B}(z)e^z$ . Comme la fge du membre droit est  $\sum_{n \geq 0} (B_n + [n=1])z^n/n! = \widehat{B}(z) + z$ , on en déduit que  $\widehat{B}(z) = z/(e^z - 1)$ . Nous venons ainsi de démontrer l'équation (6.81), qui apparaît aussi dans la table 373 sous le numéro (7.45).

Maintenant, nous allons examiner encore une fois une somme que nous commençons à bien connaître :

$$S_m(n) = 0^m + 1^m + 2^m + \dots + (n-1)^m = \sum_{0 \leq k < n} k^m.$$

Nous allons tenter de la calculer en utilisant les fonctions génératrices, en considérant que  $n$  est fixé. Notre but est donc d'évaluer les coefficients de la série

$$S(z) = S_0(n) + S_1(n)z + S_2(n)z^2 + \dots = \sum_{m \geq 0} S_m(n)z^m.$$

Nous savons que la fonction génératrice de  $\langle 1, k, k^2, \dots \rangle$  est

$$\frac{1}{1-kz} = \sum_{m \geq 0} k^m z^m,$$

donc

$$S(z) = \sum_{m \geq 0} \sum_{0 \leq k < n} k^m z^m = \sum_{0 \leq k < n} \frac{1}{1-kz}$$

en changeant l'ordre de sommation. Cette somme admet une forme close :

$$\begin{aligned} S(z) &= \frac{1}{z} \left( \frac{1}{z^{-1}-0} + \frac{1}{z^{-1}-1} + \dots + \frac{1}{z^{-1}-n+1} \right) \\ &= \frac{1}{z} (H_{z^{-1}} - H_{z^{-1}-n}). \end{aligned} \tag{7.77}$$

Hélas, nous ne savons pas développer cette expression en série entière en  $z$ .

Appelons donc les fonctions génératrices exponentielles à notre secours. Voici la fge de notre suite  $\langle S_0(n), S_1(n), S_2(n), \dots \rangle$ :

$$\widehat{S}(z, n) = S_0(n) + S_1(n) \frac{z}{1!} + S_2(n) \frac{z^2}{2!} + \dots = \sum_{m \geq 0} S_m(n) \frac{z^m}{m!}.$$

Considérons la fge de la suite  $\langle 1, k, k^2, \dots \rangle$ , qui est

$$e^{kz} = \sum_{m \geq 0} k^m \frac{z^m}{m!}.$$

On a alors

$$\widehat{S}(z, n) = \sum_{m \geq 0} \sum_{0 \leq k < n} k^m \frac{z^m}{m!} = \sum_{0 \leq k < n} e^{kz}.$$

Cette dernière somme, qui est une progression géométrique, admet la forme close

$$\widehat{S}(z, n) = \frac{e^{nz} - 1}{e^z - 1}. \quad (7.78)$$

Eurêka ! Il ne nous reste plus qu'à trouver les coefficients de cette fonction assez simple. Nous en déduirons facilement  $S_m(n)$ , car  $S_m(n) = m! [z^m] \widehat{S}(z, n)$ .

C'est maintenant qu'entrent en scène les nombres de Bernoulli. Nous avons vu très récemment la fge de ces nombres :

$$\widehat{B}(z) = \sum_{k \geq 0} B_k \frac{z^k}{k!} = \frac{z}{e^z - 1}.$$

Nous pouvons donc écrire

$$\begin{aligned} \widehat{S}(z, n) &= \widehat{B}(z) \frac{e^{nz} - 1}{z} \\ &= \left( B_0 \frac{z^0}{0!} + B_1 \frac{z^1}{1!} + B_2 \frac{z^2}{2!} + \dots \right) \left( n \frac{z^0}{1!} + n^2 \frac{z^1}{2!} + n^3 \frac{z^2}{3!} + \dots \right). \end{aligned}$$

La somme  $S_m(n)$  est égale à  $m!$  fois le coefficient de  $z^m$  dans ce produit. Par exemple,

$$\begin{aligned} S_0(n) &= 0! \left( B_0 \frac{n}{1! 0!} \right) &= n; \\ S_1(n) &= 1! \left( B_0 \frac{n^2}{2! 0!} + B_1 \frac{n}{1! 1!} \right) &= \frac{1}{2}n^2 - \frac{1}{2}n; \\ S_2(n) &= 2! \left( B_0 \frac{n^3}{3! 0!} + B_1 \frac{n^2}{2! 1!} + B_2 \frac{n}{1! 2!} \right) &= \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n. \end{aligned}$$

Nous avons donc, pour la nième fois, calculé  $\square_n = S_2(n) = \frac{1}{3}n(n-\frac{1}{2})(n-1)$ . Ce calcul est plus simple que tous ceux que nous avions fait auparavant : il nous a suffi de quelques lignes pour déterminer le comportement de  $S_m(n)$  pour tout  $m$ .

La formule générale peut s'écrire

$$S_{m-1}(n) = \frac{1}{m} (B_m(n) - B_m(0)), \quad (7.79)$$

où  $B_m(x)$  est le *polynôme de Bernoulli*, défini par

$$B_m(x) = \sum_k \binom{m}{k} B_k x^{m-k}. \quad (7.80)$$

Voici pourquoi : le polynôme de Bernoulli est égal à la convolution binomiale des suites  $\langle B_0, B_1, B_2, \dots \rangle$  et  $\langle 1, x, x^2, \dots \rangle$ ; par conséquent, la fge de  $\langle B_0(x), B_1(x), B_2(x), \dots \rangle$  est égale au produit de leurs fonctions génératrices exponentielles,

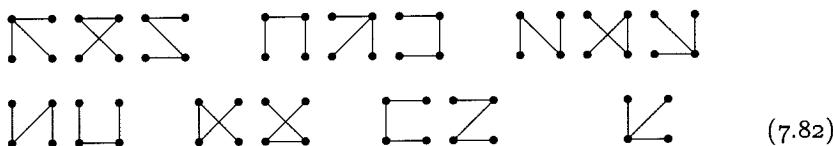
$$\widehat{B}(z, x) = \sum_{m \geq 0} B_m(x) \frac{z^m}{m!} = \frac{z}{e^z - 1} \sum_{m \geq 0} x^m \frac{z^m}{m!} = \frac{ze^{xz}}{e^z - 1}. \quad (7.81)$$

Cela entraîne l'équation (7.79) car, d'après (7.78), la fonction génératrice exponentielle de  $\langle 0, S_0(n), 2S_1(n), \dots \rangle$  est

$$z \frac{e^{nz} - 1}{e^z - 1} = \widehat{B}(z, n) - \widehat{B}(z, 0).$$

Voyons maintenant un autre problème, qui semble être fait pour les fonctions génératrices exponentielles : combien y a-t-il d'arbres couvrants dans le *graphe complet* à  $n$  sommets  $\{1, 2, \dots, n\}$ ? Convenons de noter  $t_n$  ce nombre. Le graphe complet contient  $\frac{1}{2}n(n-1)$  arêtes joignant chacune des paires de sommets distincts. Nous cherchons donc le nombre de façons de relier entre eux  $n$  objets donnés en traçant  $n-1$  segments qui les joignent deux à deux.

On voit très facilement que  $t_1 = t_2 = 1$ . Nous savons aussi que  $t_3 = 3$ , car un graphe complet à trois sommets est aussi un éventail d'ordre 2 et  $f_2 = 3$ . Si  $n = 4$ , on compte seize arbres couvrants :



Par conséquent,  $t_4 = 16$ .

D'après notre expérience du problème des éventails, il semblerait que le meilleur moyen d'opérer soit celui-ci : distinguer un sommet et observer les composantes connexes obtenues en supprimant toutes les arêtes qui touchent ce sommet. S'il y a  $m$  telles composantes connexes de tailles  $k_1, k_2, \dots, k_m$ , alors existe  $k_1 k_2 \dots k_m$  manières différentes de les connecter au sommet distingué. Prenons par exemple  $n = 4$  et considérons que le sommet en bas à gauche est distingué. On voit dans la première ligne de (7.82)  $3t_3$  cas où les trois autres sommets sont reliés entre eux de  $t_3$  manières différentes puis connectés au sommet distingué de 3 façons. La ligne du bas

contient  $2 \cdot 1 \times t_2 t_1 \times \binom{3}{2}$  solutions, dans lesquelles les trois autres sommets sont séparés en composantes de tailles 2 et 1 de  $\binom{3}{2}$  façons ; on y voit aussi le cas , où les trois sommets non distingués ne sont pas connectés les uns aux autres.

En raisonnant de cette façon, on trouve la récurrence

$$t_n = \sum_{m>0} \frac{1}{m!} \sum_{k_1+\dots+k_m=n-1} \binom{n-1}{k_1, k_2, \dots, k_m} k_1 k_2 \dots k_m t_{k_1} t_{k_2} \dots t_{k_m}$$

pour tout  $n > 1$ . En effet, il y a  $\binom{n-1}{k_1, k_2, \dots, k_m}$  façons de ranger  $n-1$  éléments dans une suite de  $m$  composantes de tailles respectives  $k_1, k_2, \dots, k_m$  ; il y a  $t_{k_1} t_{k_2} \dots t_{k_m}$  façons de faire un arbre couvrant pour chacune de ces composantes ; il y a  $k_1 k_2 \dots k_m$  façons de relier le sommet  $n$  à ces composantes ; et il faut diviser le tout par  $m!$ , car l'ordre des composantes ne doit pas être pris en compte. Par exemple, pour  $n = 4$ , la récurrence donne

$$\begin{aligned} t_4 &= 3t_3 + \frac{1}{2} \left( \binom{3}{1, 2} 2t_1 t_2 + \binom{3}{2, 1} 2t_2 t_1 \right) + \frac{1}{6} \left( \binom{3}{1, 1, 1} t_1^3 \right) \\ &= 3t_3 + 6t_2 t_1 + t_1^3. \end{aligned}$$

La récurrence des  $t_n$  peut sembler redoutable, voire épouvantable au premier abord. En fait elle n'est pas si méchante que cela. Si on définit

$$u_n = n t_n,$$

tout devient limpide :

$$\frac{u_n}{n!} = \sum_{m>0} \frac{1}{m!} \sum_{k_1+k_2+\dots+k_m=n-1} \frac{u_{k_1}}{k_1!} \frac{u_{k_2}}{k_2!} \dots \frac{u_{k_m}}{k_m!}, \quad \text{si } n > 1. \quad (7.83)$$

La somme interne est égale au coefficient de  $z^{n-1}$  dans la fge  $\widehat{U}(z)$  élevée à la puissance  $m$ ême. On peut obtenir une formule correcte même pour  $n = 1$  en ajoutant le terme  $\widehat{U}(z)^0$  qui correspond au cas  $m = 0$ . Ainsi,

$$\frac{u_n}{n!} = [z^{n-1}] \sum_{m \geq 0} \frac{1}{m!} \widehat{U}(z)^m = [z^{n-1}] e^{\widehat{U}(z)} = [z^n] z e^{\widehat{U}(z)}$$

pour tout  $n > 0$ , et on obtient l'équation

$$\widehat{U}(z) = z e^{\widehat{U}(z)}. \quad (7.84)$$

Nous avançons ! L'équation (7.84) est presque semblable à la formule

$$\mathcal{E}(z) = e^{z \mathcal{E}(z)}$$

qui définit la série exponentielle généralisée  $\mathcal{E}(z) = \mathcal{E}_1(z)$  en (5.59) et (7.71). En effet, nous avons

$$\widehat{\mathbf{U}}(z) = z \mathcal{E}(z).$$

Voici donc la réponse à notre problème :

$$t_n = \frac{u_n}{n} = \frac{n!}{n} [z^n] \widehat{\mathbf{U}}(z) = (n-1)! [z^{n-1}] \mathcal{E}(z) = n^{n-2}. \quad (7.85)$$

Pour tout  $n > 0$ , le graphe complet sur les sommets  $\{1, 2, \dots, n\}$  admet exactement  $n^{n-2}$  arbres couvrants distincts.

## 7.7 FG DE DIRICHLET

Il existe bien d'autres façons d'engendrer une suite de nombres à partir d'une série. On peut utiliser, en théorie, tout système de fonctions  $K_n(z)$  telles que

$$\sum_n g_n K_n(z) = 0 \implies g_n = 0 \text{ pour tout } n.$$

Dans le cas des fonctions génératrices ordinaires, on a  $K_n(z) = z^n$ , et, pour les fonctions génératrices exponentielles,  $K_n(z) = z^n/n!$ . Rien ne nous empêche d'essayer, par exemple, les puissances factorielles descendantes  $z^n$  ou les coefficients binomiaux  $z^n/n! = \binom{z}{n}$ .

Parmi les alternatives possibles, la plus intéressante est basée sur les fonctions  $1/n^z$ ; elle s'applique aux suites  $\langle g_1, g_2, \dots \rangle$ , qui commencent à  $n = 1$  au lieu de  $n = 0$ :

$$\tilde{G}(z) = \sum_{n \geq 1} \frac{g_n}{n^z}. \quad (7.86)$$

Les fonctions de ce type sont appelées des *fonctions génératrices de Dirichlet* (fgd), du nom du mathématicien allemand Gustav Lejeune Dirichlet (1805–1859) qui en a beaucoup fait usage.

Voici par exemple la fgd de la suite constante  $\langle 1, 1, 1, \dots \rangle$ :

$$\sum_{n \geq 1} \frac{1}{n^z} = \zeta(z). \quad (7.87)$$

C'est la *fonction zéta* de Riemann, qui se trouve être le nombre harmonique généralisé  $H_\infty^{(z)}$  lorsque  $z > 1$ .

Le produit de deux fonctions génératrices de Dirichlet correspond à une forme particulière de convolution :

$$\tilde{F}(z) \tilde{G}(z) = \sum_{l, m \geq 1} \frac{f_l}{l^z} \frac{g_m}{m^z} = \sum_{n \geq 1} \frac{1}{n^z} \sum_{l, m \geq 1} f_l g_m [l \cdot m = n].$$

Ainsi,  $\tilde{F}(z)\tilde{G}(z) = \tilde{H}(z)$  est la fgd de la suite

$$h_n = \sum_{d|n} f_d g_{n/d}. \quad (7.88)$$

Nous savons par exemple, d'après (4.55), que  $\sum_{d|n} \mu(d) = [n=1]$  ; c'est la convolution de Dirichlet de la suite de Möbius  $\langle \mu(1), \mu(2), \mu(3), \dots \rangle$  et de  $\langle 1, 1, 1, \dots \rangle$ . Par conséquent,

$$\tilde{M}(z)\zeta(z) = \sum_{n \geq 1} \frac{[n=1]}{n^z} = 1. \quad (7.89)$$

En d'autres termes, la fgd de  $\langle \mu(1), \mu(2), \mu(3), \dots \rangle$  est  $\zeta(z)^{-1}$ .

Les fonctions génératrices de Dirichlet s'avèrent particulièrement utiles lorsque la suite  $\langle g_1, g_2, \dots \rangle$  est une *fonction multiplicative*, c'est-à-dire lorsque

$$g_{mn} = g_m g_n \quad \text{pour } m \perp n.$$

Dans ce cas, la valeur de  $g_n$  pour tout  $n$  est déterminée par les valeurs de  $g_p$  pour les  $n$  qui sont puissances d'un nombre premier. On peut ainsi factoriser la fgd en un produit sur l'ensemble des nombres premiers :

$$\tilde{G}(z) = \prod_{p \text{ premier}} \left( 1 + \frac{g_p}{p^z} + \frac{g_{p^2}}{p^{2z}} + \frac{g_{p^3}}{p^{3z}} + \dots \right). \quad (7.90)$$

Si on pose, par exemple,  $g_n = 1$  pour tout  $n$ , on obtient une expression de la fonction zéta de Riemann par un produit :

$$\zeta(z) = \prod_{p \text{ premier}} \left( \frac{1}{1 - p^{-z}} \right). \quad (7.91)$$

La fonction de Möbius satisfait  $\mu(p) = -1$  et  $\mu(p^k) = 0$  pour  $k > 1$ . Par conséquent, sa fgd est

$$\tilde{M}(z) = \prod_{p \text{ premier}} (1 - p^{-z}), \quad (7.92)$$

ce qui concorde bien sûr avec (7.89) et (7.91). La fonction  $\varphi$  d'Euler satisfait  $\varphi(p^k) = p^k - p^{k-1}$ , donc sa fgd s'exprime sous la forme

$$\tilde{\Phi}(z) = \prod_{p \text{ premier}} \left( 1 + \frac{p-1}{p^z - p} \right) = \prod_{p \text{ premier}} \frac{1 - p^{-z}}{1 - p^{1-z}}. \quad (7.93)$$

Nous en concluons que  $\tilde{\Phi}(z) = \zeta(z-1)/\zeta(z)$ .

## Exercices

### Echauffements

- 1 Un collectionneur excentrique achète des pavages  $2 \times n$  par des dominos au prix de 4 francs le domino vertical et 1 franc le domino horizontal. Combien de pavages valent exactement  $m$  francs ? Par exemple, pour  $m = 6$ , il y a trois solutions :  $\square\square$ ,  $\square\square\square$  et  $\square\square\square\square$ .
- 2 Donnez des formes closes des fonctions génératrices ordinaire et exponentielle de la suite  $\langle 2, 5, 13, 35, \dots \rangle = \langle 2^n + 3^n \rangle$ .
- 3 Que vaut  $\sum_{n \geq 0} H_n / 10^n$  ?
- 4 Le théorème de développement des fonctions génératrices rationnelles  $P(z)/Q(z)$  n'est pas tout à fait général, car le degré de  $P$  doit être inférieur à celui de  $Q$ . Que se passe-t-il si le degré de  $P$  est supérieur ou égal à celui de  $Q$  ?
- 5 Trouvez une fonction génératrice  $S(z)$  telle que

$$[z^n] S(z) = \sum_k \binom{r}{k} \binom{r}{n-2k}.$$

### Exercices de base

- 6 Montrez qu'on peut résoudre la récurrence (7.32) avec la méthode du répertoire, sans faire appel aux fonctions génératrices.
- 7 Résolvez la récurrence

$$\begin{aligned} g_0 &= 1; \\ g_n &= g_{n-1} + 2g_{n-2} + \cdots + ng_0, \quad \text{pour } n > 0. \end{aligned}$$

- 8 Que vaut  $[z^n] (\ln(1-z))^2 / (1-z)^{m+1}$  ?
- 9 Utilisez le résultat de l'exercice précédent pour évaluer  $\sum_{k=0}^n H_k H_{n-k}$ .
- 10 Posez  $r = s = -1/2$  dans l'identité (7.62), puis supprimez toutes les occurrences de  $1/2$  en opérant comme dans (5.36). Quelle identité stupéfiante déduisez-vous ?
- 11 Ce problème, dont les trois parties sont indépendantes, permet de s'entraîner à la manipulation des fonctions génératrices. Nous supposons que  $A(z) = \sum_n a_n z^n$ ,  $B(z) = \sum_n b_n z^n$ ,  $C(z) = \sum_n c_n z^n$  et que les coefficients sont nuls pour tout  $n < 0$ .
  - a Si  $c_n = \sum_{j+2k \leq n} a_j b_k$ , exprimez  $C$  en fonction de  $A$  et  $B$ .
  - b Si  $nb_n = \sum_{k=0}^n 2^k a_k / (n-k)!$ , exprimez  $A$  en fonction de  $B$ .
  - c Si  $r$  est un nombre réel et si  $a_n = \sum_{k=0}^n \binom{r+k}{k} b_{n-k}$ , exprimez  $A$  en fonction de  $B$ , puis appliquez votre formule pour trouver des coefficients  $f_k(r)$  tels que  $b_n = \sum_{k=0}^n f_k(r) a_{n-k}$ .

J'en déduis  
l'identité de  
Superman :  
Clark Kent.

- 12 Combien y a-t-il de façons d'écrire les nombres  $\{1, 2, \dots, 2n\}$  dans un tableau  $2 \times n$  de sorte que les lignes et les colonnes croissent de gauche à droite et de bas en haut ? Voici par exemple une solution pour  $n = 5$  :

$$\begin{pmatrix} 3 & 6 & 7 & 9 & 10 \\ 1 & 2 & 4 & 5 & 8 \end{pmatrix}.$$

- 13 Démontrez le lemme généralisé de Raney qui est énoncé juste avant (7.70).

- 14 Résolvez la récurrence

$$g_0 = 0, \quad g_1 = 1,$$

$$g_n = -2ng_{n-1} + \sum_k \binom{n}{k} g_k g_{n-k}, \quad \text{pour } n > 1,$$

en utilisant une fonction génératrice exponentielle.

- 15 Le *nombre de Bell*  $\omega_n$  est le nombre de partitions distinctes d'un ensemble à  $n$  éléments. Par exemple,  $\omega_3 = 5$  car l'ensemble  $\{1, 2, 3\}$  peut être partitionné de cinq façons différentes :

$$\{1, 2, 3\}; \quad \{1, 2\} \cup \{3\}; \quad \{1, 3\} \cup \{2\}; \quad \{1\} \cup \{2, 3\}; \quad \{1\} \cup \{2\} \cup \{3\}.$$

Montrez que  $\omega_{n+1} = \sum_k \binom{n}{k} \omega_{n-k}$ , puis utilisez cette récurrence pour trouver une forme close de la fonction génératrice exponentielle  $P(z) = \sum_n \omega_n z^n / n!$ .

- 16 Soient deux suites  $\langle a_n \rangle$  et  $\langle b_n \rangle$  liées par la convolution

$$b_n = \sum_{k_1+k_2+\dots+n_k=n} \binom{a_1+k_1-1}{k_1} \binom{a_2+k_2-1}{k_2} \dots \binom{a_n+k_n-1}{k_n}$$

et telles que  $a_0 = 0$  et  $b_0 = 1$ . Prouvez que les fonctions génératrices correspondantes satisfont  $\ln B(z) = A(z) + \frac{1}{2}A(z^2) + \frac{1}{3}A(z^3) + \dots$ .

- 17 Montrez que la fonction génératrice exponentielle  $\widehat{G}(z)$  et la fonction génératrice ordinaire  $G(z)$  de toute suite sont liées par la formule

$$\int_0^\infty \widehat{G}(zt) e^{-t} dt = G(z)$$

si l'intégrale existe.

- 18 Trouvez les fonctions génératrices de Dirichlet des suites

- a  $g_n = \sqrt{n}$ ;
- b  $g_n = \ln n$ ;
- c  $g_n = [n \text{ est sans carré}]$ .

Exprimez vos réponses en utilisant la fonction zéta (la notion de nombre sans carré est définie dans l'exercice 4.13).

- 19 A partir de toute série  $F(z) = \sum_{n \geq 0} f_n z^n$  telle que  $f_0 = 1$ , on peut définir une suite de polynômes  $f_n(x)$  avec la règle  $F(z)^x = \sum_{n \geq 0} f_n(x)z^n$ , où  $f_n(1) = f_n$  et  $f_n(0) = [n=0]$ . En général,  $f_n(x)$  est de degré  $n$ . Montrez que ces polynômes satisfont toujours les formules

$$\sum_{k=0}^n f_k(x) f_{n-k}(y) = f_n(x+y);$$

$$(x+y) \sum_{k=0}^n k f_k(x) f_{n-k}(y) = x n f_n(x+y).$$

Qu'entendez-vous par "en général" ?  
Si  $f_1 = f_2 = \dots = f_{m-1} = 0$ , le degré de  $f_n(x)$  est au plus égal à  $\lfloor n/m \rfloor$ .

(Les identités des tables 216 et 289 sont des cas particuliers de ces formules).

- 20 Une série entière  $G(z)$  est dite *différentiablement finie*, ou plus simplement d-finie, s'il existe un ensemble fini de polynômes  $P_0(z), \dots, P_m(z)$ , non tous nuls, tels que

$$P_0(z)G(z) + P_1(z)G'(z) + \dots + P_m(z)G^{(m)}(z) = 0.$$

Une suite de nombres  $\langle g_0, g_1, g_2, \dots \rangle$  est dite *polynomialement récursive*, ou p-récessive, s'il existe un ensemble fini de polynômes  $p_0(z), \dots, p_m(z)$ , non tous nuls, tels que

$$p_0(n)g_n + p_1(n)g_{n+1} + \dots + p_m(n)g_{n+m} = 0$$

pour tout entier  $n \geq 0$ . Pouvez qu'une fonction génératrice est d-finie si et seulement si la suite de ses coefficients est p-récessive.

### Devoirs à la maison

- 21 Au cours d'un hold-up, un bandit réclame 500 francs en pièces de dix et vingt centimes. Il exige aussi de connaître le nombre de façons possibles de lui donner cet argent. Trouvez une fonction génératrice  $G(z)$  telle que le nombre cherché soit  $[z^{500}] G(z)$ , puis un autre fonction génératrice  $\check{G}(z)$ , plus compacte, telle que le nombre cherché soit  $[z^{500}] \check{G}(z)$ . Trouvez le nombre demandé (a) en utilisant des éléments simples ; (b) en appliquant une méthode du genre de (7.39).
- 22 Soit  $P$  la somme de toutes les façons possibles de "trianguler" des polygones :

$$P = \_ + \triangle + \square + \diagup \square \diagdown +$$

$$+ \begin{array}{c} \diagup \\ \square \\ \diagdown \end{array} + \cdots .$$

Accepterait-il des pavages  $2 \times n$  ?

(Le premier terme représente un polygone dégénéré qui n'a que deux côtés ; chacun des autres termes est un polygone divisé en triangles. Par exemple, un pentagone peut être triangulé de cinq façons différentes). Définissez une opération de "multiplication"  $A \Delta B$  sur les polygones triangulés  $A$  et  $B$ , de sorte que l'équation

$$P = \underline{\quad} + P \Delta P$$

soit vérifiée. Remplacez ensuite chaque triangle par la variable  $z$ . Qu'en déduisez-vous en ce qui concerne le nombre de façons de décomposer un polygone à  $n$  côtés en triangles ?

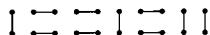
*Au tarif syndical,  
à peu près autant  
que vous pouvez en  
payer.*

- 23 De combien de façons peut-on construire un pilier  $2 \times 2 \times n$  avec des briques  $2 \times 1 \times 1$  ?
- 24 Combien y a-t-il d'arbres couvrants dans une roue à  $n$  rayons (un graphe ayant  $n$  sommets "externes" qui forment un cycle, chacun d'eux étant relié à un  $(n+1)$ ième sommet "central"), pour  $n \geq 3$  ?
- 25 Soit  $m \geq 2$  un entier. Trouvez une forme close pour la fonction génératrice de la suite  $\langle n \bmod m \rangle$  en fonction de  $z$  et  $m$ . Utilisez cette fonction génératrice pour exprimer " $n \bmod m$ " en fonction du nombre complexe  $\omega = e^{2\pi i/m}$ . (Par exemple, si  $m = 2$ ,  $\omega = -1$  et  $n \bmod 2 = \frac{1}{2} - \frac{1}{2}(-1)^n$ ).
- 26 Les nombres de Fibonacci du second ordre  $\langle \mathfrak{F}_n \rangle$  sont définis par la récurrence

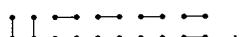
$$\begin{aligned} \mathfrak{F}_0 &= 0; & \mathfrak{F}_1 &= 1; \\ \mathfrak{F}_n &= \mathfrak{F}_{n-1} + \mathfrak{F}_{n-2} + F_n, & \text{pour } n > 1. \end{aligned}$$

Exprimez  $\mathfrak{F}_n$  en fonction des nombres de Fibonacci ordinaires  $F_n$  et  $F_{n+1}$ .

- 27 Un pavage d'un rectangle  $2 \times n$  par des dominos peut aussi être considéré comme une façon de tracer  $n$  segments disjoints dans une configuration de points  $2 \times n$  :



Si on superpose deux motifs de ce type, on obtient un ensemble de cycles, du fait que tout point se trouve à l'extrémité de deux segments exactement. Par exemple, si on superpose la figure précédente et celle qui suit,



on obtient le résultat

$$\boxed{1} \boxed{\bullet\bullet\bullet} \boxed{\square} .$$

On trouve ce même ensemble de cycles si on superpose

$$\boxed{1} \boxed{\square\square\square\square} \boxed{1} \quad \text{et} \quad \boxed{1} \boxed{\square\square\square} \boxed{1} \boxed{\square\square\square} .$$

On peut toutefois faire en sorte qu'il n'existe qu'une seule façon de retrouver les figures d'origine à partir de leur superposition. Pour cela, on oriente les segments verticaux alternativement vers le haut et vers le bas (haut/bas/haut/bas...) dans le premier motif, et alternativement vers le bas et vers le haut (bas/haut/bas/haut...) dans le second. Par exemple,

$$\boxed{\square\square\square} \boxed{\square\square\square} \boxed{\square\square\square} + \boxed{\square\square\square\square\square\square} = \boxed{1} \boxed{\bullet\bullet\bullet} \boxed{\square} .$$

On doit donc pouvoir prouver, algébriquement, que le nombre  $Q_n$  de ces motifs à cycles orientés de dimensions  $2 \times n$  est égal à  $T_n^2 = F_{n+1}^2$ . Trouvez une récurrence pour  $Q_n$ , résolvez-la à l'aide des fonctions génératrices et déduisez-en que  $Q_n = F_{n+1}^2$ .

- 28 Les coefficients de  $A(z)$  dans (7.39) satisfont  $A_r + A_{r+10} + A_{r+20} + A_{r+30} = 100$  pour tout  $0 \leq r < 10$ . Trouvez une explication "simple" de ce fait.
- 29 Que vaut la somme des produits de Fibonacci

$$\sum_{m>0} \sum_{\substack{k_1+k_2+\dots+k_m=n \\ k_1, k_2, \dots, k_m > 0}} F_{k_1} F_{k_2} \dots F_{k_m} ?$$

- 30 Si la décomposition en éléments simples de la fonction génératrice  $G(z) = 1/(1 - \alpha z)(1 - \beta z)$  est  $a/(1 - \alpha z) + b/(1 - \beta z)$ , quelle est la décomposition en éléments simples de  $G(z)^n$  ?
- 31 Trouvez la fonction  $g(n)$  sur les entiers strictement positifs  $n$  qui satisfait la récurrence

$$\sum_{d|n} g(d) \varphi(n/d) = 1 ,$$

où  $\varphi$  est la fonction d'Euler.

- 32 Une *progression arithmétique* est un ensemble infini d'entiers

$$\{an + b\} = \{b, a+b, 2a+b, 3a+b, \dots\}.$$

Un ensemble de progressions arithmétiques  $\{a_1 n + b_1\}, \dots, \{a_m n + b_m\}$  forme une partition de l'ensemble des entiers naturels si tout entier

positif ou nul apparaît dans une et une seule progression de l'ensemble. Par exemple, les trois progressions  $\{2n\}$ ,  $\{4n+1\}$  et  $\{4n+3\}$  constituent une partition de l'ensemble des entiers naturels. Montrez que si  $\{a_1n + b_1\}, \dots, \{a_mn + b_m\}$  est une partition de l'ensemble des entiers naturels telle que  $2 \leq a_1 \leq \dots \leq a_m$ , alors  $a_{m-1} = a_m$ . *Suggestion* : utilisez les fonctions génératrices.

### Problèmes d'examen

- 33** Que vaut  $[w^m z^n] (\ln(1+z))/(1-wz)$  ?
- 34** Trouvez une forme close pour la fonction génératrice  $\sum_{n \geq 0} G_n(z) w^n$ , si

$$G_n(z) = \sum_{k \leq n/m} \binom{n - mk}{k} z^{mk},$$

où  $m$  est un entier strictement positif donné.

- 35** Calculez la somme  $\sum_{0 < k < n} 1/k(n-k)$  de deux façons différentes :
- a** Développez le terme général en éléments simples.
  - b** Considérez la somme comme une convolution et utilisez les fonctions génératrices.
- 36** Soit  $A(z)$  la fonction génératrice de  $\langle a_0, a_1, a_2, a_3, \dots \rangle$ . Exprimez  $\sum_n a_{\lfloor n/m \rfloor} z^n$  en fonction de  $A$ ,  $z$  et  $m$ .
- 37** Soit  $a_n$  le nombre de façons d'écrire l'entier strictement positif  $n$  comme une somme de puissances de 2, sans se préoccuper de l'ordre. Par exemple,  $a_4 = 4$  car  $4 = 2 + 2 = 2 + 1 + 1 = 1 + 1 + 1 + 1$ . Par convention, on pose  $a_0 = 1$ . Soit  $b_n = \sum_{k=0}^n a_k$  la somme des premiers  $a_k$ .
  - a** Faites une table des  $a_n$  et des  $b_n$  jusqu'à  $n = 10$ . Quelle étonnante relation remarquez-vous ? Ne la prouvez pas encore.
  - b** Ecrivez la fonction génératrice  $A(z)$  comme un produit infini.
  - c** Utilisez l'expression de la partie (b) pour prouver le résultat de la partie (a).
- 38** Trouvez une forme close pour la fonction génératrice double

$$M(w, z) = \sum_{m, n \geq 0} \min(m, n) w^m z^n.$$

Généralisez votre réponse pour obtenir, pour  $m \geq 2$  fixé, une forme close de

$$M(z_1, \dots, z_m) = \sum_{n_1, \dots, n_m \geq 0} \min(n_1, \dots, n_m) z_1^{n_1} \dots z_m^{n_m}.$$

## 400 FONCTIONS GÉNÉRATRICES

- 39 Etant donnés deux entiers strictement positifs  $m$  et  $n$ , trouvez des formes closes pour

$$\sum_{1 \leq k_1 < k_2 < \dots < k_m \leq n} k_1 k_2 \dots k_m$$

et

$$\sum_{1 \leq k_1 \leq k_2 \leq \dots \leq k_m \leq n} k_1 k_2 \dots k_m.$$

(Par exemple, lorsque  $m = 2$  et  $n = 3$  les sommes valent  $1 \cdot 2 + 1 \cdot 3 + 2 \cdot 3$  et  $1 \cdot 1 + 1 \cdot 2 + 1 \cdot 3 + 2 \cdot 2 + 2 \cdot 3 + 3 \cdot 3$ ). *Suggestion* : regardez ce que valent les coefficients de  $z^m$  dans les fonctions génératrices  $(1 + a_1 z) \dots (1 + a_n z)$  et  $1/(1 - a_1 z) \dots (1 - a_n z)$ .

- 40 Trouvez une forme close pour

$$\sum_k \binom{n}{k} (kF_{k-1} - F_k)(n-k).$$

- 41 Une permutation alternante d'ordre  $n$  est une permutation  $a_1 a_2 \dots a_n$  des entiers  $\{1, 2, \dots, n\}$  qui monte et descend alternativement :

$$a_1 < a_2 > a_3 < a_4 > \dots$$

Par exemple, 35142 est une permutation alternante d'ordre 5. Si  $A_n$  désigne le nombre de permutations alternantes d'ordre  $n$ , montrez que la fonction génératrice exponentielle de  $\langle A_n \rangle$  est  $(1 + \sin z)/\cos z$ .

- 42 Une sonde spatiale a découvert qu'il existe sur Mars des matériaux organiques dont l'ADN est composé de cinq symboles, désignés par les lettres  $(a, b, c, d, e)$ , au lieu des quatre composants de l'ADN terrestre. Une chaîne d'ADN martien ne contient jamais l'une des paires de lettres consécutives  $cd$ ,  $ce$ ,  $ed$  et  $ee$ , mais toute autre chaîne est possible (par exemple,  $b b c d a$  est interdit mais  $b b d c a$  est correct). Combien peut-il exister de chaînes d'ADN martien de longueur  $n$  (pour  $n = 2$ , la réponse est 21) ?

- 43 On définit la série génératrice newtonienne d'une suite  $\langle g_n \rangle$  par

$$\dot{G}(z) = \sum_n g_n \binom{z}{n}.$$

Touvez une formule à base de convolution qui définit la relation entre des suites  $\langle f_n \rangle$ ,  $\langle g_n \rangle$  et  $\langle h_n \rangle$  dont les fonctions génératrices newtonniennes satisfont l'équation  $F(z)\dot{G}(z) = \dot{H}(z)$ . Essayez de trouver une formule aussi simple et symétrique que possible.

- 44** Soit  $q_n$  le nombre de résultats possibles lorsqu'on compare  $n$  nombres  $\{x_1, \dots, x_n\}$  les uns avec les autres. Par exemple,  $q_3 = 13$  car les possibilités sont

$$\begin{aligned} &x_1 < x_2 < x_3 ; \quad x_1 < x_2 = x_3 ; \quad x_1 < x_3 < x_2 ; \quad x_1 = x_2 < x_3 ; \\ &x_1 = x_2 = x_3 ; \quad x_1 = x_3 < x_2 ; \quad x_2 < x_1 < x_3 ; \\ &x_2 < x_1 = x_3 ; \quad x_2 < x_3 < x_1 ; \quad x_2 = x_3 < x_1 ; \\ &x_3 < x_1 < x_2 ; \quad x_3 < x_1 = x_2 ; \quad x_3 < x_2 < x_1 . \end{aligned}$$

Trouvez une forme close pour la fge  $\widehat{Q}(z) = \sum_n q_n z^n/n!$ . Trouvez aussi des suites  $\langle a_n \rangle$ ,  $\langle b_n \rangle$  et  $\langle c_n \rangle$  telles que

$$q_n = \sum_{k \geq 0} k^n a_k = \sum_k \binom{n}{k} b_k = \sum_k \binom{n}{k} c_k, \text{ pour tout } n > 0.$$

- 45** Calculez  $\sum_{m,n>0} [m \perp n]/m^2 n^2$ .

- 46** Trouvez une forme close de

$$\sum_{0 \leq k \leq n/2} \binom{n-2k}{k} \left(\frac{-4}{27}\right)^k.$$

*Suggestion :*  $z^3 - z^2 + \frac{4}{27} = (z + \frac{1}{3})(z - \frac{2}{3})^2$ .

- 47** Montrez que les nombres  $U_n$  et  $V_n$  de pavages de rectangles  $3 \times n$  par des dominos, donnés dans (7.34), sont étroitement liés aux fractions de l'arbre de Stern–Brocot qui convergent vers  $\sqrt{3}$ .

- 48** Une certaine suite  $\langle g_n \rangle$  satisfait la récurrence

$$ag_n + bg_{n+1} + cg_{n+2} + d = 0, \quad n \geq 0 \text{ entier},$$

pour un quadruplet d'entiers  $(a, b, c, d)$  tels que  $\text{pgcd}(a, b, c, d) = 1$ . Elle admet aussi la forme close

$$g_n = \lfloor \alpha(1 + \sqrt{2})^n \rfloor, \quad n \geq 0 \text{ entier},$$

pour un certain nombre  $\alpha$  compris entre 0 et 1. Trouvez  $a$ ,  $b$ ,  $c$ ,  $d$  et  $\alpha$ .

- 49** Voici un problème concernant les puissances et la parité.

- a** Soit la suite  $\langle a_0, a_1, a_2, \dots \rangle = \langle 2, 2, 6, \dots \rangle$  définie par la formule

$$a_n = (1 + \sqrt{2})^n + (1 - \sqrt{2})^n.$$

Trouvez une relation de récurrence satisfaite par cette suite.

- b** Prouvez que  $\lceil (1 + \sqrt{2})^n \rceil \equiv n \pmod{2}$  pour tout entier  $n > 0$ .

- c Trouvez un nombre  $\alpha$  de la forme  $(p + \sqrt{q})/2$ , où  $p$  et  $q$  sont des entiers strictement positifs, tel que  $\lfloor \alpha^n \rfloor \equiv n \pmod{2}$  pour tout entier  $n > 0$ .

### Questions subsidiaires

- 50 Continuons l'exercice 22 en considérant la somme de toutes les façons possibles de décomposer des polygones en polygones :

$$\begin{aligned} Q = & \_ + \triangle + \square + \boxtimes + \boxdot \\ & + \text{pentagone}_1 + \text{pentagone}_2 + \text{pentagone}_3 + \text{pentagone}_4 + \text{pentagone}_5 + \text{pentagone}_6 + \text{pentagone}_7 + \dots \end{aligned}$$

Trouvez une équation symbolique satisfaite par  $Q$  et appliquez-la pour trouver la fonction génératrice du nombre de façons de tracer des diagonales qui ne se coupent pas à l'intérieur d'un polygone à  $n$  côtés. On demande une forme close pour la fonction génératrice en  $z$ , mais pas pour ses coefficients.

- 51 Montrez que le produit

$$2^{mn/2} \prod_{\substack{1 \leq j \leq m \\ 1 \leq k \leq n}} \left( \left( \cos^2 \frac{j\pi}{m+1} \right) \square^2 + \left( \cos^2 \frac{k\pi}{n+1} \right) \square^2 \right)^{1/4}$$

est la fonction génératrice des pavages d'un rectangle  $m \times n$  par des dominos. (Il y a  $mn$  facteurs, que l'on peut imaginer être écrits dans les  $mn$  cellules du rectangle. Si  $mn$  est impair, le facteur du milieu est nul. Le coefficient de  $\square^j \square^k$  est le nombre de façons d'effectuer le pavage avec  $j$  dominos verticaux et  $k$  dominos horizontaux). *Suggestion* : c'est un problème difficile, qui est en réalité au-delà des objectifs de ce livre. Vous pouvez vous contenter de vérifier la formule dans le cas  $m = 3$ ,  $n = 4$ .

*C'est une suggestion ou un avertissement ?*

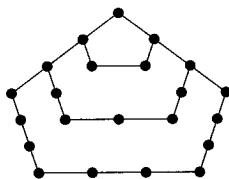
- 52 Montrez que les polynômes définis par la récurrence

$$p_n(y) = \left( y - \frac{1}{4} \right)^n - \sum_{k=0}^{n-1} \binom{2n}{2k} \left( \frac{-1}{4} \right)^{n-k} p_k(y), \quad n \geq 0 \text{ entier}$$

sont de la forme  $p_n(y) = \sum_{m=0}^n \binom{n}{m} |y^m|$ , où  $|y^m|$  est un entier strictement positif pour tout  $1 \leq m \leq n$ . *Suggestion* : cet exercice, très instructif, n'est pas vraiment facile.

- 53 La suite des  *nombres pentagonaux*  $\langle 1, 5, 12, 22, \dots \rangle$  est une générali-

sation évidente des suites des nombres triangulaires et carrés :



Soient  $T_n = n(n+1)/2$  le  $n$ ième nombre triangulaire,  $P_n = n(3n-1)/2$  le  $n$ ième nombre pentagonal et  $U_n$  le nombre de pavages  $3 \times n$  par des dominos défini en (7.38). Démontrez que le nombre triangulaire  $T_{(U_{4n+2}-1)/2}$  est aussi un nombre pentagonal. *Suggestion* :  $3U_{2n}^2 = (V_{2n-1} + V_{2n+1})^2 + 2$ .

- 54 Considérez la curieuse construction suivante :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	...
1	2	3	4		6	7	8	9		11	12	13	14		16	...
1	3	6	10		16	23	31	40		51	63	76	90		106	...
1	3	6			16	23	31			51	63	76			106	...
1	4	10			26	49	80			131	194	270			376	...
1	4				26	49				131	194				376	...
1	5				31	80				211	405				781	...
1					31					211					781	...
1					32					243					1024	...

Voici comment procéder : on part d'une ligne contenant tous les entiers strictement positifs ; puis on supprime une colonne toutes les  $m$  colonnes ( $m = 5$  dans notre exemple) ; puis on remplace les nombres qui restent par des sommes partielles ; puis on supprime une colonne toutes les  $m - 1$  colonnes ; puis on remplace à nouveau les nombres qui restent par des sommes partielles, et ainsi de suite. Utilisez les fonctions génératrices pour montrer que le résultat final est la suite des puissances  $m$ ièmes. Ainsi, dans notre exemple pour  $m = 5$ , on obtient  $(1^5, 2^5, 3^5, 4^5, \dots)$ .

- 55 Montrez que si les séries  $F(z)$  et  $G(z)$  sont  $d$ -finies (cette notion est définie dans l'exercice 20), alors  $F(z) + G(z)$  et  $F(z)G(z)$  le sont aussi.

### Problèmes de recherche

- 56 Montrez qu'il n'existe pas de "forme close simple" pour les coefficients de  $z^n$  dans  $(1 + z + z^2)^n$ , en fonction de  $n$ , même si on se place dans une large classe de "formes closes simples".

#### 404 FONCTIONS GÉNÉRATRICES

- 57 Prouvez ou réfutez : si tous les coefficients de  $G(z)$  valent 0 ou 1 et si tous les coefficients de  $G(z)^2$  sont inférieurs à une constante donnée  $M$ , alors il existe une infinité de coefficients nuls dans  $G(z)^2$ .

# Probabilités discrètes

Le hasard est un élément qui intervient dans beaucoup de nos tentatives d'appréhender le monde dans lequel nous vivons. La *théorie des probabilités* permet de calculer la probabilité de réalisation d'événements complexes en supposant qu'ils obéissent à certains axiomes. Cette théorie, qui trouve des applications dans toutes les branches scientifiques, est fortement liée aux notions que nous avons vues dans les chapitres qui précèdent.

Les probabilités sont dites “discrètes” si on peut calculer la probabilité de tout événement en effectuant une sommation (et non une intégration). Comme nous commençons à être plutôt à l'aise avec les sommes, rien d'étonnant à ce que nous soyons tout à fait prêts à appliquer nos connaissances sur des calculs de probabilités et de moyennes.

## 8.1 DÉFINITIONS

*(Les lecteurs qui ne sont pas familiers avec les probabilités pourront lire avec profit l'ouvrage bien connu de Feller, qui présente une très bonne introduction au sujet [120].)*

La base des probabilités est la notion d'*espace de probabilité*. Un espace de probabilité est constitué d'un ensemble  $\Omega$  de toutes les choses qui peuvent arriver dans un problème donné et d'une règle ou d'un ensemble de règles qui associent une probabilité  $\Pr(\omega)$  à tout événement élémentaire  $\omega \in \Omega$ . La probabilité  $\Pr(\omega)$  est toujours un nombre réel positif ou nul, et tout espace de probabilités discret doit satisfaire la condition

$$\sum_{\omega \in \Omega} \Pr(\omega) = 1. \tag{8.1}$$

Ainsi, toute valeur  $\Pr(\omega)$  est comprise dans l'intervalle  $[0..1]$ . On dit que  $\Pr$  est une *loi de probabilité*, ou une *distribution de probabilité*, car son rôle est de partager une probabilité totale de 1 entre tous les événements  $\omega$ .

Par exemple, l'ensemble  $\Omega$  des événements élémentaires lorsqu'on jette une paire de dés est  $D^2 = \{\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}, \begin{smallmatrix} \bullet & \bullet \\ \bullet & \circ \end{smallmatrix}, \dots, \begin{smallmatrix} \circ & \circ \\ \circ & \circ \end{smallmatrix}\}$ , où

$$D = \{\begin{smallmatrix} \bullet \\ \bullet \end{smallmatrix}, \begin{smallmatrix} \bullet & \circ \\ \bullet & \circ \end{smallmatrix}, \begin{smallmatrix} \bullet & \circ \\ \circ & \bullet \end{smallmatrix}, \begin{smallmatrix} \bullet & \bullet \\ \bullet & \circ \end{smallmatrix}, \begin{smallmatrix} \bullet & \bullet \\ \circ & \bullet \end{smallmatrix}, \begin{smallmatrix} \circ & \circ \\ \circ & \circ \end{smallmatrix}\}$$

désigne l'ensemble des six faces d'un dé. Nous considérerons que deux résultats comme  $\begin{array}{|c|c|}\hline \bullet & \bullet \\ \hline \end{array}$  et  $\begin{array}{|c|c|}\hline \bullet & \circ \\ \hline \end{array}$  sont distincts (en prenant par exemple deux dés de couleurs différentes). Ainsi, l'espace de probabilité contient  $6^2 = 36$  éléments.

On suppose généralement que les dés sont "justes", c'est-à-dire que chacune des faces d'un dé a une probabilité  $\frac{1}{6}$  d'apparaître, et donc que chacun des événements de  $\Omega$  a la probabilité  $\frac{1}{36}$ . On peut aussi considérer que les dés sont pipés, donc qu'ils donnent lieu à une distribution de probabilité différente. Soient par exemple

$$\begin{aligned} \Pr_1(\begin{array}{|c|c|}\hline \bullet & \\ \hline \end{array}) &= \Pr_1(\begin{array}{|c|c|}\hline \bullet & \bullet \\ \hline \end{array}) = \frac{1}{4}; \\ \Pr_1(\begin{array}{|c|c|}\hline \bullet & \circ \\ \hline \end{array}) &= \Pr_1(\begin{array}{|c|c|}\hline \circ & \bullet \\ \hline \end{array}) = \Pr_1(\begin{array}{|c|c|}\hline \bullet & \bullet \\ \hline \end{array}) = \Pr_1(\begin{array}{|c|c|}\hline \bullet & \circ \\ \hline \end{array}) = \frac{1}{8}. \end{aligned}$$

On a  $\sum_{d \in D} \Pr_1(d) = 1$ , donc  $\Pr_1$  est bien une distribution de probabilité sur l'ensemble  $D$ , et on peut associer des probabilités aux éléments de  $\Omega = D^2$  en appliquant la règle

$$\Pr_{11}(d d') = \Pr_1(d) \Pr_1(d'). \quad (8.2)$$

Par exemple,  $\Pr_{11}(\begin{array}{|c|c|}\hline \bullet & \bullet \\ \hline \end{array} \begin{array}{|c|c|}\hline \bullet & \circ \\ \hline \end{array}) = \frac{1}{4} \cdot \frac{1}{8} = \frac{1}{32}$ . Cette distribution est correcte car

$$\begin{aligned} \sum_{\omega \in \Omega} \Pr_{11}(\omega) &= \sum_{dd' \in D^2} \Pr_{11}(dd') = \sum_{d, d' \in D} \Pr_1(d) \Pr_1(d') \\ &= \sum_{d \in D} \Pr_1(d) \sum_{d' \in D} \Pr_1(d') = 1 \cdot 1 = 1. \end{aligned}$$

On peut aussi considérer le cas où l'un des dés est juste tandis que l'autre est pipé,

$$\Pr_{01}(d d') = \Pr_0(d) \Pr_1(d'), \quad \text{où } \Pr_0(d) = \frac{1}{6}, \quad (8.3)$$

auquel cas  $\Pr_{01}(\begin{array}{|c|c|}\hline \bullet & \bullet \\ \hline \end{array} \begin{array}{|c|c|}\hline \bullet & \circ \\ \hline \end{array}) = \frac{1}{6} \cdot \frac{1}{8} = \frac{1}{48}$ . Dans le "monde réel", les dés ne sont jamais tout à fait justes car il est impossible d'obtenir une symétrie parfaite ; toutefois, la valeur  $\frac{1}{6}$  se trouve généralement très proche de la réalité.

Un événement est un sous-ensemble de  $\Omega$ . Lorsqu'on joue aux dés, par exemple, l'ensemble

$$\{\begin{array}{|c|c|}, \begin{array}{|c|c|}, \begin{array}{|c|c|}, \begin{array}{|c|c|}, \begin{array}{|c|c|}, \begin{array}{|c|c|} \end{array} \end{array} \end{array} \end{array} \}$$

*Attention, ça va se gâter.*

*Si toutes les faces d'un cube étaient identiques, comment pourrions-nous savoir quelle face se trouve au-dessus ?*

constitue l'événement "on a tiré un double". Les éléments  $\omega$  de  $\Omega$  sont appelés *événements élémentaires* car ils ne peuvent pas décomposés ; l'élément  $\omega$  peut être considéré comme l'événement singleton  $\{\omega\}$ .

La probabilité d'un événement  $A$  est définie par la formule

$$\Pr(\omega \in A) = \sum_{\omega \in A} \Pr(\omega). \quad (8.4)$$

Si  $R(\omega)$  est une proposition concernant  $\omega$ , on désigne par " $\Pr(R(\omega))$ " la somme de toutes les probabilités  $\Pr(\omega)$  telles que  $R(\omega)$  est vraie. Ainsi, par exemple, la probabilité de tirer un double avec des dés justes est égale à  $\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}$ ; en revanche, si les deux dés sont lancés avec la distribution de probabilité  $\Pr_1$ , alors on trouve  $\frac{1}{16} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} = \frac{3}{16} > \frac{1}{6}$ . En pipant les dés de cette manière, on rend l'événement "tirer un double" plus probable.

Vous avez peut-être remarqué que nous utilisons ici la notation  $\sum$  dans un sens plus général que celui que nous avons défini au chapitre 2 : les sommes de (8.1) et (8.4) portent sur tous les éléments  $\omega$  d'un ensemble qui n'est pas nécessairement un ensemble de nombres entiers. Malgré tout, il n'y a pas de raison de s'inquiéter : il suffit, comme nous l'avons fait, d'utiliser une notation particulière sous le signe  $\sum$  chaque fois qu'on somme sur autre chose que des entiers pour qu'il n'y ait pas de confusion possible. Les autres définitions du chapitre 2 sont toujours correctes dans ce nouveau contexte. En particulier, les sommes infinies, telles qu'elles ont été définies au chapitre 2, correspondent bien à l'interprétation des sommes qui nous intéressent actuellement lorsque l'ensemble  $\Omega$  est infini. Comme toutes les probabilités sont positives ou nulles et comme leur somme est majorée par la constante 1, la probabilité de l'événement  $A$  de (8.4) est bien définie pour tout  $A \subseteq \Omega$ .

Une *variable aléatoire* est une fonction définie sur les événements élémentaires  $\omega$  d'un espace de probabilité. Par exemple, si  $\Omega = D^2$ , on peut définir  $S(\omega)$  comme étant la somme des valeurs des deux dés du tirage  $\omega$ , de sorte que  $S(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}) = 6 + 3 = 9$ . La probabilité de tirer un sept est égale à la probabilité de l'événement  $S(\omega) = 7$ , soit

$$\Pr(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}) + \Pr(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}) + \Pr(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}) \\ + \Pr(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}) + \Pr(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}) + \Pr(\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}).$$

Si on jette des dés justes ( $\Pr = \Pr_{00}$ ), cela arrive avec une probabilité  $\frac{1}{6}$ ; si les dés sont pipés ( $\Pr = \Pr_{11}$ ), la probabilité devient  $\frac{1}{16} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} = \frac{3}{16}$ , la même que pour les doubles.

On peut généralement se permettre de ne pas écrire " $(\omega)$ " lorsqu'on manipule des variables aléatoires car, dans un problème donné, on n'a affaire dans la plupart des cas qu'à un seul espace de probabilité. Ainsi, nous pouvons simplement écrire " $S = 7$ " pour désigner l'événement "on a tiré 7", et " $S = 4$ " pour désigner l'événement  $\{\begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}, \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}, \begin{smallmatrix} \bullet & \bullet \\ \bullet & \bullet \end{smallmatrix}\}$ .

Toute variable aléatoire peut être caractérisée par la distribution de probabilité de ses valeurs. Ainsi, par exemple, on peut écrire dans un tableau les probabilités des onze valeurs  $\{2, 3, \dots, 12\}$  que peut prendre  $S$  :

$s$	2	3	4	5	6	7	8	9	10	11	12
$Pr_{00}(S = s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$Pr_{11}(S = s)$	$\frac{4}{64}$	$\frac{4}{64}$	$\frac{5}{64}$	$\frac{6}{64}$	$\frac{7}{64}$	$\frac{12}{64}$	$\frac{7}{64}$	$\frac{6}{64}$	$\frac{5}{64}$	$\frac{4}{64}$	$\frac{4}{64}$

Si on travaille sur un problème concernant la variable aléatoire  $S$ , mais qui ne préjuge d'aucune autre propriété des dés, les seules probabilités de la table ci-dessus suffisent pour trouver la réponse, sans qu'on ait à se préoccuper de la structure de l'ensemble  $\Omega = D^2$ . En fait, on pourrait considérer que l'espace de probabilité est l'ensemble  $\Omega = \{2, 3, \dots, 12\}$ , associé à la distribution de probabilité  $Pr(s)$  désirée. Dans ce cas, " $S = 4$ " constituerait un événement élémentaire. Retenons donc qu'on peut très souvent se permettre d'ignorer l'espace de probabilité sous-jacent  $\Omega$  et travailler directement sur les variables aléatoires et leurs distributions.

Si deux variables aléatoires  $X$  et  $Y$  sont définies sur le même espace de probabilité  $\Omega$ , on peut déterminer leur comportement sans rien savoir de  $\Omega$ , à condition de connaître la "distribution conjointe"

$$Pr(X = x \text{ et } Y = y)$$

pour toute valeur  $x$  que peut prendre  $X$  et toute valeur  $y$  que peut prendre  $Y$ . On dit que  $X$  et  $Y$  sont des variables aléatoires *indépendantes* si

$$Pr(X = x \text{ et } Y = y) = Pr(X = x) \cdot Pr(Y = y) \quad (8.5)$$

pour tout  $x$  et tout  $y$ . Intuitivement, cela signifie que la valeur de  $X$  n'a aucun effet sur celle de  $Y$ , et réciproquement.

Par exemple, si  $\Omega$  est l'ensemble  $D^2$  des lancers de deux dés, soit  $S_1$  le chiffre tiré sur le premier dé et  $S_2$  le chiffre tiré sur le second. Alors les variables aléatoires  $S_1$  et  $S_2$  sont indépendantes, pour chacune des trois distributions de probabilité  $Pr_{00}$ ,  $Pr_{11}$  et  $Pr_{01}$ . En effet, nous avons défini la probabilité de chaque événement élémentaire  $dd'$  comme le produit de la probabilité de  $S_1 = d$  et de la probabilité de  $S_2 = d'$ . Nous aurions pu décider d'une définition différente, pour que, par exemple,

$$Pr(\square \blacksquare) / Pr(\square \blacksquare) \neq Pr(\square \bullet \blacksquare) / Pr(\square \bullet \blacksquare);$$

nous ne l'avons pas fait car deux dés différents ne sont pas sensés s'influencer mutuellement. Avec les définitions dont nous avons convenu, ces rapports valent tous deux  $Pr(S_2 = 5) / Pr(S_2 = 6)$ .

La variable aléatoire  $S$  est la somme  $S_1 + S_2$  des valeurs des dés. Considérons maintenant une autre variable aléatoire  $P$ , définie par le produit

$S_1 S_2$ . Peut-on dire que  $S$  et  $P$  sont indépendantes ? Intuitivement non, car si  $S = 2$ , alors  $P$  est forcément égal à 1. Formellement non plus, car, à l'évidence, la condition d'indépendance (8.5) n'est pas respectée (au moins dans le cas où les dés sont justes) : pour chaque valeur possible de  $s$  et  $p$ , on a  $0 < \Pr_{00}(S=s) \cdot \Pr_{00}(P=p) \leq \frac{1}{6} \cdot \frac{1}{9}$  ; cela ne peut pas être égal à  $\Pr_{00}(S=s \text{ et } P=p)$  qui est un multiple de  $\frac{1}{36}$ .

Pour avoir une idée générale du comportement d'une variable aléatoire donnée, on cherche souvent sa moyenne. La *moyenne* d'une suite de nombres est égale à la somme de ses valeurs divisée par le nombre de valeurs. Il existe deux autres notions proches de celle-ci : la *médiane*, qui est la valeur du "milieu" de la suite, c'est-à-dire le nombre de la suite tel qu'il existe dans la suite autant de nombres qui lui sont supérieurs ou égaux que de nombres qui lui sont inférieurs ou égaux (si la suite contient un nombre pair d'éléments, la médiane est égale à la moyenne des deux nombres "centraux") ; et le *mode*, qui est la valeur la plus fréquente dans la suite. Par exemple, la moyenne de la suite  $(3, 1, 4, 1, 5)$  est  $\frac{3+1+4+1+5}{5} = 2,8$  ; sa médiane vaut 3, et son mode est 1.

Ces trois notions peuvent être définies aussi sur des variables aléatoires plutôt que sur des suites. La *moyenne* d'une variable aléatoire  $X$  à valeurs réelles sur un espace de probabilité  $\Omega$  est égale, par définition, à

$$\sum_{x \in X(\Omega)} x \cdot \Pr(X=x) \quad (8.6)$$

à condition que cette somme éventuellement infinie existe (ici,  $X(\Omega)$  désigne l'ensemble des valeurs que peut prendre  $X$ ). La *médiane* de  $X$  est l'ensemble de tous les  $x$  tels que

$$\Pr(X \leq x) \geq \frac{1}{2} \quad \text{et} \quad \Pr(X \geq x) \geq \frac{1}{2}. \quad (8.7)$$

Enfin, le *mode* de  $X$  est l'ensemble de tous les  $x$  tels que

$$\Pr(X=x) \geq \Pr(X=x') \quad \text{pour tout } x' \in X(\Omega). \quad (8.8)$$

Remarquez que ces trois définitions, concernant des variables aléatoires, s'accordent bien avec celles concernant les suites de nombres. En effet, supposons qu'on répète une même expérience (par exemple jeter deux dés) un grand nombre de fois, de sorte que chaque valeur de  $X$  (ou  $S$  ou  $P$ , dans le cas des dés) finisse par apparaître avec une fréquence à peu près proportionnelle à sa probabilité. Alors les moyenne, médiane et mode de la suite de nombres obtenue sont proches des moyenne, médiane et mode de la variable aléatoire  $X$ .

Dans notre exemple du jeter de dés, la moyenne de  $S$  vaut  $2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \dots + 12 \cdot \frac{1}{36} = 7$  dans la distribution  $\Pr_{00}$  ; elle est aussi égale à 7 dans  $\Pr_{11}$ .

La médiane et le mode valent {7} aussi dans les deux distributions. La moyenne de  $P$  dans la distribution  $\Pr_{00}$  est égale à  $\frac{49}{4} = 12,25$ , sa médiane vaut {10} et son mode {6, 12}. Dans la distribution de  $\Pr_{11}$ , sa moyenne reste la même mais sa médiane tombe à {8} tandis que son mode devient le singleton {6}.

Les théoriciens des probabilités ont un autre nom pour la moyenne d'une variable aléatoire : ils l'appellent l'*espérance* et l'écrivent

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega) \Pr(\omega). \quad (8.9)$$

Dans notre exemple, cette somme contient 36 termes (un pour chaque élément de  $\Omega$ ), alors que (8.6) est une somme de onze termes seulement. Toutefois, elles sont toutes deux égales à

$$\sum_{\substack{\omega \in \Omega \\ x \in X(\Omega)}} x \Pr(\omega) [x = X(\omega)].$$

Parmi les trois notions que nous venons d'introduire, la plus utile est sans conteste la moyenne. C'est pourquoi, à partir de maintenant, nous allons complètement oublier tout ce qui concerne la médiane et le mode. Au cours du chapitre, nous utiliserons indifféremment les termes "moyenne" et "espérance".

Si  $X$  et  $Y$  sont deux variables aléatoires définies sur un même espace de probabilité, alors  $X + Y$  est une variable aléatoire sur cet espace. D'après la formule (8.9), la moyenne de leur somme est égale à la somme de leurs moyennes :

$$\begin{aligned} \mathbb{E}(X + Y) &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \Pr(\omega) \\ &= \mathbb{E}X + \mathbb{E}Y. \end{aligned} \quad (8.10)$$

De même, pour toute constante  $\alpha$  on a

$$\mathbb{E}(\alpha X) = \alpha \mathbb{E}X. \quad (8.11)$$

La règle correspondante pour la multiplication de deux variables aléatoires est en général plus compliquée, car l'espérance est définie comme une somme sur des événements élémentaires et la plupart des sommes de produits sont difficiles à simplifier. Néanmoins, dans le cas particulier où les deux variables aléatoires sont indépendantes, on a une jolie formule :

$$\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y), \quad \text{si } X \text{ et } Y \text{ sont indépendantes.} \quad (8.12)$$

*J'ai compris, du moins je l'espère : en moyenne, "moyenne" signifie "espérance".*

Ceci peut se prouver à l'aide de la règle de distributivité :

$$\begin{aligned}
 E(XY) &= \sum_{\omega \in \Omega} X(\omega)Y(\omega) \cdot \Pr(\omega) \\
 &= \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xy \cdot \Pr(X=x \text{ et } Y=y) \\
 &= \sum_{\substack{x \in X(\Omega) \\ y \in Y(\Omega)}} xy \cdot \Pr(X=x) \Pr(Y=y) \\
 &= \sum_{x \in X(\Omega)} x \Pr(X=x) \cdot \sum_{y \in Y(\Omega)} y \Pr(Y=y) = (EX)(EY).
 \end{aligned}$$

Par exemple, nous savons que  $S = S_1 + S_2$  et  $P = S_1S_2$ , où  $S_1$  et  $S_2$  sont les chiffres respectifs donnés par le premier et le second dé après un jet. Comme  $ES_1 = ES_2 = \frac{7}{2}$ , on a  $ES = 7$ ; de plus, du fait que  $S_1$  et  $S_2$  sont indépendantes,  $EP$  vaut bien  $\frac{7}{2} \cdot \frac{7}{2} = \frac{49}{4}$ , comme nous le prétendions il y a peu. D'autre part,  $E(S + P) = ES + EP = 7 + \frac{49}{4}$ . Cependant, comme  $S$  et  $P$  ne sont pas indépendantes, on n'a pas le droit d'affirmer que  $E(SP) = 7 \cdot \frac{49}{4} = \frac{343}{4}$ . En réalité, l'espérance de  $SP$  vaut  $\frac{637}{6}$  dans la distribution  $\Pr_{00}$  et 112 (exactement) dans la distribution  $\Pr_{11}$ .

## 8.2 MOYENNE ET VARIANCE

La deuxième caractéristique importante d'une variable aléatoire, après son espérance, est sa *variance*

$$VX = E((X - EX)^2). \quad (8.13)$$

Si on pose  $\mu = EX$ , la variance  $VX$  est égale à la moyenne de  $(X - \mu)^2$ . Elle permet de se faire une idée de la "répartition" de la distribution de  $X$ .

Voici un exemple de calcul de variance. Supposons qu'on nous donne de quoi acheter deux tickets de loterie, et que c'est une offre que nous ne pouvons pas refuser. Les organisateurs de cette loterie impriment 100 tickets par semaine. L'un de ces tickets, dont le numéro est tiré uniformément au hasard (cela signifie que tous les tickets ont même probabilité), permet à son possesseur de gagner la coquette somme de cent millions de francs ; les 99 autres ne gagnent rien du tout.

Nous avons le choix entre deux possibilités : soit prendre nos deux tickets pour le même tirage (la même semaine), soit en prendre un pour un tirage donné et l'autre pour un tirage différent (donc ne pas les acheter la même semaine). Quelle est la meilleure stratégie ? Soient  $X_1$  et  $X_2$  les variables aléatoires qui représentent la somme que nous gagnons avec le

*(C'est un tout petit peu subtil : il y a deux espaces de probabilité différents selon la stratégie que nous choisissons, mais  $EX_1$  et  $EX_2$  sont égales).*

premier et le second ticket respectivement. L'espérance de  $X_1$  est

$$\mathbb{E}X_1 = \frac{99}{100} \cdot 0 + \frac{1}{100} \cdot 100 = 1,$$

exprimée en millions de francs ; on trouve exactement la même chose pour l'espérance de  $X_2$ . Par conséquent, d'après (8.10), notre gain moyen sera

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2 = 2 \text{ millions},$$

quelle que soit la stratégie adoptée.

Pourtant, ces deux stratégies ont l'air différentes. Pour nous en assurer, allons plus loin en étudiant l'exakte distribution de probabilité de  $X_1 + X_2$  :

*Au-delà de nos espérances...*

		gains (en millions)		
		0	100	200
même tirage	0,9800	0,0200		
	0,9801	0,0198	0,0001	

Si nous achetons les deux tickets la même semaine, nous avons 98% de chances de ne rien gagner et 2% de chances de gagner 100 millions. Si nous les achetons pendant deux semaines différentes, nous avons 98,01% de chances de ne rien gagner ce qui fait un tout petit peu plus que précédemment, 0,01% de chances de gagner 200 millions ce qui est aussi un tout petit peu supérieur à ce que nous avions avant, tandis que nos chances de gagner 100 millions diminuent pour atteindre et 1,98%. Dans ce deuxième cas, la distribution de  $X_1 + X_2$  est donc plus "étalée" ; la valeur du milieu, 100 millions, est un peu moins probable mais les valeurs extrêmes sont un peu plus probables.

La variance permet de quantifier cette notion de répartition plus ou moins étalée d'une variable aléatoire. Nous la mesurons en considérant le carré de la déviation de la variable aléatoire par rapport à sa moyenne. Dans le premier cas, la variance vaut

$$0,98(0M - 2M)^2 + 0,02(100M - 2M)^2 = 196M^2$$

alors que dans le deuxième cas elle est égale à

$$0,9801(0M - 2M)^2 + 0,0198(100M - 2M)^2 + 0,0001(200M - 2M)^2 = 198M^2.$$

Nous trouvons bien une variance plus grande dans ce dernier cas, du fait que la distribution est légèrement plus étalée que dans le premier cas.

Lorsqu'on travaille avec des variances, les nombres peuvent devenir très grands car tout est élevé au carré (le facteur  $M^2$  vaut un billion, ce qui est

*Voilà qui est intéressant : la variance d'une somme en franc s'exprime en francs carrés.*

plutôt impressionnant, même pour des joueurs invétérés). Pour revenir à une échelle plus raisonnable, on considère habituellement la racine carrée de la variance. On obtient ainsi ce qu'on appelle l'*écart-type*, que l'on désigne généralement par la lettre grecque  $\sigma$  :

$$\sigma = \sqrt{VX}. \quad (8.14)$$

Les écarts-types des variables aléatoires  $X_1 + X_2$  de nos deux stratégies valent respectivement  $\sqrt{196M^2} = 14,00M$  et  $\sqrt{198M^2} \approx 14,071247M$ . Dans un certain sens, la seconde stratégie est plus risquée, et le risque supplémentaire "vaut" en quelque sorte 71 247 francs.

Comment la variance peut-elle nous aider à choisir une stratégie ? Il n'y a pas de réponse catégorique à cette question. Bien sûr, la stratégie qui admet la plus grande variance est un peu plus risquée ; mais vaut-il mieux jouer "pépère" ou prendre un peu plus de risques ? Supposez que nous ayons la possibilité d'acheter 100 tickets au lieu de deux seulement. Dans ce cas, le gain est garanti si nous les achetons tous au même tirage (et la variance est nulle). D'un autre côté, si nous participons à cent tirages différents, nous ne gagnerons rien avec une probabilité  $0,99^{100} \approx 0,366$ , mais nous aurons une probabilité non nulle de gagner jusqu'à 10 milliards de francs. Notre but n'est pas de décider quelle est la meilleure alternative ; contentons-nous simplement de faire les calculs.

Il existe en fait une manière plus simple de calculer la variance que celle donnée dans la définition (8.13). Comme  $(EX)$  est une constante, on a

$$\begin{aligned} E((X - EX)^2) &= E(X^2 - 2X(EX) + (EX)^2) \\ &= E(X^2) - 2(EX)(EX) + (EX)^2. \end{aligned}$$

Par conséquent,

$$VX = E(X^2) - (EX)^2. \quad (8.15)$$

"La variance est égale à la moyenne des carrés moins le carré de la moyenne".

Il n'y a donc rien d'étonnant à ce que les variances de notre loterie aient été exactement des nombres entiers de billions de francs. Par exemple, la moyenne de  $(X_1 + X_2)^2$  vaut  $0,98(0M)^2 + 0,02(100M)^2 = 200M^2$  ou  $0,9801(0M)^2 + 0,0198(100M)^2 + 0,0001(200M)^2 = 202M^2$ , selon le cas. En soustrayant  $4M^2$  (le carré de la moyenne) nous retombons sur les résultats précédemment calculés.

Voyons maintenant un résultat particulièrement utile pour calculer  $V(X+Y)$  si  $X$  et  $Y$  sont indépendantes : comme  $E(XY) = (EX)(EY)$  dans ce cas, on a

$$E((X+Y)^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2(EX)(EY) + E(Y^2).$$

*Pour réduire les risques, il y a un bon moyen : souoyer les organisateurs. On obtient alors des probabilités indiscrètes.*

*(N.B. : la direction décline toute responsabilité quant aux opinions exprimées dans ces marges).*

Par conséquent,

$$\begin{aligned}
 V(X+Y) &= E((X+Y)^2) - (EX+EY)^2 \\
 &= E(X^2) + 2(EX)(EY) + E(Y^2) \\
 &\quad - (EX)^2 - 2(EX)(EY) - (EY)^2 \\
 &= E(X^2) - (EX)^2 + E(Y^2) - (EY)^2 \\
 &= VX + VY.
 \end{aligned} \tag{8.16}$$

“La variance d'une somme de variables aléatoires indépendantes est égale à la somme de leurs variances”. Par exemple, la variance de la somme que nous pouvons gagner avec un unique ticket vaut

$$E(X_1^2) - (EX_1)^2 = 0,99(0M)^2 + 0,01(100M)^2 - (1M)^2 = 99M^2.$$

Ainsi, la variance du gain total de deux tickets dans deux tirages différents (donc indépendants) est égale à  $2 \times 99M^2 = 198M^2$ . Plus généralement, la variance pour  $n$  tickets dans  $n$  tirages indépendants vaut  $n \times 99M^2$ .

Revenons un instant à nos jets de dés. La variance de la somme des valeurs des dés  $S$  se calcule avec la formule ci-dessus, car  $S = S_1 + S_2$  est la somme de deux variables aléatoires indépendantes. Comme

$$VS_1 = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

si les dés sont justes, on a  $VS = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}$ . Si on lance un dé pipé, alors

$$VS_1 = \frac{1}{8}(2 \cdot 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 2 \cdot 6^2) - \left(\frac{7}{2}\right)^2 = \frac{45}{12};$$

par conséquent,  $VS = \frac{45}{6} = 7,5$  si les deux dés sont pipés. Remarquez que la variance de  $S$  est plus élevée si les dés sont pipés, mais que tirer la valeur 7, moyenne dans les deux cas, est plus probable si les dés sont pipés que s'ils sont justes. Si le but du jeu est de tirer le plus possible de 7, la variance n'est absolument pas le meilleur critère de réussite.

Bon. Nous savons maintenant calculer des variances. Par contre, nous ne savons pas vraiment pourquoi nous calculons ce paramètre. Par quelle vertu est-il naturel que d'autres que nous pourrions imaginer ? La réponse réside dans l'*inégalité de Tchebychev* ([29] et [57]), qui met en évidence une propriété importante de la variance :

$$\Pr((X-EX)^2 \geq \alpha) \leq VX/\alpha, \quad \text{pour tout } \alpha > 0 \tag{8.17}$$

(à ne pas confondre avec les inégalités monotones de Tchebychev que nous avons vues au chapitre 2).

En gros, la formule (8.17) signifie que si la variance  $VX$  d'une variable aléatoire  $X$  est petite, sa valeur sera rarement loin de sa moyenne  $EX$ . La preuve de cette propriété est étonnamment simple :

$$\begin{aligned} VX &= \sum_{\omega \in \Omega} (X(\omega) - EX)^2 \Pr(\omega) \\ &\geq \sum_{\substack{\omega \in \Omega \\ (X(\omega) - EX)^2 \geq \alpha}} (X(\omega) - EX)^2 \Pr(\omega) \\ &\geq \sum_{\substack{\omega \in \Omega \\ (X(\omega) - EX)^2 \geq \alpha}} \alpha \Pr(\omega) = \alpha \cdot \Pr((X - EX)^2 \geq \alpha); \end{aligned}$$

il suffit de diviser par  $\alpha$  pour conclure.

Si on note  $\mu$  la moyenne et  $\sigma$  l'écart-type et si on remplace  $\alpha$  par  $c^2VX$  dans (8.17), la condition  $(X - EX)^2 \geq c^2VX$  devient  $(X - \mu)^2 \geq (c\sigma)^2$ ; donc, (8.17) entraîne que

$$\Pr(|X - \mu| \geq c\sigma) \leq 1/c^2. \quad (8.18)$$

Ainsi, la valeur de  $X$  se situe presque toujours à moins de  $c$  fois l'écart-type de sa valeur moyenne ; elle n'est en dehors de l'intervalle ainsi défini qu'avec une probabilité  $1/c^2$ . Toute variable aléatoire a au moins 75% de chances de se trouver à moins de  $2\sigma$  de  $\mu$  ; elle a au moins 99% de chances de se situer entre  $\mu - 10\sigma$  et  $\mu + 10\sigma$ . Ces résultats constituent les cas particuliers  $\alpha = 4VX$  et  $\alpha = 100VX$  de l'inégalité de Tchebychev.

Si on jette deux dés  $n$  fois de suite, où  $n$  est un grand nombre, la somme des valeurs des jets sera presque toujours proche de  $7n$ . En effet, la variance de  $n$  jets indépendants est égale à  $\frac{35}{6}n$ , donc l'écart-type vaut

$$\sqrt{\frac{35}{6}n}.$$

L'inégalité de Tchebychev entraîne que la somme finale sera entre

$$7n - 10\sqrt{\frac{35}{6}n} \text{ et } 7n + 10\sqrt{\frac{35}{6}n}$$

dans au moins 99% des cas si on utilise des dés justes. Par exemple, si on jette un million de fois les dés, il y a 99% de chances que le total soit compris entre 6,975 millions et 7,025 millions.

Voyons quelque chose de plus général. Soit  $X$  une variable aléatoire quelconque sur un espace de probabilité  $\Omega$ , admettant une moyenne  $\mu$  et un écart-type  $\sigma$  finis. Considérons alors l'espace de probabilité  $\Omega^n$ , dont les événements élémentaires sont des  $n$ -uplets  $(\omega_1, \omega_2, \dots, \omega_n)$  tels que

$\omega_k \in \Omega$  pour tout  $k$ , et dont les probabilités sont définies par

$$\Pr(\omega_1, \omega_2, \dots, \omega_n) = \Pr(\omega_1) \Pr(\omega_2) \dots \Pr(\omega_n).$$

Si on définit un ensemble de variables aléatoires  $X_k$  par la formule

$$X_k(\omega_1, \omega_2, \dots, \omega_n) = X(\omega_k),$$

alors la quantité

$$X_1 + X_2 + \dots + X_n$$

désigne une somme de  $n$  variables aléatoires indépendantes. Cela revient à prendre  $n$  "échantillons" indépendants de  $X$  sur  $\Omega$  et les additionner. Comme la moyenne de  $X_1 + X_2 + \dots + X_n$  est égale à  $n\mu$  et son écart-type est  $\sqrt{n}\sigma$ , la moyenne des  $n$  échantillons,

$$\frac{1}{n}(X_1 + X_2 + \dots + X_n),$$

se trouvera entre  $\mu - 10\sigma/\sqrt{n}$  et  $\mu + 10\sigma/\sqrt{n}$  pendant au moins 99% du temps. En d'autres termes, si  $n$  est choisi assez grand, alors la moyenne de  $n$  échantillons indépendants sera presque toujours extrêmement proche de l'espérance  $EX$ . (Il existe même un résultat plus fort, la "loi forte des grands nombres", qui est démontré dans tous les livres de théorie des probabilités ; toutefois, pour notre propos, le résultat simple que nous venons de déduire de l'inégalité de Tchebychev suffit largement).

Il arrive parfois qu'on ait besoin d'estimer la moyenne d'une variable aléatoire  $X$  sans rien connaître des caractéristiques de son espace de probabilité (par exemple, si on veut connaître la température moyenne à midi en janvier à San Francisco, ou la durée de vie moyenne des agents d'assurances). Pour cela, on observe des échantillons de  $X$  : si on a récolté des valeurs empiriques indépendantes  $X_1, X_2, \dots, X_n$ , on peut raisonnablement considérer que la véritable moyenne de  $X$  est à peu près égale à

$$\hat{E}X = \frac{X_1 + X_2 + \dots + X_n}{n}. \quad (8.19)$$

(Cela signifie que, pour tout  $n$  fixé, la moyenne d'une ensemble de  $n$  échantillons indépendants sera entre les limites données dans au moins 99% des cas. Ne croyez surtout pas qu'il s'agisse de la moyenne d'une suite infinie  $X_1, X_2, X_3, \dots$  lorsque  $n$  varie).

On peut aussi estimer la variance avec la formule

$$\hat{V}X = \frac{X_1^2 + X_2^2 + \dots + X_n^2}{n-1} - \frac{(X_1 + X_2 + \dots + X_n)^2}{n(n-1)}. \quad (8.20)$$

Notez bien les facteurs  $(n-1)$  dans les dénominateurs. A priori, on aurait envie de les remplacer par  $n$ , comme dans (8.19), parce que la véritable variance  $VX$  se calcule, dans (8.15), en fonction de l'espérance. Pourtant,

on obtient une meilleure estimation avec  $n - 1$  qu'avec  $n$ , car la définition (8.20) implique que

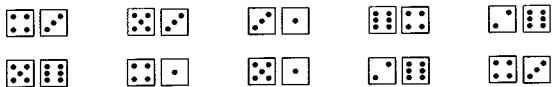
$$\mathbb{E}(\widehat{V}X) = VX. \quad (8.21)$$

Voici pourquoi :

$$\begin{aligned}\mathbb{E}(\widehat{V}X) &= \frac{1}{n-1} \mathbb{E} \left( \sum_{k=1}^n X_k^2 - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n X_j X_k \right) \\ &= \frac{1}{n-1} \left( \sum_{k=1}^n \mathbb{E}(X_k^2) - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \mathbb{E}(X_j X_k) \right) \\ &= \frac{1}{n-1} \left( \sum_{k=1}^n \mathbb{E}(X^2) - \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n (\mathbb{E}(X)^2[j \neq k] + \mathbb{E}(X^2)[j = k]) \right) \\ &= \frac{1}{n-1} \left( n\mathbb{E}(X^2) - \frac{1}{n} (n\mathbb{E}(X^2) + n(n-1)\mathbb{E}(X)^2) \right) \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = VX.\end{aligned}$$

Remarquez que, dans ce calcul, on fait usage de l'hypothèse d'indépendance des observations lorsqu'on remplace  $\mathbb{E}(X_j X_k)$  par  $(\mathbb{E}X)^2[j \neq k] + \mathbb{E}(X^2)[j = k]$ .

En pratique, les résultats expérimentaux sur une variable aléatoire  $X$  sont généralement obtenus en calculant une moyenne  $\hat{\mu} = \widehat{E}X$  et un écart-type  $\hat{\sigma} = \sqrt{\widehat{V}X}$  empiriques, et en présentant le résultat sous la forme " $\hat{\mu} \pm \hat{\sigma}/\sqrt{n}$ ". Voici par exemple dix jets de deux dés supposés justes :



La moyenne empirique de la somme  $S$  vaut

$$\hat{\mu} = (7 + 11 + 8 + 5 + 4 + 6 + 10 + 8 + 8 + 7)/10 = 7,4$$

et la variance empirique est égale à

$$(7^2 + 11^2 + 8^2 + 5^2 + 4^2 + 6^2 + 10^2 + 8^2 + 8^2 + 7^2 - 10\hat{\mu}^2)/9 \approx 2,1^2.$$

Sur la base de ces résultats expérimentaux, nous obtenons une valeur estimée de  $7,4 \pm 2,1/\sqrt{10} \approx 7,4 \pm 0,7$ , pour la moyenne de la somme  $S$ .

Voyons encore un exemple de calcul, théorique cette fois, de la moyenne et de la variance d'une variable aléatoire. Nous avons considéré au chapitre 5 le problème de la victoire au football ; il s'agissait de  $n$  chapeaux lancés

en l'air et d'une permutation aléatoire de ces chapeaux. Nous avons montré dans l'équation (5.51) que la probabilité qu'aucun des supporters ne retrouve son chapeau vaut  $n!/n! \approx 1/e$ . Nous avons aussi calculé la probabilité que  $k$  personnes exactement retrouvent leurs propres chapeaux :

$$P(n, k) = \frac{1}{n!} \binom{n}{k} (n-k)! = \frac{1}{k!} \frac{(n-k)!}{(n-k)!}. \quad (8.22)$$

Présentons ces résultats avec le formalisme que nous venons de découvrir. Soit  $\Pi_n$  l'espace de probabilité contenant la totalité des  $n!$  permutations  $\pi$  de  $\{1, 2, \dots, n\}$ , avec  $Pr(\pi) = 1/n!$  pour tout  $\pi \in \Pi_n$ . La variable aléatoire

$$F_n(\pi) = \text{nombre de "points fixes" de } \pi, \quad \text{pour } \pi \in \Pi_n,$$

*A ne pas confondre avec un nombre de Fibonacci.*

compte le nombre d'appariements corrects (chapeau, propriétaire) dans le problème de la victoire au football. Bien que nous sachions pertinemment que l'équation (8.22) implique que  $Pr(F_n = k)$ , faisons comme si nous ne connaissions pas cette formule et calculons la valeur moyenne et l'écart-type de  $F_n$ .

En fait, le calcul de la moyenne est beaucoup plus facile à effectuer maintenant qu'au chapitre 5. Remarquons simplement que

$$F_n(\pi) = F_{n,1}(\pi) + F_{n,2}(\pi) + \cdots + F_{n,n}(\pi),$$

où  $F_{n,k}(\pi) = [\text{la } k\text{ème position de } \pi \text{ est un point fixe}]$ , pour  $\pi \in \Pi_n$ .

Par conséquent,  $E F_n = E F_{n,1} + E F_{n,2} + \cdots + E F_{n,n}$ . Or, l'espérance de  $F_{n,k}$  est tout bonnement égale à la probabilité que  $F_{n,k} = 1$ , qui vaut  $1/n$  car, parmi les  $n!$  permutations  $\pi = \pi_1 \pi_2 \dots \pi_n \in \Pi_n$ , il y en a exactement  $(n-1)!$  qui satisfont  $\pi_k = k$ . Par conséquent,

$$E F_n = n/n = 1, \quad \text{pour } n > 0. \quad (8.23)$$

En moyenne, il y aura un chapeau qui se retrouvera sur sa tête attitrée. "Le nombre moyen de points fixes d'une permutation aléatoire est égal à 1".

La question de l'écart type est plus délicate, car les  $F_{n,k}$  ne sont pas indépendants les uns par rapport aux autres. Nous pouvons néanmoins calculer la variance en analysant leurs dépendances mutuelles :

$$\begin{aligned} E(F_n^2) &= E\left(\left(\sum_{k=1}^n F_{n,k}\right)^2\right) = E\left(\sum_{j=1}^n \sum_{k=1}^n F_{n,j} F_{n,k}\right) \\ &= \sum_{j=1}^n \sum_{k=1}^n E(F_{n,j} F_{n,k}) = \sum_{1 \leq k \leq n} E(F_{n,k}^2) + 2 \sum_{1 \leq j < k \leq n} E(F_{n,j} F_{n,k}) \end{aligned}$$

(c'est une astuce du même genre que nous avons utilisée lors du calcul de (2.33) au chapitre 2). Comme  $F_{n,k}$  vaut toujours 0 ou 1, on a  $F_{n,k}^2 = F_{n,k}$ , donc  $E(F_{n,k}^2) = EF_{n,k} = 1/n$  comme précédemment. De plus, si  $j < k$ , alors  $E(F_{n,j} F_{n,k}) = \Pr(j \text{ et } k \text{ sont tous deux des points fixes de } \pi) = (n-2)!/n! = 1/n(n-1)$ . Par conséquent,

$$E(F_n^2) = \frac{n}{n} + \binom{n}{2} \frac{2}{n(n-1)} = 2, \quad \text{pour } n \geq 2. \quad (8.24)$$

Par exemple, pour  $n = 3$ , on a  $\frac{2}{6}0^2 + \frac{3}{6}1^2 + \frac{0}{6}2^2 + \frac{1}{6}3^2 = 2$ . La variance est  $E(F_n^2) - (EF_n)^2 = 1$ , donc l'écart-type vaut aussi 1 (tout comme la moyenne). "Une permutation aléatoire de  $n \geq 2$  éléments a  $1 \pm 1$  points fixes en moyenne."

### 8.3 FG DE PROBABILITE

Si  $X$  est une variable aléatoire à valeurs entières positives ou nulles, les techniques présentées au chapitre 7 offrent un moyen efficace de calculer sa distribution de probabilité. La *fonction génératrice de probabilité* (ou fgp) de  $X$  est définie par

$$G_X(z) = \sum_{k \geq 0} \Pr(X=k) z^k. \quad (8.25)$$

Cette série entière en  $z$  contient toutes les informations concernant la variable aléatoire  $X$ . Elle peut aussi s'exprimer de deux autres façons :

$$G_X(z) = \sum_{\omega \in \Omega} \Pr(\omega) z^{X(\omega)} = E(z^X). \quad (8.26)$$

Les coefficients de  $G_X(z)$  sont positifs ou nuls et leur somme vaut 1 ; cette dernière condition peut s'écrire

$$G_X(1) = 1. \quad (8.27)$$

Réciproquement, toute série entière  $G(z)$  à coefficients positifs ou nuls et telle que  $G(1) = 1$  est la fgp d'une certaine variable aléatoire.

Généralement, l'usage des fgp simplifie le calcul de moyennes et de variances. La moyenne, notamment, s'exprime sans difficulté :

$$\begin{aligned} EX &= \sum_{k \geq 0} k \cdot \Pr(X=k) \\ &= \sum_{k \geq 0} \Pr(X=k) \cdot kz^{k-1} \Big|_{z=1} \\ &= G'_X(1). \end{aligned} \quad (8.28)$$

Il suffit de dériver la fgp par rapport à  $z$ , puis de poser  $z = 1$ .

La variance est un tout petit peu plus compliquée :

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{k \geq 0} k^2 \cdot \Pr(X=k) \\ &= \sum_{k \geq 0} \Pr(X=k) \cdot (k(k-1)z^{k-2} + kz^{k-1}) \Big|_{z=1} \\ &= G_X''(1) + G_X'(1).\end{aligned}$$

Par conséquent,

$$VX = G_X''(1) + G_X'(1) - G_X'(1)^2. \quad (8.29)$$

Donc, selon les équations (8.28) et (8.29), il suffit, pour pouvoir calculer la moyenne et la variance, de savoir calculer les valeurs  $G_X'(1)$  et  $G_X''(1)$  des deux dérivées. Nous n'avons pas besoin de connaître une forme close des probabilités, ni même une forme close de  $G_X(z)$ .

Pour simplifier les calculs ultérieurs, convenons d'écrire, pour toute fonction  $G$ ,

$$\text{Moy}(G) = G'(1), \quad (8.30)$$

$$\text{Var}(G) = G''(1) + G'(1) - G'(1)^2. \quad (8.31)$$

Les fgp sont appréciables aussi parce que, dans la plupart des cas, elles se trouvent être des fonctions relativement simples de  $z$ . Observons par exemple la *distribution uniforme* d'ordre  $n$ , dans laquelle la variable aléatoire prend chacune des valeurs  $\{0, 1, \dots, n-1\}$  avec la probabilité  $1/n$ . La fgp correspondante

$$U_n(z) = \frac{1}{n}(1+z+\dots+z^{n-1}) = \frac{1}{n} \frac{1-z^n}{1-z}, \quad \text{pour } n \geq 1. \quad (8.32)$$

admet une forme close en tant que série géométrique.

Pourtant, cette forme close s'avère quelque peu embarrassante : si on pose  $z = 1$ , on obtient le rapport indéfini  $0/0$ , bien que  $U_n(z)$  soit un polynôme parfaitement défini pour toute valeur de  $z$ . A partir de la forme non close  $(1+z+\dots+z^{n-1})/n$ , la valeur  $U_n(1) = 1$  est immédiate ; si on veut la déduire de la forme close, il semble qu'il faille recourir à la règle de L'Hospital pour calculer  $\lim_{z \rightarrow 1} U_n(z)$ . L'évaluation de  $U_n'(1)$  par cette même règle sera plus délicate, car il y aura un facteur  $(z-1)^2$  au dénominateur ;  $U_n''(1)$  sera plus difficile encore.

Fort heureusement, il y a un moyen d'éviter ce dilemme. Si une série  $G(z) = \sum_{n \geq 0} g_n z^n$  converge au moins pour une valeur de  $z$  telle que  $|z| > 1$ , la série  $G'(z) = \sum_{n \geq 0} n g_n z^{n-1}$  a la même propriété, ainsi que  $G''(z)$ ,

$G'''(z)$ , etc. Par conséquent, on peut appliquer la formule de Taylor pour écrire

$$G(1+t) = G(1) + \frac{G'(1)}{1!}t + \frac{G''(1)}{2!}t^2 + \frac{G'''(1)}{3!}t^3 + \dots \quad (8.33)$$

Lorsqu'on développe  $G(1+t)$  en série, toutes les dérivées de  $G(z)$  en  $z=1$  apparaissent dans les coefficients.

Par exemple, on peut facilement trouver, par ce biais, les dérivées de la fgp uniforme  $U_n(z)$  :

$$\begin{aligned} U_n(1+t) &= \frac{1}{n} \frac{(1+t)^n - 1}{t} \\ &= \frac{1}{n} \binom{n}{1} + \frac{1}{n} \binom{n}{2} t + \frac{1}{n} \binom{n}{3} t^2 + \dots + \frac{1}{n} \binom{n}{n} t^{n-1}. \end{aligned}$$

En comparant ceci à (8.33), on obtient

$$U_n(1) = 1; \quad U'_n(1) = \frac{n-1}{2}; \quad U''_n(1) = \frac{(n-1)(n-2)}{3}. \quad (8.34)$$

Plus généralement,  $U_n^{(m)}(1) = (n-1)^m/(m+1)$  (mais les cas  $m=1$  et  $m=2$  nous suffisent pour calculer la moyenne et la variance). La moyenne de la distribution uniforme est donc égale à

$$U'_n(1) = \frac{n-1}{2}, \quad (8.35)$$

et voici sa variance :

$$\begin{aligned} U''_n(1) + U'_n(1) - U'_n(1)^2 &= 4 \frac{(n-1)(n-2)}{12} + 6 \frac{(n-1)}{12} - 3 \frac{(n-1)^2}{12} \\ &= \frac{n^2 - 1}{12}. \end{aligned} \quad (8.36)$$

Voici maintenant une autre raison d'apprécier les fgp. Nous avons vu aux chapitres 5 et 7 que le produit de deux fonctions génératrices correspond à la convolution des suites correspondantes. Voici un fait plus important encore, qui concerne les fgp : si deux variables aléatoires sont indépendantes, le produit de leurs fgp correspond à leur somme. En effet, si  $X$  et  $Y$  sont des variables aléatoires qui ne prennent que des valeurs entières, la probabilité que  $X+Y=n$  est

$$\Pr(X+Y=n) = \sum_k \Pr(X=k \text{ et } Y=n-k).$$

Si  $X$  et  $Y$  sont indépendantes, on en déduit que

$$\Pr(X + Y = n) = \sum_k \Pr(X = k) \Pr(Y = n - k),$$

ce qui constitue une convolution. Voici donc le résultat que nous obtenons :

$$G_{X+Y}(z) = G_X(z) G_Y(z), \quad \text{si } X \text{ et } Y \text{ sont indépendantes.} \quad (8.37)$$

Un peu plus haut dans ce chapitre, nous avons établi que  $V(X+Y) = V(X) + V(Y)$  lorsque  $X$  et  $Y$  sont indépendantes. Soient  $F(z)$  et  $G(z)$  les fgp de  $X$  et  $Y$  et  $H(z)$  la fgp de  $X + Y$ . Alors

$$H(z) = F(z)G(z),$$

et on déduit des formules (8.28) à (8.31) que

$$\text{Moy}(H) = \text{Moy}(F) + \text{Moy}(G); \quad (8.38)$$

$$\text{Var}(H) = \text{Var}(F) + \text{Var}(G). \quad (8.39)$$

Ces deux propriétés de  $\text{Moy}(H) = H'(1)$  et  $\text{Var}(H) = H''(1) + H'(1) - H'(1)^2$  ne sont pas vraies pour n'importe quel produit de fonctions  $H(z) = F(z)G(z)$ . Dans le cas général, on a

$$H'(z) = F'(z)G(z) + F(z)G'(z),$$

$$H''(z) = F''(z)G(z) + 2F'(z)G'(z) + F(z)G''(z).$$

Toutefois, si on pose  $z = 1$ , on voit qu'il suffit que

$$F(1) = G(1) = 1 \quad (8.40)$$

et que les dérivées existent pour que (8.38) et (8.39) soient vraies. Les "probabilités" n'ont même pas besoin d'être dans l'intervalle  $[0..1]$ . Si  $F(1)$  et  $G(1)$  ne valent pas 1, il suffit de normaliser les fonctions  $F(z)$  et  $G(z)$  en les divisant respectivement par  $F(1)$  et  $G(1)$  pour que la condition adéquate soit respectée, pourvu, bien sûr, que les valeurs de  $F(1)$  et de  $G(1)$  soient non nulles.

La moyenne et la variance font partie d'une série infinie de paramètres, appelés *cumulants*, qui ont été introduits par l'astronome danois Thorvald Nicolai Thiele [351] en 1903. Le deux premiers cumulants,  $\kappa_1$  et  $\kappa_2$ , d'une variable aléatoire sont exactement les paramètres que nous avons appelé moyenne et variance. Les cumulants d'ordre plus élevé permettent d'exprimer des propriétés plus subtiles d'une distribution. Si  $G(z)$  est la fgp d'une variable aléatoire, la série de ses cumulants est définie par la formule

$$\ln G(e^t) = \frac{\kappa_1}{1!}t + \frac{\kappa_2}{2!}t^2 + \frac{\kappa_3}{3!}t^3 + \frac{\kappa_4}{4!}t^4 + \dots \quad (8.41)$$

Regardons cela de plus près. Si  $G(z)$  est la fgp de  $X$ , on a

$$\begin{aligned} G(e^t) &= \sum_{k \geq 0} \Pr(X=k)e^{kt} = \sum_{k,m \geq 0} \Pr(X=k) \frac{k^m t^m}{m!} \\ &= 1 + \frac{\mu_1}{1!}t + \frac{\mu_2}{2!}t^2 + \frac{\mu_3}{3!}t^3 + \dots, \quad (8.42) \end{aligned}$$

où  $\mu_m$ , défini par

$$\mu_m = \sum_{k \geq 0} k^m \Pr(X=k) = E(X^m), \quad (8.43)$$

est appelé le “moment d’ordre  $m$ ” de  $X$ . Si on prend l’exponentielle de chacun des deux membres de (8.41), on obtient une nouvelle formule pour  $G(e^t)$  :

$$\begin{aligned} G(e^t) &= 1 + \frac{(\kappa_1 t + \frac{1}{2}\kappa_2 t^2 + \dots)}{1!} + \frac{(\kappa_1 t + \frac{1}{2}\kappa_2 t^2 + \dots)^2}{2!} + \dots \\ &= 1 + \kappa_1 t + \frac{1}{2}(\kappa_2 + \kappa_1^2)t^2 + \dots \end{aligned}$$

Il ne reste plus qu’à mettre en équations les coefficients des puissances de  $t$  pour en déduire une suite de formules

$$\kappa_1 = \mu_1, \quad (8.44)$$

$$\kappa_2 = \mu_2 - \mu_1^2, \quad (8.45)$$

$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3, \quad (8.46)$$

$$\kappa_4 = \mu_4 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 3\mu_2^2 - 6\mu_1^4, \quad (8.47)$$

$$\begin{aligned} \kappa_5 &= \mu_5 - 5\mu_1\mu_4 + 20\mu_1^2\mu_3 - 10\mu_2\mu_3 \\ &\quad + 30\mu_1\mu_2^2 - 60\mu_1^3\mu_2 + 24\mu_1^5, \end{aligned} \quad (8.48)$$

⋮

qui exprime les cumulants en fonction des moments. Notez que  $\kappa_2$  est bien égal à la variance  $E(X^2) - (EX)^2$ .

On peut immédiatement déduire de l’équation (8.41) que les cumulants définis par le produit  $F(z)G(z)$  de deux fgp sont égaux aux sommes des cumulants correspondants de  $F(z)$  et  $G(z)$ , tout simplement parce que le logarithme d’un produit est une somme. Par conséquent, tout cumulant d’une somme de variables aléatoires indépendantes est additif, comme le sont la moyenne et la variable. En raison de cette propriété, les cumulants se trouvent être des paramètres plus importants que les moments.

Changeons maintenant de point de vue et écrivons

$$G(1+t) = 1 + \frac{\alpha_1}{1!}t + \frac{\alpha_2}{2!}t^2 + \frac{\alpha_3}{3!}t^3 + \dots$$

*For these higher half-invariants we shall propose no special names.*

—T. N. Thiele [351]

Les  $\alpha_m$  sont des "moments factoriels" car, selon (8.33),

$$\begin{aligned}
 \alpha_m &= G^{(m)}(1) \\
 &= \sum_{k \geq 0} \Pr(X=k) k^m z^{k-m} \Big|_{z=1} \\
 &= \sum_{k \geq 0} k^m \Pr(X=k) \\
 &= E(X^m).
 \end{aligned} \tag{8.49}$$

Il s'ensuit que

$$\begin{aligned}
 G(e^t) &= 1 + \frac{\alpha_1}{1!}(e^t - 1) + \frac{\alpha_2}{2!}(e^t - 1)^2 + \dots \\
 &= 1 + \frac{\alpha_1}{1!}(t + \frac{1}{2}t^2 + \dots) + \frac{\alpha_2}{2!}(t^2 + t^3 + \dots) + \dots \\
 &= 1 + \alpha_1 t + \frac{1}{2}(\alpha_2 + \alpha_1)t^2 + \dots,
 \end{aligned}$$

et on peut exprimer les cumulants en fonction des dérivées  $G^{(m)}(1)$  :

$$\kappa_1 = \alpha_1, \tag{8.50}$$

$$\kappa_2 = \alpha_2 + \alpha_1 - \alpha_1^2, \tag{8.51}$$

$$\kappa_3 = \alpha_3 + 3\alpha_2 + \alpha_1 - 3\alpha_2\alpha_1 - 3\alpha_1^2 + 2\alpha_1^3, \tag{8.52}$$

⋮

Ces formules constituent une généralisation de (8.38) et (8.39) à tous les cumulants.

Revenons sur terre et appliquons ces notions nouvelles sur des exemples simples. La variable aléatoire la plus simple possible est la "constante aléatoire" :  $X$  a une valeur fixée  $x$  avec probabilité 1. Dans ce cas,  $G_X(z) = z^x$  et  $\ln G_X(e^t) = xt$ , donc la moyenne est égale à  $x$  et tous les autres cumulants sont nuls. Il s'ensuit que l'action de multiplier une fgp par  $z^x$  a pour effet d'augmenter la moyenne de  $x$  en laissant inchangée la variance et tous les autres cumulants.

Voyons maintenant comment appliquer les fonctions génératrices de probabilité à notre jeu de dés. La fgp de la distribution des valeurs d'un dé est

$$G(z) = \frac{z + z^2 + z^3 + z^4 + z^5 + z^6}{6} = zU_6(z),$$

où  $U_6$  est la fgp de la distribution uniforme d'ordre 6. Comme le facteur " $z$ " ajoute 1 à la moyenne, celle-ci est égale à 3,5 au lieu de la valeur  $\frac{n-1}{2} = 2,5$  calculée selon (8.35). Par contre, ce facteur " $z$ " ne modifie pas le calcul de la variance (8.36), donc celle-ci est égale à  $\frac{35}{12}$ .

La fgp de la somme des valeurs de deux dés indépendants est égale au carré de la fgp d'un unique dé :

$$\begin{aligned} G_S(z) &= \frac{z^2 + 2z^3 + 3z^4 + 4z^5 + 5z^6 + 6z^7 + 5z^8 + 4z^9 + 3z^{10} + 2z^{11} + z^{12}}{36} \\ &= z^2 U_6(z)^2. \end{aligned}$$

Si on jette deux dés  $n$  fois de suite, la probabilité d'obtenir une somme totale égale à  $k$  vaut donc

$$\begin{aligned} [z^k] G_S(z)^n &= [z^k] z^{2n} U_6(z)^{2n} \\ &= [z^{k-2n}] U_6(z)^{2n}. \end{aligned}$$

*La distribution de chapeaux est une sorte de distribution d'uniformes.*

Dans le problème de la victoire au football, ou problème d'énumération des points fixes d'une permutation aléatoire, on sait d'après (5.49) que la fgp est

$$F_n(z) = \sum_{0 \leq k \leq n} \frac{(n-k)_i z^k}{(n-k)! k!}, \quad \text{pour } n \geq 0. \quad (8.53)$$

Par conséquent,

$$\begin{aligned} F'_n(z) &= \sum_{1 \leq k \leq n} \frac{(n-k)_i}{(n-k)!} \frac{z^{k-1}}{(k-1)!} \\ &= \sum_{0 \leq k \leq n-1} \frac{(n-1-k)_i}{(n-1-k)!} \frac{z^k}{k!} \\ &= F_{n-1}(z). \end{aligned}$$

Inutile de calculer les coefficients pour conclure de cette récurrence  $F'_n(z) = F_{n-1}(z)$  que  $F_n^{(m)}(z) = F_{n-m}(z)$ , et donc que

$$F_n^{(m)}(1) = F_{n-m}(1) = [n \geq m]. \quad (8.54)$$

Avec cette formule, il nous est facile de calculer la moyenne et la variance. Nous trouvons, comme précédemment (mais plus rapidement), qu'elles sont toutes deux égales à 1 pour tout  $n \geq 2$ .

En fait, nous pouvons même montrer que le  $m$ ième cumulant  $\kappa_m$  de cette variable aléatoire est égal à 1 pour tout  $n \geq m$ . En effet, le  $m$ ième cumulant ne dépend que de  $F'_n(1), F''_n(1), \dots, F_n^{(m)}(1)$ , qui sont tous égaux à 1 ; on obtient le même résultat pour le  $m$ ième cumulant en remplaçant  $F_n(z)$  par la fgp limite

$$F_\infty(z) = e^{z-1}, \quad (8.55)$$

qui satisfait  $F_\infty^{(m)}(1) = 1$  pour tout  $m$ . Les cumulants de  $F_\infty$  sont tous égaux à 1 du fait que

$$\ln F_\infty(e^t) = \ln e^{e^t - 1} = e^t - 1 = \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots$$

## 8.4 PILE OU FACE

Nous allons maintenant nous intéresser à des processus qui n'ont que deux résultats possibles. Si on lance une pièce de monnaie en l'air, on a une probabilité  $p$  d'obtenir pile et une probabilité  $q$  d'obtenir face, avec  $p + q = 1$ . (nous supposons que la pièce ne tombe pas sur la tranche, ni dans un trou, etc). Dans cette section, la somme des nombres  $p$  et  $q$  sera toujours égale à 1. Si la pièce est équilibrée, on a  $p = q = \frac{1}{2}$ ; autrement, nous dirons que la pièce est biaisée.

Voici la fonction génératrice de probabilité pour le nombre de piles après un unique lancer :

$$H(z) = q + pz. \quad (8.56)$$

Si on lance la pièce  $n$  fois de suite, en supposant que deux lancers différents sont indépendants, la fonction génératrice de probabilité du nombre de piles est, selon la formule du binôme,

$$H(z)^n = (q + pz)^n = \sum_{k \geq 0} \binom{n}{k} p^k q^{n-k} z^k. \quad (8.57)$$

Ainsi, la probabilité d'obtenir exactement  $k$  fois pile dans  $n$  lancers est égale à  $\binom{n}{k} p^k q^{n-k}$ . Cette suite de probabilités est appelée la *loi binomiale*.

Supposons maintenant que nous lancions une pièce jusqu'à ce que nous obtenions pile. Quelle est la probabilité que  $k$  lancers exactement soient nécessaires ? Il est évident que  $k = 1$  avec probabilité  $p$ ; et la probabilité que  $k = 2$  vaut  $qp$  (c'est la probabilité de trouver face d'abord, puis pile). De façon générale, la probabilité d'effectuer  $k$  lancers est égale à  $q^{k-1}p$ . La fonction génératrice correspondante est donc

$$pz + qpz^2 + q^2 pz^3 + \dots = \frac{pz}{1 - qz}. \quad (8.58)$$

Si on répète le processus jusqu'à obtenir  $n$  fois pile, on trouve la fgp

$$\begin{aligned} \left(\frac{pz}{1 - qz}\right)^n &= p^n z^n \sum_k \binom{n+k-1}{k} (qz)^k \\ &= \sum_k \binom{k-1}{k-n} p^n q^{k-n} z^k. \end{aligned} \quad (8.59)$$

*Les escrocs patentés savent bien que  $p \approx 0,9$  si on joue à pile ou face avec un penny américain tout neuf sur une surface lisse : à cause de la position du centre de gravité, la tête d'Abraham Lincoln se trouve presque toujours dessous.*

Remarquons en passant qu'elle est égale à  $z^n$  multiplié par

$$\left(\frac{p}{1-qz}\right)^n = \sum_k \binom{n+k-1}{k} p^n q^k z^k, \quad (8.60)$$

la fonction génératrice de la *loi binomiale négative*.

L'espace de probabilité de l'exemple (8.59) est tout à fait différent de ceux que nous avons rencontrés jusqu'à présent dans ce chapitre : il est infini, alors que les précédents étaient finis. Chacun de ses éléments est une suite de finie de piles et/ou de faces, contenant exactement  $n$  piles, et finissant par pile. La probabilité d'une telle suite est égale à  $p^n q^{k-n}$ , où  $k - n$  est le nombre de faces. Ainsi, par exemple, si  $n = 3$  et si on note P pour pile et F pour face, la suite FPFFFPP appartient à l'espace de probabilité, et sa probabilité est égale à  $qpqqqpp = p^3 q^4$ .

Pile je gagne,  
face tu perds.

Non ? D'accord ;  
face tu perds, pile je  
gagne.

Non ? Bon, alors  
pile tu perds,  
face je gagne.

Soient  $X$  une variable aléatoire de loi binomiale (8.57) et  $Y$  une variable aléatoire de loi binomiale négative (8.60). Ces distributions dépendent de  $n$  et de  $p$ . La moyenne de  $X$  est  $nH'(1) = np$ , car sa fgp est  $H(z)^n$  ; sa variance est

$$n(H''(1) + H'(1) - H'(1)^2) = n(0 + p - p^2) = npq. \quad (8.61)$$

L'écart-type est donc égal à  $\sqrt{npq}$  : si on lance  $n$  fois une pièce de monnaie, on doit donc obtenir pile à peu près  $np \pm \sqrt{npq}$  fois. La moyenne et la variance de  $Y$  se calculent de manière similaire : si on pose

$$G(z) = \frac{p}{1-qz},$$

alors on a

$$G'(z) = \frac{pq}{(1-qz)^2},$$

$$G''(z) = \frac{2pq^2}{(1-qz)^3};$$

par conséquent,  $G'(1) = pq/p^2 = q/p$  et  $G''(1) = 2pq^2/p^3 = 2q^2/p^2$ . Il s'ensuit que la moyenne et la variance de  $Y$  sont respectivement égales à  $nq/p$  et  $nq/p^2$ .

Il y a une façon plus simple de trouver ce résultat, en utilisant la fonction génératrice inverse

$$F(z) = \frac{1-qz}{p} = \frac{1}{p} - \frac{q}{p}z, \quad (8.62)$$

et en écrivant

$$G(z)^n = F(z)^{-n}. \quad (8.63)$$

Ce polynôme  $F(z)$  n'est pas une fonction génératrice de probabilité car il a des coefficients négatifs. Cependant, il satisfait la condition fondamentale  $F(1) = 1$ . Formellement,  $F(z)$  correspond donc à une pièce avec laquelle la "probabilité" d'obtenir pile est égale à  $-q/p$ , et  $G(z)$  équivaut à lancer cette pièce  $-1$  fois(!). Ainsi, la loi binomiale négative de paramètres  $(n, p)$  peut être considérée comme une loi binomiale ordinaire de paramètres  $(n', p') = (-n, -q/p)$ . Formellement, la moyenne doit donc être  $n'p' = (-n)(-q/p) = nq/p$ , et la variance  $n'p'q' = (-n)(-q/p)(1 + q/p) = nq/p^2$ . Ce calcul strictement formel, où apparaissent des probabilités négatives, est tout à fait correct. En effet, les résultats que nous y appliquons sont basés sur des identités de séries formelles pour lesquelles il n'est pas nécessaire que  $0 \leq p \leq 1$ .

*La probabilité que je rajeunisse est négative.*

*Ah ? Alors la probabilité que tu vieillisses ou que tu ne changes pas est > 1.*

Passons à un autre exemple : Combien de fois faut-il lancer une pièce pour obtenir pile deux fois de suite ? Maintenant, l'espace de probabilité est constitué de toutes les suites de P et de F qui finissent par PP mais qui ne contiennent pas d'autre occurrence de deux P consécutifs :

$$\Omega = \{PP, FPP, FFPP, PFPP, FFFFPP, FPFPP, PFFPP, \dots\}.$$

La probabilité d'une séquence donnée s'obtient en remplaçant P par  $p$  et F par  $q$ . Par exemple, la suite FPFPP apparaît avec la probabilité

$$\Pr(FPFPP) = qpqpp = p^3q^2.$$

Soit  $S$  la somme infinie

$$S = PP + FPP + FFPP + PFPP + FFFFPP + FPFPP + PFFPP + \dots$$

de tous les éléments de  $\Omega$ . Si on remplace chaque P par  $pz$  et chaque F par  $qz$ , on obtient la fonction génératrice de probabilité du nombre de lancers nécessaires pour tomber deux fois de suite sur pile.

Il existe une curieuse relation entre  $S$  et la somme des pavages par des dominos

$$T = 1 + \square + \square\square + \square + \square\square + \square\square + \square\square + \dots$$

de l'équation (7.1) :  $S$  peut être obtenue à partir de  $T$  en remplaçant chaque  $\square$  par F et chaque  $\square\square$  par PF, puis en ajoutant PP à la fin. Il est facile de démontrer cette correspondance, car tout élément de  $\Omega$  est de la forme  $(F + PF)^n PP$ , pour un certain  $n \geq 0$ , et tout terme de  $T$  est de la forme  $(\square + \square\square)^n$ . Par conséquent, d'après (7.4),

$$S = (1 - F - PF)^{-1} PP,$$

et la fonction génératrice de probabilité que nous cherchons est donc

$$\begin{aligned} G(z) &= (1 - qz - (pz)(qz))^{-1} (pz)^2 \\ &= \frac{p^2 z^2}{1 - qz - pqz^2}. \end{aligned} \quad (8.64)$$

Ce que nous avons fait récemment avec la loi binomiale négative nous incite à calculer la moyenne et la variance de (8.64) en écrivant

$$G(z) = \frac{z^2}{F(z)},$$

où

$$F(z) = \frac{1 - qz - pqz^2}{p^2},$$

et en calculant la “moyenne” et la “variance” de cette pseudo-fgp  $F(z)$  (qui satisfait  $F(1) = 1$ ). On a

$$F'(1) = (-q - 2pq)/p^2 = 2 - p^{-1} - p^{-2};$$

$$F''(1) = -2pq/p^2 = 2 - 2p^{-1}.$$

Par conséquent, comme  $z^2 = F(z)G(z)$ ,  $\text{Moy}(z^2) = 2$ , et  $\text{Var}(z^2) = 0$ , voici la moyenne et la variance de la distribution  $G(z)$  :

$$\text{Moy}(G) = 2 - \text{Moy}(F) = p^{-2} + p^{-1}; \quad (8.65)$$

$$\text{Var}(G) = -\text{Var}(F) = p^{-4} + 2p^{-3} - 2p^{-2} - p^{-1}. \quad (8.66)$$

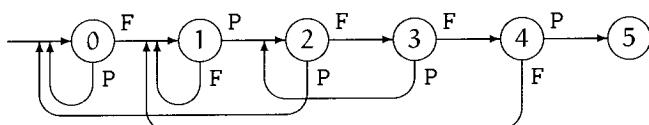
Lorsque  $p = \frac{1}{2}$ , la moyenne et la variance valent 6 et 22 respectivement. (L'exercice 4 est consacré aux calculs de moyenne et de variance par sous-traction).

Tentons maintenant une expérience plus compliquée : nous allons lancer une pièce en l'air jusqu'à obtenir la configuration FPFFP. La somme des suites gagnantes est

$$\begin{aligned} S &= \text{FPFFP} + \text{PFPFFP} + \text{FFPFFP} \\ &\quad + \text{PPFPFFP} + \text{PFFPFFP} + \text{FPFPFFP} + \text{FFFPPFP} + \dots; \end{aligned}$$

““Vous êtes un automate, une machine à calculer”, m'écriai-je. “Il y a parfois en vous quelque chose de positivement inhumain”.”  
— J. H. Watson [83]

elle est plus difficile à décrire que la précédente. Pour résoudre le problème, nous pouvons considérer  $S$  comme un “langage” défini par l’“automate fini” suivant :



Les événements élémentaires de l'espace de probabilité sont exactement les suites de P et de F qui mènent de l'état 0 à l'état 5. Supposons par exemple que nous venions de faire FPF ; alors nous nous trouvons dans l'état 3. A ce moment, si nous tirons face, nous allons dans l'état 4 ; et si nous tirons pile, toujours depuis l'état 3, nous nous retrouvons en 2 (et non en 0, car il suffit que le FP que nous venons de faire soit suivi par FFP pour que nous gagnions).

Soit  $S_k$  la somme de toutes les suites de P et F qui conduisent de l'état 0 à l'état k. Alors

$$S_0 = 1 + S_0 P + S_2 P,$$

$$S_1 = S_0 F + S_1 F + S_4 F,$$

$$S_2 = S_1 P + S_3 P,$$

$$S_3 = S_2 F,$$

$$S_4 = S_3 F,$$

$$S_5 = S_4 P.$$

Notre somme S est égale à  $S_5$ . Pour l'obtenir, il suffit de résoudre ce système de six équations à six inconnues  $S_0, S_1, \dots, S_5$ . Si on remplace P par  $p_z$  et F par  $q_z$  dans l'un des  $F_k$  ainsi trouvés, on trouve une fonction génératrice dont le coefficient de  $z^n$  représente la probabilité d'être dans l'état k après n tirages.

Ce procédé peut s'appliquer à tout diagramme de transitions entre états, où la transition entre l'état j et l'état k est associée à une probabilité  $p_{j,k}$ . On obtient ainsi un système d'équations linéaires dont les solutions sont les fonctions génératrices des probabilités des états après n transitions. Les systèmes de ce type sont appelés des *processus de Markov* et l'étude de leur comportement est très liée à la théorie des équations linéaires.

Il existe aussi une manière bien plus simple de résoudre notre problème de pile ou face, qui évite le passage par l'automate fini. Les six équations à six inconnues  $S_0, S_1, \dots, S_5$ , peuvent être avantageusement remplacées par seulement deux équations à deux inconnues. Voici l'astuce : considérons la somme  $N = S_0 + S_1 + S_2 + S_3 + S_4$  de toutes les suites qui ne contiennent aucune occurrence de FPFFP,

$$N = 1 + P + F + PP + \dots + FPFFP + FPFFF + \dots$$

Alors

$$1 + N(P + F) = N + S, \quad (8.67)$$

car le membre gauche contient les termes qui finissent par FPFFP (et donc appartiennent à S) et les autres (qui appartiennent à N) ; inversement, tout

terme de droite non vide appartient soit à  $N_P$  soit à  $N_F$ . De plus, l'équation

$$N_{FPFFP} = S + S_{FFP} \quad (8.68)$$

est vérifiée, car, si un terme du membre gauche n'est pas élément de  $S$ , alors il le devient si on supprime ses trois dernières lettres ; et tout terme du membre droit se retrouve dans le membre gauche.

Il est facile de résoudre ce système de deux équations : d'après (8.67),  $N = (1 - S)(1 - P - F)^{-1}$ , donc

$$(1 - S)(1 - F - P)^{-1} FPFFP = S(1 + FFP).$$

Il suffit, comme précédemment, de remplacer  $P$  par  $pz$  et  $F$  par  $qz$  pour obtenir la fonction génératrice de probabilité  $G(z)$  du nombre de lancers. Après quelques simplifications (parce que  $p + q = 1$ ), on trouve

$$\frac{(1 - G(z)) p^2 q^3 z^5}{1 - z} = G(z)(1 + pq^2 z^3);$$

par conséquent, la solution est

$$G(z) = \frac{p^2 q^3 z^5}{p^2 q^3 z^5 + (1 + pq^2 z^3)(1 - z)}. \quad (8.69)$$

Remarquez que  $G(1) = 1$  si  $pq \neq 0$ . Nous sommes donc sûrs d'obtenir la configuration FPFFP avec probabilité 1, sauf si la pièce est truquée au point de toujours tomber du même côté.

Pour trouver la moyenne et la variance de la distribution (8.69), inversons  $G(z)$  comme nous l'avons fait dans le problème précédent en écrivant  $G(z) = z^5/F(z)$ , où  $F$  est un polynôme :

$$F(z) = \frac{p^2 q^3 z^5 + (1 + pq^2 z^3)(1 - z)}{p^2 q^3}. \quad (8.70)$$

Nous en déduisons les dérivées

$$\begin{aligned} F'(1) &= 5 - (1 + pq^2)/p^2 q^3, \\ F''(1) &= 20 - 6pq^2/p^2 q^3, \end{aligned}$$

et, si  $X$  désigne le nombre de lancers,

$$EX = \text{Moy}(G) = 5 - \text{Moy}(F) = p^{-2} q^{-3} + p^{-1} q^{-1}; \quad (8.71)$$

$$\begin{aligned} VX = \text{Var}(G) &= -\text{Var}(F) \\ &= -25 + p^{-2} q^{-3} + 7p^{-1} q^{-1} + \text{Moy}(F)^2 \\ &= (EX)^2 - 9p^{-2} q^{-3} - 3p^{-1} q^{-1}. \end{aligned} \quad (8.72)$$

Pour  $p = \frac{1}{2}$ , la moyenne et la variance valent respectivement 36 et 996.

Nous allons maintenant voir les choses de façon bien plus générale. Le problème que nous venons de résoudre nous donne de bonnes pistes pour nous attaquer au cas où nous lançons la pièce jusqu'à obtenir une configuration donnée *quelconque* A de P et de F. Soient donc S la somme de toutes les suites gagnantes et N la somme des suites qui n'ont pas encore abouti au motif A recherché. Alors l'équation (8.67) ne change pas, tandis que l'équation (8.68) devient

$$NA = S(1 + A^{(1)} [A^{(m-1)} = A_{(m-1)}] + A^{(2)} [A^{(m-2)} = A_{(m-2)}] \\ + \cdots + A^{(m-1)} [A^{(1)} = A_{(1)}]), \quad (8.73)$$

où m est la longueur de A, et  $A^{(k)}$  et  $A_{(k)}$  désignent respectivement les k dernières lettres et les k premières lettres de A. Par exemple, si A est le motif FPFFP que nous venons d'étudier, alors

$$A^{(1)} = P, \quad A^{(2)} = FP, \quad A^{(3)} = FFP, \quad A^{(4)} = PFFP; \\ A_{(1)} = F, \quad A_{(2)} = FP, \quad A_{(3)} = FPF, \quad A_{(4)} = FPFF.$$

Comme la seule égalité vérifiée est  $A^{(2)} = A_{(2)}$ , l'équation (8.73) se réduit à (8.68).

Soit  $\tilde{A}$  le résultat obtenu en remplaçant P par  $p^{-1}$  et F par  $q^{-1}$  dans A. Alors il n'est pas difficile de généraliser nos calculs de (8.71) et (8.72) pour conclure (voir l'exercice 20) que la moyenne vaut

$$EX = \sum_{k=1}^m \tilde{A}_{(k)} [A^{(k)} = A_{(k)}]; \quad (8.74)$$

et que la variance est donnée par

$$VX = (EX)^2 - \sum_{k=1}^m (2k-1) \tilde{A}_{(k)} [A^{(k)} = A_{(k)}]. \quad (8.75)$$

Si  $p = \frac{1}{2}$ , il existe une interprétation particulièrement simple de ces formules. Etant donné un motif A de longueur m, soit

$$A:A = \sum_{k=1}^m 2^{k-1} [A^{(k)} = A_{(k)}]. \quad (8.76)$$

Nous allons voir maintenant comment trouver la représentation en base deux de ce nombre : commencez par superposer deux occurrences de A, puis déplacez la seconde occurrence vers la droite, lettre par lettre ; écrivez “1” sous la première lettre de cette seconde occurrence chaque fois que les

lettres des deux occurrences coïncident exactement, écrivez "0" dans le cas contraire. La figure suivante présente une illustration de ce procédé :

D'après (8.74), le nombre moyen de lancers nécessaires pour que le motif A apparaisse est exactement égal à  $2(A:A)$ , car  $\tilde{A}_{(k)} = 2^k$  lorsque  $p = q = \frac{1}{2}$ . Ce résultat, découvert par le mathématicien soviétique A. D. Solov'ev en 1966 [331], peut sembler paradoxal à première vue : les motifs qui ne coïncident pas lorsqu'on les superpose surviennent plus tôt que les autres ! Il faut presque deux fois plus de temps pour rencontrer PPPPP que pour rencontrer PPPPF ou FPPPP.

Voyons maintenant un jeu inventé par Walter Penney [289] en 1969. Alice et Bill jouent à pile ou face jusqu'à ce qu'ils obtiennent PPF ou PFF. Si c'est PPF, c'est Alice qui gagne, sinon c'est Bill. Ce jeu semble bien être équitable si la pièce n'est pas truquée, car les deux motifs PPF et PFF présentent les mêmes caractéristiques si on le considère séparément : la fonction génératrice de probabilité du nombre de lancers nécessaires pour obtenir PPF est

$$G(z) = \frac{z^3}{z^3 - 8(z-1)},$$

*Bien sûr que non !  
Sur qui auraient-ils  
l'avantage ?*

et on trouve exactement la même pour PFF. Par conséquent, ni Alice ni Bill ne prend l'avantage s'ils jouent en solitaire.

En revanche, lorsqu'on considère les deux motifs simultanément, leur interaction donne quelque chose d'intéressant. Soient A et B les sommes des configurations gagnantes d'Alice et de Bill respectivement :

$$S_A = PPF + PPPF + FPPF + PPPPF + PFPPF + FPPPFF + \dots ;$$

$$S_B = PFF + FPFF + PFPFF + FFPFF + FPFPFF + FFFFPPF + \dots .$$

## 434 PROBABILITÉS DISCRÈTES

Soit aussi  $N$  la somme de toutes les suites pour lesquelles aucun des deux joueurs n'a encore gagné :

$$N = 1 + P + F + PP + PF + FP + FF + PPP + PFP + FPP + \dots \quad (8.77)$$

On peut alors facilement vérifier les équations suivantes :

$$\begin{aligned} 1 + N(P + F) &= N + S_A + S_B ; \\ N PPF &= S_A ; \\ N PFF &= S_A F + S_B . \end{aligned} \quad (8.78)$$

Si on pose  $P = F = \frac{1}{2}$ , la valeur de  $S_A$  est égale à la probabilité qu'Alice gagne, et la valeur de  $S_B$  est égale à la probabilité que Bill gagne. Les trois équations se réduisent à

$$1 + N = N + S_A + S_B ; \quad \frac{1}{8}N = S_A ; \quad \frac{1}{8}N = \frac{1}{2}S_A + S_B .$$

Nous en déduisons que  $S_A = \frac{2}{3}$  et  $S_B = \frac{1}{3}$ . Alice gagnera en moyenne deux fois plus souvent que Bill !

On peut généraliser ce jeu en permettant à Alice et Bill de choisir leurs motifs  $A$  et  $B$  au lieu de les leur imposer. Les deux motifs ne sont pas nécessairement de même longueur, mais il ne faut pas que  $A$  soit contenu dans  $B$  ou inversement, sinon le jeu dégénérerait (par exemple, si  $A = PF$  et  $B = FPFP$ , le pauvre Bill ne peut jamais gagner, et si  $A = PFP$  et  $B = FP$ , les deux joueurs peuvent prétendre à la victoire en même temps). On peut alors écrire trois équations analogues à (8.73) et (8.78) :

$$\begin{aligned} 1 + N(P + F) &= N + S_A + S_B ; \\ NA &= S_A \sum_{k=1}^l A^{(1-k)} [A^{(k)} = A_{(k)}] + S_B \sum_{k=1}^{\min(l, m)} A^{(1-k)} [B^{(k)} = A_{(k)}] ; \\ NB &= S_A \sum_{k=1}^{\min(l, m)} B^{(m-k)} [A^{(k)} = B_{(k)}] + S_B \sum_{k=1}^m B^{(m-k)} [B^{(k)} = B_{(k)}] . \end{aligned} \quad (8.79)$$

Ici,  $l$  et  $m$  désignent respectivement les longueurs de  $A$  et  $B$ . Par exemple, si  $A = PFFPFPFP$  et  $B = FPFPFFP$ , les deux dernières équations se réduisent à

$$\begin{aligned} N PFFPFPFP &= S_A FFPFPFP + S_A + S_B FFPFPFP + S_B FPFP ; \\ N FPFPFFP &= S_A FPFPFP + S_A FFP + S_B FPFPFP + S_B . \end{aligned}$$

Si on suppose que la pièce n'est pas truquée, on obtient les probabilités de

victoire en posant  $P = F = \frac{1}{2}$ . Nos deux équations deviennent alors

$$\begin{aligned} N &= S_A \sum_{k=1}^l 2^k [A^{(k)} = A_{(k)}] + S_B \sum_{k=1}^{\min(l,m)} 2^k [B^{(k)} = A_{(k)}]; \\ N &= S_A \sum_{k=1}^{\min(l,m)} 2^k [A^{(k)} = B_{(k)}] + S_B \sum_{k=1}^m 2^k [B^{(k)} = B_{(k)}]. \end{aligned} \quad (8.80)$$

Nous verrons mieux ce qui se passe si nous généralisons l'opération  $A:A$  de (8.76) à une fonction de deux motifs indépendants  $A$  et  $B$  :

$$A:B = \sum_{k=1}^{\min(l,m)} 2^{k-1} [A^{(k)} = B_{(k)}]. \quad (8.81)$$

Les équations (8.80) se réécrivent alors simplement en

$$S_A(A:A) + S_B(B:A) = S_A(A:B) + S_B(B:B),$$

et l'avantage d'Alice vaut

$$\frac{S_A}{S_B} = \frac{B:B - B:A}{A:A - A:B}. \quad (8.82)$$

Cette belle formule est due à John Horton Conway [137].

Par exemple, si  $A = PFFPFPFP$  et  $B = FPFPFFP$  comme précédemment, alors  $A:A = (10000001)_2 = 129$ ,  $A:B = (0001010)_2 = 10$ ,  $B:A = (0001001)_2 = 9$  et  $B:B = (1000010)_2 = 66$ . Donc, le rapport  $S_A/S_B$  vaut  $(66 - 9)/(129 - 10) = 57/119$ . Alice ne gagnera ce jeu que 57 fois sur 176 en moyenne.

Il peut se passer des choses étranges dans le jeu de Penney. Par exemple, le motif PPFP gagne contre le motif PFPP à 3 contre 2, et PFPP gagne contre FPPP à 7 contre 5. Par conséquent, PPFP devrait être bien meilleur que FPPP. Eh bien non, car FPPP gagne contre PPFP, à 7 contre 5 ! La relation entre les motifs n'est pas transitive. L'exercice 57 montre que si Alice choisit un quelconque motif  $\tau_1\tau_2\dots\tau_l$  de longueur  $l \geq 3$ , Bill est sûr d'avoir plus de chances de gagner s'il choisit le motif  $\bar{\tau}_2\tau_1\tau_2\dots\tau_{l-1}$ , où  $\bar{\tau}_2$  est égal à  $F$  si  $\tau_2$  est égal à  $P$  et inversement.

## 8.5 HACHAGE

Pour conclure ce chapitre, nous allons appliquer la théorie des probabilités à l'informatique. Un bon nombre d'algorithmes de stockage et de recherche d'informations dans un ordinateur sont basés sur une technique que l'on appelle le "hachage". Voici le problème qu'elle permet de

*Vous avez dit bizarre ? Comme c'est bizarre !*

résoudre : il s'agit de gérer un ensemble d'informations, appelées enregistrements ; chaque enregistrement contient une valeur  $K$  appelée la "clé" et des données  $D(K)$  correspondant à cette clé ; on veut pouvoir trouver rapidement  $D(K)$  lorsque  $K$  est donné. Par exemple, les clés peuvent être des noms d'étudiants, et les données associées les notes de leurs devoirs.

En pratique, les ordinateurs ne disposent pas d'assez de mémoire pour résERVER un emplacement pour chaque clé possible. Il peut y avoir des milliards de clés possibles, dont quelques-une seulement sont utilisées dans une application donnée. Une manière de résoudre ce problème consiste à gérer deux tableaux CLE et DONNEE indicés de 1 à  $N$ , où  $N$  est le nombre maximum d'enregistrements que l'on peut mémoriser ; une autre variable,  $n$ , compte le nombre d'enregistrements effectivement présents en mémoire. La recherche d'une clé donnée  $K$  peut alors s'effectuer, de façon évidente, en parcourant la table séquentiellement :

- S1** Poser  $j := 1$  (nous avons parcouru toutes les cases de position  $< j$ ).
- S2** Si  $j > n$ , arrêter (nous n'avons pas trouvé).
- S3** Si  $CLE[j] = K$ , arrêter (nous avons trouvé).
- S4** Incrémenter  $j$  de 1 et revenir en S2 (nous essayons à nouveau).

Si l'information cherchée a été trouvée, les données  $D(K)$  sont contenues dans la case  $DONNEE[j]$  du tableau DONNEE. Sinon, on peut ajouter  $K$  et  $D(K)$  dans la table en exécutant

$$n := j, \quad CLE[n] := K, \quad DONNEE[n] := D(K),$$

pourvu que le tableau ne soit pas déjà entièrement plein.

Cette méthode, qui marche dans tous les cas, peut cependant être affreusement lente : chaque fois qu'on cherche une clé qui n'existe pas dans le tableau, l'instruction S2 est effectuée  $n+1$  fois ; et  $n$  peut être très grand.

C'est pour accélérer ce traitement qu'on a inventé le hachage. En gros, l'idée consiste à utiliser  $m$  listes séparées au lieu d'une seule très grande. Pour cela, on définit une "fonction de hachage" qui associe à chaque clé  $K$  possible un numéro de liste  $h(K)$  entre 1 et  $m$ . On utilise deux tableau auxiliaires : le tableau PREMIER, indicé de 1 à  $m$ , dont chaque case  $PREMIER[i]$  pointe sur le premier enregistrement de la liste  $i$  ; et le tableau SUIVANT, indicé de 1 à  $N$ , dont chaque case  $SUIVANT[j]$  pointe sur l'enregistrement qui suit l'enregistrement numéro  $j$  dans sa liste. On convient que

$$\begin{aligned} PREMIER[i] &= -1 && \text{si la liste } i \text{ est vide;} \\ SUIVANT[j] &= 0 && \text{si l'enregistrement } j \text{ est le dernier de sa liste.} \end{aligned}$$

Il y a aussi, comme pour la première méthode, une variable  $n$  qui compte le nombre d'enregistrements en mémoire.

"Somehow the verb 'to hash' magically became standard terminology for  $k$  transformation during the mid-1960 yet nobody was rich enough to use such an undignified word publicly until 1966  
— D. E. Knuth [2]

Supposons par exemple que les clés sont des noms, que  $m = 4$  et que l'appartenance d'une donnée à l'une des quatre listes est déterminée par la première lettre de sa clé :

$$h(\text{nom}) = \begin{cases} 1, & \text{de A à F ;} \\ 2, & \text{de G à L ;} \\ 3, & \text{de M à R ;} \\ 4, & \text{de S à Z.} \end{cases}$$

On commence avec des listes vides et avec  $n = 0$ . Supposons que la clé du premier enregistrement à stocker soit Nora. Alors  $h(\text{Nora}) = 3$ , donc la clé Nora et les données correspondantes sont stockées dans la liste 3. Si les deux noms qui suivent sont Glenn et Jim, on les met tous deux dans la liste 2. Voici alors l'état de nos tableaux dans la mémoire de l'ordinateur :

```
PREMIER[1] = -1, PREMIER[2] = 2, PREMIER[3] = 1, PREMIER[4] = -1.  
CLE[1] = Nora,      SUIVANT[1] = 0;  
CLE[2] = Glenn,     SUIVANT[2] = 3;  
CLE[3] = Jim,        SUIVANT[3] = 0;      n = 3.
```

*Elles concernent les trois étudiants du premier rang du cours de Mathématiques Concrètes qui ont prêté leurs noms pour la circonstance.*

(Nous n'indiquons pas les valeurs des tableaux DONNEE[1], DONNEE[2] et DONNEE[3], qui sont confidentielles). Après insertion de 18 enregistrements, les listes peuvent par exemple contenir les noms

liste 1	liste 2	liste 3	liste 4
Dianne	Glenn	Nora	Scott
Ari	Jim	Mike	Tina
Brian	Jennifer	Michael	
Fran	Joan	Ray	
Doug	Jerry	Paula	
	Jean		

qui apparaissent probablement en désordre dans la tableau CLE, mais avec les valeurs de SUIVANT adéquates pour que les listes soient correctement séparées. Si on cherche John, il faut parcourir les six noms de la liste 2 (qui se trouve être la plus longue). C'est quand même plus rapide qu'examiner la totalité de clés, soit 18 noms.

Voici l'algorithme de recherche de la clé K :

- H1** Poser  $i := h(K)$  et  $j := \text{PREMIER}[i]$ .
- H2** Si  $j \leq 0$ , arrêter (nous n'avons pas trouvé).
- H3** Si  $\text{CLE}[j] = K$ , arrêter (nous avons trouvé).
- H4** Poser  $i := j$ , puis  $j := \text{SUIVANT}[i]$  et revenir en H2 (nous essayons à nouveau).

Par exemple, voici ce qui se passe si on cherche Jennifer dans l'exemple donné. A l'étape H1, on pose  $i := 2$  et  $j := 2$  ; à l'étape H3, on trouve que

Glenn  $\neq$  Jennifer ; à l'étape H4, on pose  $j := 3$  ; à l'étape H3, on trouve que Jim  $\neq$  Jennifer. Après une autre exécution des étapes H4 et H3, on trouve finalement Jennifer.

*Je parie que leurs parents en sont heureux.*

Comme dans le premier algorithme que nous avons vu, les données correspondant à la clé cherchée, une fois celle-ci trouvée, peuvent être lues dans DONNEE[j]. Si la recherche a été infructueuse, on peut ajouter K et D(K) dans la table en effectuant les opérations suivantes :

```
n := n + 1;
si j < 0 alors PREMIER[i] := n sinon SUIVANT[i] := n;
CLE[n] := K;
DONNEE[n] := D(K);
SUIVANT[n] := 0.                                (8.83)
```

L'idéal serait d'obtenir des listes de longueurs à peu près égales, car dans ce cas la recherche serait environ  $m$  fois plus rapide qu'avec le premier algorithme. Comme la valeur de  $m$  est généralement bien plus grande que 4, un facteur de  $1/m$  présente une amélioration qui est loin d'être négligeable.

On ne peut pas savoir à l'avance quelles clés devront être stockées dans la table. En revanche, il est possible, dans la plupart des cas, de choisir la fonction de hachage  $h$  de sorte que  $h(K)$  puisse être considérée comme une variable aléatoire de distribution uniforme entre 1 et  $m$ . Dans ce genre de cas, le calcul de la fonction de hachage pour une clé donnée est équivalent au jet d'un dé à  $m$  faces. Il n'est bien sûr pas impossible que toutes les clés se retrouvent dans la même liste, comme il est possible qu'un dé tombe toujours sur  ; la théorie des probabilités nous assure cependant que les listes seront *presque toujours* équilibrées.

### Analyse du hachage : introduction

L'“analyse des algorithmes” est une branche de l'informatique qui étudie quantitativement l'efficacité des méthodes informatiques. Effectuer une “analyse probabiliste” d'un algorithme consiste à étudier sa “complexité en temps”, c'est-à-dire son temps d'exécution, en considérant ce temps comme une variable aléatoire qui dépend des caractéristiques des données fournies en entrée. Ce type d'analyse est particulièrement bien adapté à la méthode de hachage, car celle-ci est extrêmement efficace en moyenne, bien que son comportement dans le pire des cas soit catastrophique (lorsque tous les clés ont la même valeur par la fonction de hachage). C'est pourquoi les informaticiens qui utilisent le hachage ont tout intérêt à croire en la théorie des probabilités.

Soit  $P$  le nombre de fois que le test de l'étape H3 est effectué au cours de la recherche d'une clé avec l'algorithme que nous avons décrit. Si on connaît  $P$ , on connaît exactement le nombre d'exécutions de chaque étape,

selon que la clé a été finalement trouvée ou non :

Etape	Clé non trouvée	Clé trouvée
H1	1 fois	1 fois
H2	$P + 1$ fois	$P$ fois
H3	$P$ fois	$P$ fois
H4	$P$ fois	$P - 1$ fois

C'est donc bien ce nombre de tests  $P$  qui détermine le temps d'exécution de l'algorithme de recherche.

Pour bien nous représenter le problème, imaginons un carnet d'adresses organisé de façon très particulière. On ne peut écrire qu'une adresse par page. On indique sur la couverture du carnet le numéro de la page de la première adresse de chacune des  $m$  listes, et on peut déterminer, à partir de chaque nom  $K$ , le numéro de la liste à laquelle il appartient. De plus, on indique en bas de chaque page le numéro de page de l'adresse suivante de la même liste. Le nombre de tests  $P$  correspond, dans ce cas, au nombre de pages à consulter pour trouver l'adresse correspondant à un nom.

Si on insère un à un  $n$  enregistrements, leurs positions dans la table ne dépendent que des valeurs de la fonction de hachage  $\langle h_1, h_2, \dots, h_n \rangle$  qui leurs correspondent. Nous considérerons que les  $m^n$  suites possibles  $\langle h_1, h_2, \dots, h_n \rangle$  sont équiprobables, et que  $P$  est une variable aléatoire dont la valeur dépend de la suite  $\langle h_1, h_2, \dots, h_n \rangle$ .

Regarde sous le paillasson.

### Cas 1 : la clé n'est pas dans la table.

Commençons par observer le comportement de  $P$  dans le cas d'une recherche infructueuse, en supposant que la table de hachage contient  $n$  enregistrements. Dans ce cas, l'espace de probabilité est constitué de  $m^{n+1}$  événements élémentaires  $\omega = (h_1, h_2, \dots, h_n, h_{n+1})$ , où  $h_j$  désigne la valeur de la fonction de hachage pour la  $j$ ème clé insérée, et  $h_{n+1}$  désigne sa valeur pour la clé qu'on cherche et qui ne se trouve pas dans la table. Nous supposerons que la fonction  $h$  a été convenablement choisie, de sorte que  $\Pr(\omega) = 1/m^{n+1}$  pour tout  $\omega$ .

Par exemple, si  $m = n = 2$ , il y a huit possibilités équiprobables :

$h_1$	$h_2$	$h_3$ :	$P$
1	1	1 :	2
1	1	2 :	0
1	2	1 :	1
1	2	2 :	1
2	1	1 :	1
2	1	2 :	1
2	2	1 :	0
2	2	2 :	2

Si  $h_1 = h_2 = h_3$ , on effectue deux tests avant de conclure que la clé K n'existe pas dans la table ; si  $h_1 = h_2 \neq h_3$ , aucun test n'est nécessaire ; et ainsi de suite. On peut déduire de cette liste de possibilités que la distribution de probabilité de  $P$  est donnée par la fgp  $(\frac{2}{8} + \frac{4}{8}z + \frac{2}{8}z^2) = (\frac{1}{2} + \frac{1}{2}z)^2$  lorsque  $m = n = 2$ .

Voyons maintenant le problème général. Comme toute recherche infructueuse donne lieu à un test par enregistrement dans la liste numéro  $h_{n+1}$ , on a

$$P = [h_1 = h_{n+1}] + [h_2 = h_{n+1}] + \cdots + [h_n = h_{n+1}]. \quad (8.84)$$

La probabilité que  $h_j = h_{n+1}$  est égale à  $1/m$ , pour  $1 \leq j \leq n$ . Il s'ensuit que

$$EP = E[h_1 = h_{n+1}] + E[h_2 = h_{n+1}] + \cdots + E[h_n = h_{n+1}] = \frac{n}{m}.$$

Cela mérite peut-être d'être un peu plus détaillé : soit  $X_j$  la variable aléatoire

$$X_j = X_j(\omega) = [h_j = h_{n+1}].$$

Alors  $P = X_1 + \cdots + X_n$ , et  $EX_j = 1/m$  pour tout  $j \leq n$ . Par conséquent,

$$EP = EX_1 + \cdots + EX_n = n/m.$$

Parfait. Le nombre moyen de tests est exactement comme nous l'espérions :  $m$  fois plus petit que sans le hachage. De plus, les variables aléatoires  $X_j$  sont indépendantes et elles ont toutes la même fonction génératrice de probabilité

$$X_j(z) = \frac{m-1+z}{m}.$$

La fgp du nombre total de tests lors d'une recherche infructueuse vaut donc

$$P(z) = X_1(z) \dots X_n(z) = \left( \frac{m-1+z}{m} \right)^n. \quad (8.85)$$

C'est une loi binomiale, avec  $p = 1/m$  et  $q = (m-1)/m$ . Ainsi, le nombre de tests lors d'une recherche infructueuse se comporte exactement comme le nombre de piles lorsqu'on lance une pièce truquée de façon que la probabilité de tirer pile soit égale à  $1/m$  à chaque lancer. D'après l'équation (8.61), la variance de  $P$  est donc égale à

$$npq = \frac{n(m-1)}{m^2}.$$

Lorsque  $m$  est grand, la variance de  $P$  vaut à peu près  $n/m$ , donc son écart-type est proche de  $\sqrt{n/m}$ .

**Cas 2 : la clé est dans la table.**

Examinons maintenant le cas où la recherche se termine bien. L'espace de probabilité correspondant est un peu plus compliqué que précédemment :  $\Omega$  est l'ensemble de tous les événements élémentaires

$$\omega = (h_1, \dots, h_n; k), \quad (8.86)$$

où  $h_j$  désigne toujours la valeur de la fonction de hachage pour la  $j$ ème clé, et où  $k$  est l'indice de la clé que l'on recherche (celle qui correspond à  $h_k$ ). Ainsi, on a  $1 \leq h_j \leq m$  pour tout  $1 \leq j \leq n$ , et  $1 \leq k \leq n$ ; il y a  $m^n \cdot n$  événements élémentaires  $\omega$  en tout.

Soit  $s_j$  la probabilité que nous cherchions la  $j$ ème clé à avoir été insérée dans la table. Alors

$$\Pr(\omega) = s_k/m^n \quad (8.87)$$

si  $\omega$  est l'événement (8.86). Selon les applications, on cherche plus souvent les enregistrements qui ont été insérés les premiers ou bien ceux qui ont été insérés parmi les derniers ; c'est pourquoi nous ne supposerons pas que  $s_j = 1/n$  pour tout  $j$ . Notez que  $\sum_{\omega \in \Omega} \Pr(\omega) = \sum_{k=1}^n s_k = 1$  et que, par conséquent, (8.87) définit bien une distribution de probabilité.

Si la clé  $K$  est la  $j$ ème dans sa liste, Le nombre  $P$  de tests lors de sa recherche est égal à  $P$ . Par conséquent,

$$P(h_1, \dots, h_n; k) = [h_1 = h_k] + [h_2 = h_k] + \cdots + [h_k = h_k]. \quad (8.88)$$

Si  $X_j$  désigne la variable aléatoire  $[h_j = h_k]$ , on peut écrire, de manière équivalente,

$$P = X_1 + X_2 + \cdots + X_k. \quad (8.89)$$

Supposons par exemple que  $m = 10$ ,  $n = 16$ , et que les valeurs de la fonction de hachage soient réparties "aléatoirement" comme suit :

$$(h_1, \dots, h_{16}) = 3 \ 1 \ 4 \ 1 \ 5 \ 9 \ 2 \ 6 \ 5 \ 3 \ 5 \ 8 \ 9 \ 7 \ 9 \ 3;$$

$$(P_1, \dots, P_{16}) = 1 \ 1 \ 1 \ 2 \ 1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 3 \ 1 \ 2 \ 1 \ 3 \ 3.$$

Le nombre  $P_j$  de tests nécessaires pour trouver la  $j$ ème clé est indiqué sous chaque  $h_j$ .

Dans l'équation (8.89),  $P$  est représenté par une somme de variables aléatoires. Cependant, on n'a pas le droit de dire que la moyenne  $E P$  est égale à  $E X_1 + \cdots + E X_k$ , car  $k$  est aussi une variable aléatoire. Quelle est la fonction génératrice de  $P$ ? Pour répondre à cette question, il nous faut d'abord faire une petite digression pour parler de *probabilités conditionnelles*.

J'ai déjà vu cette suite de chiffres quelque part.

L'équation (8.43) était aussi une petite digression.

Soient A et B deux événements d'un espace de probabilité donné. La probabilité conditionnelle de A sachant que B est réalisé est

$$\Pr(\omega \in A | \omega \in B) = \frac{\Pr(\omega \in A \cap B)}{\Pr(\omega \in B)}. \quad (8.90)$$

Par exemple, si X et Y sont des variables aléatoires, alors la probabilité de l'événement  $X = x$  sachant que  $Y = y$  est réalisé est

$$\Pr(X=x | Y=y) = \frac{\Pr(X=x \text{ et } Y=y)}{\Pr(Y=y)}. \quad (8.91)$$

Pour tout  $y \in Y$  donné, la somme de ces probabilités conditionnelles sur tous les  $x \in X$  vaut  $\Pr(Y=y)/\Pr(Y=y) = 1$ . Par conséquent, (8.91) définit une distribution de probabilité, et on peut définir une nouvelle variable aléatoire " $X|y$ " telle que  $\Pr((X|y)=x) = \Pr(X=x | Y=y)$ .

Si X et Y sont indépendantes, la variable aléatoire  $X|y$  sera exactement la même que X, quelle que soit la valeur de y, car, d'après (8.5),  $\Pr(X=x | Y=y)$  est égale à  $\Pr(X=x)$ ; c'est là la réelle signification de l'indépendance. Par contre, si X et Y ne sont pas indépendantes, les variables aléatoires  $X|y$  et  $X|y'$  n'ont absolument rien en commun lorsque  $y \neq y'$ .

Si X ne prend que des valeurs entières positives ou nulles, on peut décomposer sa fgp en une somme de fgp conditionnelles en fonction de n'importe quelle autre variable aléatoire Y :

$$G_X(z) = \sum_{y \in Y(\Omega)} \Pr(Y=y) G_{X|y}(z). \quad (8.92)$$

En effet, dans le membre gauche, le coefficient de  $z^x$  est égal à  $\Pr(X=x)$  pour tout  $x \in X(\Omega)$ , tandis que dans le membre droit il vaut

$$\begin{aligned} & \sum_{y \in Y(\Omega)} \Pr(Y=y) \Pr(X=x | Y=y) \\ &= \sum_{y \in Y(\Omega)} \Pr(X=x \text{ et } Y=y) \\ &= \Pr(X=x). \end{aligned}$$

Par exemple, si X est le produit des valeurs de deux dés après un jet et si Y désigne la somme de ces valeurs, alors la fgp de  $X|6$  est

$$G_{X|6}(z) = \frac{2}{5}z^5 + \frac{2}{5}z^8 + \frac{1}{5}z^9$$

car, si  $Y = 6$ , l'ensemble des événements possibles est  $\{\boxed{\bullet}\boxed{\bullet}, \boxed{\bullet}\boxed{\circ}, \boxed{\circ}\boxed{\bullet}, \boxed{\circ}\boxed{\circ}\}$ , et ils ont tous la même probabilité. Dans ce cas, l'équation

(8.92) se réduit à

$$\begin{aligned} G_X(z) = & \frac{1}{36}G_{X|2}(z) + \frac{2}{36}G_{X|3}(z) + \frac{3}{36}G_{X|4}(z) + \frac{4}{36}G_{X|5}(z) \\ & \frac{5}{36}G_{X|6}(z) + \frac{6}{36}G_{X|7}(z) + \frac{5}{36}G_{X|8}(z) + \frac{4}{36}G_{X|9}(z) \\ & \frac{3}{36}G_{X|10}(z) + \frac{2}{36}G_{X|11}(z) + \frac{1}{36}G_{X|12}(z). \end{aligned}$$

*Je comprends maintenant ce que veulent dire les mathématiciens lorsqu'ils déclarent que quelque chose est "évident", "clair" ou "trivial".*

*"By clearly, I mean a good freshman should be able to do it, although it's not completely trivial."*

—Paul Erdős [94].

Cette formule est évidente une fois qu'on l'a comprise. Fin de la digression.

Dans le cas du hachage, on peut trouver la fgp du nombre de tests lors d'une recherche fructueuse en posant  $X = P$  et  $Y = K$  dans (8.92). Pour tout  $k$  fixé entre 1 et  $n$ , la variable aléatoire  $P|k$  est définie comme une somme de variables aléatoires indépendantes  $X_1 + \dots + X_k$ , comme indiqué dans (8.89). Sa fgp est donc

$$G_{P|k}(z) = \left(\frac{m-1+z}{m}\right)^{k-1} z.$$

Par conséquent, la fgp de  $P$  vaut clairement

$$\begin{aligned} G_P(z) &= \sum_{k=1}^n s_k G_{P|k}(z) \\ &= \sum_{k=1}^n s_k \left(\frac{m-1+z}{m}\right)^{k-1} z \\ &= z S\left(\frac{m-1+z}{m}\right), \end{aligned} \tag{8.93}$$

où

$$S(z) = s_1 + s_2 z + s_3 z^2 + \dots + s_n z^{n-1} \tag{8.94}$$

désigne la fgp des probabilités  $s_k$  (que l'on a divisée par  $z$  pour des raisons pratiques).

Bien. Nous avons enfin notre fonction génératrice. En la dérivant, nous allons trouver sa moyenne et sa variance. En fait il vaut mieux, comme nous l'avons déjà fait en d'autres circonstances, supprimer d'abord le facteur  $z$ , pour trouver, au lieu des moyenne et variance de  $P$ , celles de  $P - 1$  :

$$\begin{aligned} F(z) &= G_P(z)/z = S\left(\frac{m-1+z}{m}\right); \\ F'(z) &= \frac{1}{m} S'\left(\frac{m-1+z}{m}\right); \quad F''(z) = \frac{1}{m^2} S''\left(\frac{m-1+z}{m}\right). \end{aligned}$$

Par conséquent,

$$\mathbb{E}P = 1 + \text{Moy}(F) = 1 + F'(1) = 1 + m^{-1} \text{Moy}(S); \quad (8.95)$$

$$\begin{aligned} \text{VP} &= \text{Var}(F) = F''(1) + F'(1) - F'(1)^2 \\ &= m^{-2}S''(1) + m^{-1}S'(1) - m^{-2}S'(1)^2 \\ &= m^{-2} \text{Var}(S) + (m^{-1} - m^{-2}) \text{Moy}(S). \end{aligned} \quad (8.96)$$

Ces formules expriment la moyenne et la variance du nombre P de tests en fonction de la moyenne et de la variance de la distribution S.

Supposons par exemple que  $s_k = 1/n$  pour tout  $1 \leq k \leq n$ . Cela signifie qu'on effectue une recherche fructueuse totalement "aléatoire", en ce sens que toutes les clés de la table ont la même probabilité d'être recherchées. Alors  $S(z)$  est la distribution de probabilité uniforme  $U_n(z)$  de (8.32), et on a  $\text{Moy}(S) = (n-1)/2$  et  $\text{Var}(S) = (n^2-1)/12$ . Par conséquent,

$$\mathbb{E}P = \frac{n-1}{2m} + 1; \quad (8.97)$$

$$\text{VP} = \frac{n^2-1}{12m^2} + \frac{(m-1)(n-1)}{2m^2} = \frac{(n-1)(6m+n-5)}{12m^2}. \quad (8.98)$$

Ici aussi, nous gagnons en vitesse d'un facteur de  $1/m$ . Si  $m \approx n/\ln n$  et  $n \rightarrow \infty$ , le nombre moyen de tests par recherche fructueuse vaut à peu près  $\frac{1}{2} \ln n$ , et l'écart-type est asymptotiquement égal à  $(\ln n)/\sqrt{12}$ .

Voyons ce qui se passe si on suppose que  $s_k = (kH_n)^{-1}$  pour  $1 \leq k \leq n$ . Cette distribution est appelée la "loi de Zipf". Dans ce cas,  $\text{Moy}(S) = n/H_n - 1$  et  $\text{Var}(S) = \frac{1}{2}n(n+1)/H_n - n^2/H_n^2$ . Le nombre moyen de tests, pour  $m \approx n/\ln n$  et  $n \rightarrow \infty$ , est à peu près 2, et l'écart type tend asymptotiquement vers  $\sqrt{\ln n}/\sqrt{2}$ .

Dans les deux cas, ceux d'entre nous qui craignent le pire des cas peuvent se rassurer : l'inégalité de Tchebychev nous assure que les listes seront courtes, sauf sans des cas extrêmement rares.

### **Cas 2, suite : variations sur la variance.**

Nous venons de calculer la variance du nombre de tests lors d'une recherche fructueuse. Pour cela, nous avons considéré P comme une variable aléatoire sur un espace de probabilité à  $m^n \cdot n$  éléments qui s'écrivent  $(h_1, \dots, h_n; k)$ . Nous aurions aussi pu adopter un autre point de vue : chaque configuration  $(h_1, \dots, h_n)$  de valeurs de la fonction de hachage définit une variable aléatoire  $P|(h_1, \dots, h_n)$  qui représente le nombre de tests effectués lors d'une recherche fructueuse dans une table de hachage parti-

*Les gars, encore une fois, vous pouvez lire ça en diagonale.*

*— Votre sympathique chargé de TD*

culière de  $n$  clés données. La valeur moyenne de  $P|(h_1, \dots, h_n)$ ,

$$A(h_1, \dots, h_n) = \sum_{p=1}^n p \cdot \Pr((P|(h_1, \dots, h_n)) = p), \quad (8.99)$$

représente le temps moyen d'exécution d'une telle recherche. C'est une variable aléatoire qui ne dépend que de  $(h_1, \dots, h_n)$ , et pas de  $k$ . On peut l'écrire sous la forme

$$A(h_1, \dots, h_n) = \sum_{k=1}^n s_k P(h_1, \dots, h_n; k),$$

où  $P(h_1, \dots, h_n; k)$  est défini comme en (8.88). En effet, la probabilité que  $P|(h_1, \dots, h_n) = p$  est

$$\begin{aligned} \frac{\sum_{k=1}^n \Pr(P(h_1, \dots, h_n; k) = p)}{\sum_{k=1}^n \Pr(h_1, \dots, h_n; k)} &= \frac{\sum_{k=1}^n m^{-n} s_k [P(h_1, \dots, h_n; k) = p]}{\sum_{k=1}^n m^{-n} s_k} \\ &= \sum_{k=1}^n s_k [P(h_1, \dots, h_n; k) = p]. \end{aligned}$$

La valeur moyenne de  $A(h_1, \dots, h_n)$ , que l'on peut obtenir en sommant sur les  $m^n$  configurations  $(h_1, \dots, h_n)$  possibles puis en divisant par  $m^n$ , est évidemment la même que celle que nous avons déjà calculée en (8.95). En revanche, la *variance* de  $A(h_1, \dots, h_n)$  est différente : ici, il s'agit de la variance de  $m^n$  moyennes, et non la variance de  $m^n \cdot n$  nombres de tests. Par exemple, si  $m = 1$  (il n'y a qu'une liste), la valeur "moyenne"  $A(h_1, \dots, h_n) = A(1, \dots, 1)$  est en fait une constante, donc sa variance  $VA$  est nulle. Par contre, comme le nombre de tests lors d'une recherche fructueuse n'est pas constant, la variance  $VP$  n'est pas nulle.

Illustrons cette différence en effectuant les calculs pour  $m$  et  $n$  quelconques dans le cas le plus simple, celui où  $s_k = 1/n$  pour tout  $1 \leq k \leq n$ . Nous supposons donc, pour cette fois, que la distribution des clés cherchées est uniforme. Toute suite de valeurs de la fonction de hachage  $(h_1, \dots, h_n)$  détermine  $m$  listes qui contiennent respectivement  $(n_1, n_2, \dots, n_m)$  enregistrements, où les nombres  $n_j$  sont tels que  $n_1 + n_2 + \dots + n_m = n$ . Une recherche fructueuse donnera lieu à un nombre moyen de tests égal à

$$\begin{aligned} A(h_1, \dots, h_n) &= \frac{(1+\dots+n_1) + (1+\dots+n_2) + \dots + (1+\dots+n_m)}{n} \\ &= \frac{n_1(n_1+1) + n_2(n_2+1) + \dots + n_m(n_m+1)}{2n} \\ &= \frac{n_1^2 + n_2^2 + \dots + n_m^2 + n}{2n}. \end{aligned}$$

Nous voulons calculer la variance de  $A(h_1, \dots, h_n)$  sur l'espace de probabilité constitué des  $m^n$  suites  $(h_1, \dots, h_n)$ .

Nous allons en fait, pour simplifier les calculs, calculer la variance de quelque chose de légèrement différent :

$$B(h_1, \dots, h_n) = \binom{n_1}{2} + \binom{n_2}{2} + \dots + \binom{n_m}{2}.$$

Comme

$$A(h_1, \dots, h_n) = 1 + B(h_1, \dots, h_n)/n,$$

la moyenne et la variance de  $A$  satisfont

$$EA = 1 + \frac{EB}{n}; \quad VA = \frac{VB}{n^2}. \quad (8.100)$$

La probabilité que les longueurs respectives des listes soient  $n_1, n_2, \dots, n_m$  est égale au coefficient multinomial

$$\binom{n}{n_1, n_2, \dots, n_m} = \frac{n!}{n_1! n_2! \dots n_m!}$$

divisé par  $m^n$ . Voici par conséquent la fgp de  $B(h_1, \dots, h_n)$  :

$$B_n(z) = \sum_{\substack{n_1, n_2, \dots, n_m \geq 0 \\ n_1 + n_2 + \dots + n_m = n}} \binom{n}{n_1, n_2, \dots, n_m} z^{\binom{n_1}{2} + \binom{n_2}{2} + \dots + \binom{n_m}{2}} m^{-n}.$$

Cette somme effraierait certainement des moins expérimentés que nous. Quant à nous, grâce à notre expérience acquise au chapitre 7, nous reconnaissons bien sûr une convolution de  $m$  suites. En effet, si on considère la double fonction génératrice exponentielle

$$G(w, z) = \sum_{n \geq 0} B_n(z) \frac{m^n w^n}{n!},$$

on vérifie facilement que  $G(w, z)$  est tout simplement une puissance mième :

$$G(w, z) = \left( \sum_{k \geq 0} z^{\binom{k}{2}} \frac{w^k}{k!} \right)^m.$$

Juste pour vérifier, posons  $z = 1$ . Nous obtenons  $G(w, 1) = (e^w)^m$ , donc le coefficient de  $m^n w^n/n!$  est  $B_n(1) = 1$ .

Pour trouver  $\text{Var}(B_n)$ , il nous faut connaître les valeurs de  $B'_n(1)$  et  $B''_n(1)$ . Nous allons donc calculer les dérivées partielles de  $G(w, z)$  par

rapport à  $z$  :

$$\begin{aligned}\frac{\partial}{\partial z} G(w, z) &= \sum_{n \geq 0} B'_n(z) \frac{m^n w^n}{n!} \\&= m \left( \sum_{k \geq 0} z^{\binom{k}{2}} \frac{w^k}{k!} \right)^{m-1} \sum_{k \geq 0} \binom{k}{2} z^{\binom{k}{2}-1} \frac{w^k}{k!}; \\ \frac{\partial^2}{\partial z^2} G(w, z) &= \sum_{n \geq 0} B''_n(z) \frac{m^n w^n}{n!} \\&= m(m-1) \left( \sum_{k \geq 0} z^{\binom{k}{2}} \frac{w^k}{k!} \right)^{m-2} \left( \sum_{k \geq 0} \binom{k}{2} z^{\binom{k}{2}-1} \frac{w^k}{k!} \right)^2 \\&\quad + m \left( \sum_{k \geq 0} z^{\binom{k}{2}} \frac{w^k}{k!} \right)^{m-1} \sum_{k \geq 0} \binom{k}{2} \left( \binom{k}{2} - 1 \right) z^{\binom{k}{2}-2} \frac{w^k}{k!}.\end{aligned}$$

C'est assez compliqué, avouons le, mais tout se simplifie très bien si on pose  $z = 1$ . Par exemple, on a

$$\begin{aligned}\sum_{n \geq 0} B'_n(1) \frac{m^n w^n}{n!} &= m e^{(m-1)w} \sum_{k \geq 2} \frac{w^k}{2(k-2)!} \\&= m e^{(m-1)w} \sum_{k \geq 0} \frac{w^{k+2}}{2k!} \\&= \frac{mw^2 e^{(m-1)w}}{2} e^w = \sum_{n \geq 0} \frac{(mw)^{n+2}}{2m n!} = \sum_{n \geq 0} \frac{n(n-1)m^n w^n}{2m n!},\end{aligned}$$

et il s'ensuit que

$$B'_n(1) = \binom{n}{2} \frac{1}{m}. \tag{8.101}$$

D'après (8.100), nous obtenons  $E A = 1 + (n-1)/2m$ , ce qui concorde avec (8.97).

En ce qui concerne  $B''_n(1)$ , on trouve une somme similaire,

$$\begin{aligned}\sum_{k \geq 0} \binom{k}{2} \left( \binom{k}{2} - 1 \right) \frac{w^k}{k!} &= \frac{1}{4} \sum_{k \geq 0} \frac{(k+1)k(k-1)(k-2)w^k}{k!} \\&= \frac{1}{4} \sum_{k \geq 3} \frac{(k+1)w^k}{(k-3)!} = \frac{1}{4} \sum_{k \geq 0} \frac{(k+4)w^{k+3}}{k!} = (\frac{1}{4}w^4 + w^3)e^w.\end{aligned}$$

Par conséquent,

$$\begin{aligned}
 & \sum_{n \geq 0} B_n''(1) \frac{m^n w^n}{n!} \\
 &= m(m-1)e^{w(m-2)} \left(\frac{1}{2}w^2 e^w\right)^2 + m e^{w(m-1)} \left(\frac{1}{4}w^4 + w^3\right) e^w \\
 &= m e^{wm} \left(\frac{1}{4}mw^4 + w^3\right); \\
 B_n''(1) &= \binom{n}{2} \left( \binom{n}{2} - 1 \right) \frac{1}{m^2}. \tag{8.102}
 \end{aligned}$$

Nous avons maintenant tous les éléments nécessaires pour calculer la variance  $V\lambda$ . Le résultat est étonnamment simple :

$$\begin{aligned}
 V\lambda &= \frac{VB}{n^2} = \frac{B_n''(1) + B_n'(1) - B_n'(1)^2}{n^2} \\
 &= \frac{n(n-1)}{m^2 n^2} \left( \frac{(n+1)(n-2)}{4} + \frac{m}{2} - \frac{n(n-1)}{4} \right) \\
 &= \frac{(m-1)(n-1)}{2m^2 n}. \tag{8.103}
 \end{aligned}$$

Quand une formule est si simple, on peut raisonnablement soupçonner qu'il y a une bonne raison mathématique à cela ; il doit y avoir une autre façon d'attaquer le problème qui mène directement à ce résultat. En effet, on voit dans l'exercice 61 que, par une autre approche, on peut montrer que la variance d'une recherche fructueuse moyenne peut s'écrire sous la forme

$$V\lambda = \frac{m-1}{m^2} \sum_{k=1}^n s_k^2(k-1), \tag{8.104}$$

où  $s_k$  est la probabilité que le  $k$ ième élément qui a été inséré soit celui qu'on cherche. L'équation (8.103) correspond au cas particulier  $s_k = 1/n$  pour tout  $1 \leq k \leq n$ .

Jusqu'à présent, nous avons considéré la variance de la moyenne ; passons maintenant à la moyenne de la variance. Soyons plus précis : chaque suite  $(h_1, \dots, h_n)$  qui définit une table de hachage définit aussi une distribution de probabilité pour la recherche fructueuse ; la variance de cette distribution décrit la répartition du nombre de tests nécessaires pour les recherches fructueuses dans la table. Revenons par exemple à notre table de  $n = 16$  enregistrements dans  $m = 10$  listes :

$$(h_1, \dots, h_{16}) = 3 \ 1 \ 4 \ 1 \ 5 \ 9 \ 2 \ 6 \ 5 \ 3 \ 5 \ 8 \ 9 \ 7 \ 9 \ 3$$

$$(P_1, \dots, P_{16}) = 1 \ 1 \ 1 \ 2 \ 1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 3 \ 1 \ 2 \ 1 \ 3 \ 3$$

*J'ai déjà vu cette suite de chiffres quelque part.*

*J'ai déjà vu ce graffiti quelque part.*

La fgp d'une recherche fructueuse dans la table de hachage correspondante est

$$\begin{aligned} G(3, 1, 4, 1, \dots, 3) &= \sum_{k=1}^{16} s_k z^{P(3, 1, 4, 1, \dots, 3; k)} \\ &= s_1 z + s_2 z + s_3 z + s_4 z^2 + \dots + s_{16} z^3. \end{aligned}$$

Nous avons déjà considéré le nombre moyen de tests lors d'une recherche fructueuse dans cette table,  $A(3, 1, 4, 1, \dots, 3) = \text{Moy}(G(3, 1, 4, 1, \dots, 3))$ . On peut aussi considérer la variance

$$\begin{aligned} s_1 \cdot 1^2 + s_2 \cdot 1^2 + s_3 \cdot 1^2 + s_4 \cdot 2^2 + \dots + s_{16} \cdot 3^2 \\ - (s_1 \cdot 1 + s_2 \cdot 1 + s_3 \cdot 1 + s_4 \cdot 2 + \dots + s_{16} \cdot 3)^2. \end{aligned}$$

Comme cette variance est une variable aléatoire qui dépend de  $(h_1, \dots, h_n)$ , il est naturel de se demander quelle est sa valeur moyenne.

Récapitulons tout cela. Il y a trois sortes de variance qui nous intéressent dans le problème de la recherche fructueuse : la *variance totale* du nombre de tests, prise sur toutes les suites  $(h_1, \dots, h_n)$  et sur tous les  $k$ ; la *variance de la moyenne* du nombre de tests, où la moyenne est faite sur tous les  $k$ , et où la variance est donc prise sur toutes les suites  $(h_1, \dots, h_n)$ ; enfin, la *moyenne de la variance* du nombre de tests, où la variance est prise sur tous les  $k$ , et où la moyenne est donc faite sur toutes les suites  $(h_1, \dots, h_n)$ . De façon plus formelle, voici respectivement la variance totale, la variance de la moyenne et la moyenne de la variance :

$$\begin{aligned} VP &= \sum_{1 \leq h_1, \dots, h_n \leq m} \sum_{k=1}^n \frac{s_k}{m^n} P(h_1, \dots, h_n; k)^2 \\ &\quad - \left( \sum_{1 \leq h_1, \dots, h_n \leq m} \sum_{k=1}^n \frac{s_k}{m^n} P(h_1, \dots, h_n; k) \right)^2; \\ VA &= \sum_{1 \leq h_1, \dots, h_n \leq m} \frac{1}{m^n} \left( \sum_{k=1}^n s_k P(h_1, \dots, h_n; k) \right)^2 \\ &\quad - \left( \sum_{1 \leq h_1, \dots, h_n \leq m} \frac{1}{m^n} \sum_{k=1}^n s_k P(h_1, \dots, h_n; k) \right)^2; \\ AV &= \sum_{1 \leq h_1, \dots, h_n \leq m} \frac{1}{m^n} \left( \sum_{k=1}^n s_k P(h_1, \dots, h_n; k)^2 \right. \\ &\quad \left. - \left( \sum_{k=1}^n s_k P(h_1, \dots, h_n; k) \right)^2 \right). \end{aligned}$$

Ces trois quantités sont liées de façon très simple par la formule

$$VX = VA + AV. \quad (8.105)$$

En fait, plus généralement, pour toutes variable aléatoire  $X$  à valeurs réelles et toute variable aléatoire  $Y$ , on a l'identité

$$VX = V(E(X|Y)) + E(V(X|Y)) \quad (8.106)$$

(on la démontre dans l'exercice 22). L'équation (8.105) en est un cas particulier, dans lequel  $X$  représente le nombre d'essais lors d'une recherche fructueuse et  $Y$  est la suite  $(h_1, \dots, h_n)$  des valeurs de la fonction de hachage.

L'équation (8.106) mérite d'être examinée de très près, car il n'est pas facile au premier abord de bien voir quelles sont les différentes variables aléatoires et sur quels espaces de probabilité sont calculées les moyennes et variances. Pour toute valeur  $y$  de  $Y$ , la variable aléatoire  $X|y$  est définie comme en (8.91), et sa moyenne  $E(X|y)$  dépend de  $y$ . Alors  $E(X|Y)$  désigne la variable aléatoire qui prend les valeurs  $E(X|y)$  lorsque  $y$  parcourt toutes les valeurs de  $Y$ , et  $V(E(X|Y))$  est la variance de cette variable aléatoire selon la distribution de probabilité de  $Y$ . De même,  $E(V(X|Y))$  désigne la moyenne des variables aléatoires  $V(X|y)$  lorsque  $y$  varie. Comme une variance est toujours positive ou nulle, on déduit de (8.106) deux propriétés de  $VX$ , la variance inconditionnelle de  $X$  :

$$VX \geq V(E(X|Y)) \quad \text{et} \quad VX \geq E(V(X|Y)). \quad (8.107)$$

*(C'est le moment de s'échauffer avec l'exercice 6).*

### Retour au cas 1 : recherche infructueuse (bis).

Terminons notre étude du hachage par un autre calcul typique de ce qu'on fait en analyse d'algorithmes. Cette fois, nous allons nous intéresser au temps total d'exécution d'une recherche infructueuse et d'une insertion de la nouvelle clé (et de son enregistrement) dans la table.

Dans la procédure d'insertion de (8.83), il y a deux cas possibles, selon que  $j$  est strictement négatif ou nul. Il est strictement négatif si et seulement si  $P = 0$ , car une valeur négative ne peut provenir que de la case du tableau PREMIER associée à une liste vide. Par conséquent, si la liste est vide, on a  $P = 0$  et il faut exécuter  $\text{PREMIER}[h_{n+1}] := n + 1$ , pour insérer le nouvel enregistrement en position  $n + 1$ . Dans le cas contraire, on a  $P > 0$ , et c'est dans une case du tableau SUIVANT qu'il faut mettre la valeur  $n + 1$ . Les temps d'exécution de ces deux cas peuvent être différents. Par conséquent, le temps total d'exécution d'une recherche infructueuse suivie d'une insertion s'écrit

$P$  désigne toujours le nombre de tests.

$$T = \alpha + \beta P + \delta[P = 0], \quad (8.108)$$

où  $\alpha$ ,  $\beta$  et  $\delta$  sont des constantes qui dépendent de l'ordinateur qu'on utilise et de la façon dont la méthode de hachage est programmée dans son langage interne. Il serait intéressant de connaître la moyenne et la variance de  $T$ , car ce genre d'informations est plus utile en pratique que la moyenne et la variance de  $P$ .

Jusqu'ici, nous n'avons utilisé les fonctions génératrices de probabilité que pour des variables aléatoires à valeurs positives ou nulles. On peut aussi travailler de la même manière avec

$$G_X(z) = \sum_{\omega \in \Omega} \Pr(\omega) z^{X(\omega)}$$

pour toute variable  $X$  à valeurs réelles, car les caractéristiques essentielles de  $X$  ne dépendent que du comportement de  $G_X$  au voisinage de  $z = 1$ , voisinage dans lequel les puissances de  $z$  sont bien définies. Par exemple, le temps d'exécution (8.108) d'une recherche infructueuse est une variable aléatoire définie sur l'espace des valeurs équiprobables  $(h_1, \dots, h_n, h_{n+1})$  de la fonction de hachage, avec  $1 \leq h_j \leq m$ . On peut considérer que la série

$$G_T(z) = \frac{1}{m^{n+1}} \sum_{h_1=1}^m \cdots \sum_{h_n=1}^m \sum_{h_{n+1}=1}^m z^{\alpha + \beta P(h_1, \dots, h_{n+1}) + \delta [P(h_1, \dots, h_{n+1}) = 0]}$$

est une fgp même si  $\alpha$ ,  $\beta$  et  $\delta$  ne sont pas des entiers. (En fait,  $\alpha$ ,  $\beta$  et  $\delta$  ne sont même pas des nombres purs, ce sont des paramètres physiques qui représentent des unités de temps ; cela ne nous empêche pas de nous en servir dans l'exposant de  $z$ ). Nous pouvons alors trouver la moyenne et la variance de  $T$ , comme d'habitude, au moyen de  $G'_T(1)$  et  $G''_T(1)$ .

La fonction génératrice de  $P$  est

$$P(z) = \left( \frac{m-1+z}{m} \right)^n = \sum_{p \geq 0} \Pr(P=p) z^p.$$

Par conséquent,

$$\begin{aligned} G_T(z) &= \sum_{p \geq 0} \Pr(P=p) z^{\alpha + \beta p + \delta [p=0]} \\ &= z^\alpha \left( (z^\delta - 1) \Pr(P=0) + \sum_{p \geq 0} \Pr(P=p) z^{\beta p} \right) \\ &= z^\alpha \left( (z^\delta - 1) \left( \frac{m-1}{m} \right)^n + \left( \frac{m-1+z^\beta}{m} \right)^n \right). \end{aligned}$$

C'est maintenant pure routine que de calculer  $\text{Moy}(G_T)$  et  $\text{Var}(G_T)$  :

$$\text{Moy}(G_T) = G'_T(1) = \alpha + \beta \frac{n}{m} + \delta \left( \frac{m-1}{m} \right)^n; \quad (8.109)$$

$$\begin{aligned} G_T''(1) &= \alpha(\alpha - 1) + 2\alpha\beta\frac{n}{m} + \beta(\beta - 1)\frac{n}{m} + \beta^2\frac{n(n-1)}{m^2} \\ &\quad + 2\alpha\delta\left(\frac{m-1}{m}\right)^n + \delta(\delta - 1)\left(\frac{m-1}{m}\right)^n; \end{aligned}$$

$$\begin{aligned} \text{Var}(G_T) &= G_T''(1) + G_T'(1) - G_T'(1)^2 \\ &= \beta^2\frac{n(m-1)}{m^2} - 2\beta\delta\left(\frac{m-1}{m}\right)^n\frac{n}{m} \\ &\quad + \delta^2\left(\left(\frac{m-1}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right). \quad (8.110) \end{aligned}$$

Dans le chapitre 9, nous apprendrons à estimer des quantités de ce genre lorsque  $m$  et  $n$  sont grands. Par exemple, si  $m = n$  et  $n \rightarrow \infty$ , on peut montrer que la moyenne et la variance de  $T$  valent respectivement  $\alpha + \beta + \delta e^{-1} + O(n^{-1})$  et  $\beta^2 - 2\beta\delta e^{-1} + \delta^2(e^{-1} - e^{-2}) + O(n^{-1})$ . Si  $m = n/\ln n + O(1)$  et  $n \rightarrow \infty$ , on trouve

$$\begin{aligned} \text{Moy}(G_T) &= \beta \ln n + \alpha + O((\log n)^2/n); \\ \text{Var}(G_T) &= \beta^2 \ln n + O((\log n)^2/n). \end{aligned}$$

## Exercices

### Echauffements

- 1 Quelle est la probabilité de tirer un double dans la distribution de probabilité  $\Pr_{01}$  de (8.3), si l'un des dés est juste tandis que l'autre est pipé ? Quelle est la probabilité d'obtenir  $S = 7$  ?
- 2 Quelle est la probabilité que la carte du dessus et la carte du dessous d'un paquet de cartes soient tous deux des as ? (On suppose que les 52! permutations possibles ont toutes la même probabilité  $1/52!$ ).
- 3 En 1979, on a demandé aux étudiants en mathématiques concrètes de Stanford de tirer à pile ou face jusqu'à obtenir deux piles à la suite et d'indiquer le nombre de lancers qu'il leur avait fallu pour cela. Voici leurs réponses :

3, 2, 3, 5, 10, 2, 6, 6, 9, 2.

En 1987, les étudiants de Princeton ont fait la même expérience, obtenant les résultats

10, 2, 10, 7, 5, 2, 10, 6, 10, 2.

Calculez la moyenne et la variance de (a) l'échantillon de Stanford ; (b) l'échantillon de Princeton.

*Pourquoi n'y a-t-il que dix nombres ?*

*Peut-être que les autres étudiants n'étaient pas des expérimentateurs, ou qu'ils ne voulaient pas faire face... à leurs obligations.*

- 4 Soit  $H(z) = F(z)/G(z)$ , avec  $F(1) = G(1) = 1$ . Montrez que, de façon analogue à (8.38) et (8.39),

$$\begin{aligned} \text{Moy}(H) &= \text{Moy}(F) - \text{Moy}(G), \\ \text{Var}(H) &= \text{Var}(F) - \text{Var}(G), \end{aligned}$$

si les dérivées correspondantes existent en  $z = 1$ .

- 5 Supposez qu'Alice et Bill jouent au jeu (8.78) avec une pièce truquée telle que la probabilité de faire pile soit  $p$ . Existe-t-il une valeur de  $p$  pour laquelle le jeu devienne équitable ?
- 6 A quoi se réduit la règle (8.106) si  $X$  et  $Y$  sont des variables aléatoires indépendantes ?

### *Exercices de base*

- 7 Montrez que si on jette deux dés qui ont la même distribution de probabilité, alors la probabilité de tirer un double est toujours supérieure ou égale à  $\frac{1}{6}$ .
- 8 Soient  $A$  et  $B$  deux événements tels que  $A \cup B = \Omega$ . Montrez que

$$\Pr(\omega \in A \cap B) = \Pr(\omega \in A) \Pr(\omega \in B) - \Pr(\omega \notin A) \Pr(\omega \notin B).$$

- 9 Prouvez ou réfutez : si  $X$  et  $Y$  sont des variables aléatoires indépendantes, alors, pour toutes fonctions  $F$  et  $G$ ,  $F(X)$  et  $G(Y)$  sont aussi des variables aléatoires indépendantes.
- 10 Selon la définition (8.7), quel est le nombre maximum d'éléments de la médiane d'une variable aléatoire  $X$  ?
- 11 Construisez une variable aléatoire dont la moyenne est finie et dont la variance est infinie.
- 12 a Si  $P(z)$  désigne la fgp de la variable aléatoire  $X$ , montrez que

$$\begin{aligned} \Pr(X \leq r) &\leq x^{-r}P(x) \quad \text{pour } 0 < x \leq 1; \\ \Pr(X \geq r) &\leq x^{-r}P(x) \quad \text{pour } x \geq 1. \end{aligned}$$

- b Dans le cas particulier  $P(z) = (1+z)^{-n}/2^n$ , utilisez la première inégalité pour prouver que  $\sum_{k \leq \alpha n} \binom{n}{k} \leq 1/\alpha^{\alpha n} (1-\alpha)^{(1-\alpha)n}$  lorsque  $0 < \alpha < \frac{1}{2}$ .
- 13 Si  $X_1, \dots, X_{2n}$  sont des variables aléatoires indépendantes de même distribution, et si  $\alpha$  est un nombre réel quelconque, montrez que

$$\Pr\left(\left|\frac{X_1 + \dots + X_{2n}}{2n} - \alpha\right| \leq \left|\frac{X_1 + \dots + X_n}{n} - \alpha\right|\right) \geq \frac{1}{2}.$$

- 14 Soient  $F(z)$  et  $G(z)$  des fonctions génératrices de probabilité, et soit

$$H(z) = p F(z) + q G(z)$$

avec  $p+q=1$  (cela correspond à choisir la distribution  $F$  ou  $G$  en tirant à pile ou face). Calculez la moyenne et la variance de  $H$  en fonction de  $p$ , de  $q$  et des moyennes et variances de  $F$  et  $G$ .

- 15 Si  $F(z)$  et  $G(z)$  sont deux fonctions génératrices de probabilité, on peut définir une autre fgp  $H(z)$  par composition :

$$H(z) = F(G(z)).$$

Exprimez  $\text{Moy}(H)$  et  $\text{Var}(H)$  en fonction de  $\text{Moy}(F)$ ,  $\text{Var}(F)$ ,  $\text{Moy}(G)$  et  $\text{Var}(G)$ . (L'équation (8.93) en est un cas particulier).

- 16 Trouvez une forme close de la série génératrice double  $\sum_{n \geq 0} F_n(z)w^n$ , où  $F_n(z)$  est la fonction génératrice définie en (8.53).

- 17 Soient  $X_{n,p}$  et  $Y_{n,p}$  deux variables aléatoires qui obéissent respectivement à la loi binomiale et à la loi binomiale négative de paramètres  $(n, p)$  (ces distributions sont définies en (8.57) et (8.60)). Montrez que  $\Pr(Y_{n,p} \leq m) = \Pr(X_{m+n,p} \geq n)$ . Quelle identité de coefficients binomiaux cela entraîne-t-il ?

- 18 On dit qu'une variable aléatoire  $X$  est soumise à la *loi de Poisson* de moyenne  $\mu$  si  $\Pr(X=k) = e^{-\mu}\mu^k/k!$  pour tout  $k \geq 0$ .

- a Quelle est la fgp d'une telle variable ?  
b Que valent sa moyenne, sa variance et ses autres cumulants ?

*Le nombre de poisons par unité de volume d'eau.*

- 19 Pour continuer l'exercice précédent, soient  $X_1$  et  $X_2$  deux variables aléatoires indépendantes qui obéissent à des lois de Poisson de moyennes respectives  $\mu_1$  et  $\mu_2$ .

- a Quelle est la probabilité que  $X_1 + X_2 = n$  ?  
b Calculez la moyenne, la variance et les autres cumulants de  $2X_1 + 3X_2$ .

- 20 Démontrez les formules (8.74) et (8.75) qui donnent la moyenne et la variance du nombre de lancers nécessaires pour obtenir une suite donnée de piles et de faces.

- 21 Que représente la valeur de  $N$  si  $P$  et  $F$  sont tous deux égaux à  $\frac{1}{2}$  dans (8.77) ?

- 22 Démontrez la formule (8.106).

#### *Devoirs à la maison*

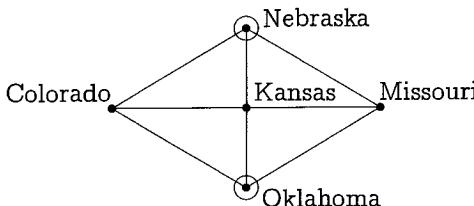
- 23 Soit  $\Pr_{00}$  la distribution de probabilité de deux dés justes et  $\Pr_{11}$  la distribution de probabilité de deux dés pipés donnée en (8.2). Trouvez

tous les événements  $A$  tels que  $\Pr_{00}(A) = \Pr_{11}(A)$ . Quels sont ceux qui ne dépendent que de la variable aléatoire  $S$ ? (Un espace de probabilité tel que  $\Omega = D^2$  contient  $2^{36}$  événements ; parmi ceux-ci,  $2^{11}$  éléments ne dépendent que de  $S$ ).

- 24 Le joueur J lance  $2n + 1$  dés (non pipés) et supprime tous ceux qui ont fait  $\boxed{\bullet\bullet}$ . Le joueur K annonce alors un nombre entre 1 et 6, lance les dés qui restent et supprime tous ceux qui donnent le nombre annoncé. Et ainsi de suite jusqu'à ce qu'il ne reste aucun dé. Le gagnant est celui qui a supprimé le plus de dés ( $n + 1$  ou plus).
- Trouvez la moyenne et la variance du nombre de dés supprimés par J. *Suggestion* : les dés sont indépendants.
  - Quelle est la probabilité que J gagne, si  $n = 2$ ?
- 25 Considérez le jeu suivant : vous misez une somme donnée  $A$  et lancez un dé (non pipé) ; si le résultat est  $k$ , vous gagnez votre mise multipliée par  $2(k - 1)/5$  (en particulier, vous doublez la mise si vous tirez  $\boxed{\bullet\bullet}$ , mais vous perdez tout si vous tirez  $\boxed{\bullet}$ ). Vous pouvez arrêter à tout moment et prendre la somme qu'il vous reste. Que valent la moyenne et la variance de cette somme après  $n$  lancers ?
- 26 Trouvez la moyenne et la variance du nombre de cycles de longueur  $l$  dans une permutation aléatoire de  $n$  éléments (le problème de la victoire au football, que nous avons vu en (8.23), (8.24) et (8.53), constitue le cas particulier  $l = 1$ ).
- 27 Soient  $X_1, X_2, \dots, X_n$  des échantillons indépendants de la variable aléatoire  $X$ . Les équations (8.19) et (8.20) montrent comment estimer la moyenne et la variance de  $X$  sur la base de ces observations. Donnez une formule analogue pour estimer le troisième cumulant  $\kappa_3$  (votre formule doit donner une estimation "juste", c'est-à-dire que son espérance doit être égale à  $\kappa_3$ ).
- 28 Quelle est la durée moyenne du jeu de pile ou face (8.78),
- sachant qu'Alice a gagné ?
  - sachant que Bill a gagné ?
- 29 Alice, Bill et le Calculateur tirent à pile ou face jusqu'à ce que l'un des motifs A = PPFP, B = PFPP ou C = FPPP apparaisse. (Si on ne considérait que deux de ces motifs, nous pourrions déduire de (8.82) que A battrait probablement B, que B battrait probablement C et que C battrait probablement A ; mais il faut considérer les trois motifs en même temps). Quels sont les chances de gagner de chaque joueur ?
- 30 On considère dans le texte du chapitre trois types de variance associées à la recherche fructueuse dans une table de hachage. En fait, il en existe deux de plus : on peut considérer la moyenne (sur  $k$ ) des variances

(sur  $h_1, \dots, h_n$ ) de  $P(h_1, \dots, h_n; k)$  ; et aussi la variance (sur  $k$ ) des moyennes (sur  $h_1, \dots, h_n$ ). Calculez ces quantités.

- 31** Une pomme est posée sur le sommet A d'un pentagone ABCDE, et un ver se trouve à deux sommets de là, en C. Chaque jour, le ver rampe, avec égales probabilités, vers l'un des sommets adjacents à celui sur lequel il se trouve. Ainsi, à la fin du premier jour, il se trouve sur le sommet B ou sur le sommet D avec probabilité  $\frac{1}{2}$ . Après deux jours, il peut se retrouver en C car il n'a aucune mémoire de son parcours. Lorsqu'il atteint le sommet A, il s'arrête pour dîner.
- a Que valent la moyenne et la variance du nombre de jours jusqu'au dîner ?
- b Soit  $p$  la probabilité que le nombre de jours soit supérieur ou égal à 100. Que dit l'inégalité de Tchebychev à propos de  $p$  ?
- c Que disent les inégalités de l'exercice 12 à propos de  $p$  ?
- 32** Alice et Bill sont à l'armée. Alice est basée dans le Nebraska et Bill dans l'Oklahoma. Il y a aussi trois autres bases dans le Kansas, le Missouri et le Colorado. Chaque mois, chaque militaire est réaffecté dans un état voisin de celui dans lequel il se trouve ; tous les états voisins à un état donné sont équiprobables. Voici le graphe des voisinages :



On y a entouré les états initiaux. Par exemple, après le premier mois, Alice peut être affectée soit dans le Colorado, soit dans le Kansas, soit dans le Missouri, avec probabilité  $1/3$  pour chacune des possibilités. Trouvez la moyenne et la variance du nombre de mois écoulés jusqu'à ce qu'Alice et Bill se retrouvent (vous avez le droit de recruter un ordinateur).

*Le ver de Schrödinger.*

*C'est vraiment un cas d'états finis.*

- 33** Les variables aléatoires  $X_1$  et  $X_2$  de (8.89) sont-elles indépendantes ?
- 34** Gina, joueuse de golf, a une probabilité  $p = 0,05$  à chaque coup de faire un tir exceptionnel, qui lui permet de gagner un point sur le par, une probabilité  $q = 0,91$  de faire un tir ordinaire, et un probabilité  $r = .04$  de rater son coup, ce qui lui fait perdre un point par rapport au par. Voici quelques explications pour ceux qui ne jouent pas au golf : à chaque tour, elle se rapproche de 2, 1 ou 0 unités de son but avec les probabilités respectives  $p$ ,  $q$  ou  $r$ . Sur un trou par-m, son score est

(N'hésitez pas à prendre une calculatrice pour la partie numérique de ce problème).

égal au nombre minimum  $n$  de tours qui lui ont été nécessaires pour avancer de  $m$  unités ou plus. Elle doit s'efforcer d'obtenir un score aussi faible que possible.

- a Montrez que si Gina joue sur un trou par-4 contre quelqu'un qui fait le par (c'est-à-dire qui atteint le but en 4 coups exactement), elle gagne plus souvent qu'elle ne perd (plus précisément, la probabilité que son score soit inférieur à 4 est supérieure à la probabilité qu'il soit supérieur à 4).
- b Montrez que son score moyen sur un trou par-4 est supérieur à 4 (donc, en moyenne, elle doit perdre contre un joueur "régulier" si on compte le nombre de points, bien qu'elle gagne si on compte le nombre de victoires trou par trou).

### Problèmes d'examen

35 Un dé a été pipé de façon à obéir à la distribution de probabilité

$$\Pr(\square \bullet) = p_1; \quad \Pr(\bullet \square) = p_2; \quad \dots; \quad \Pr(\square \square) = p_6.$$

Soit  $S_n$  la somme des valeurs de  $n$  jets consécutifs de ce dé. Trouvez une condition nécessaire et suffisante sur la distribution pour que les deux variables aléatoires  $S_n \bmod 2$  et  $S_n \bmod 3$  soient indépendantes pour tout  $n$ .

36 Soit un dé dont les six faces sont marquées



au lieu de contenir les motifs habituels de  $\square \bullet$  à  $\square \square$ .

- a Montrez qu'il est possible de fabriquer un second dé à six faces tel que, si on les jette tous les deux, la distribution de probabilité de leur somme sera la même que celle correspondant à deux dés ordinaires.
- b Généralisez cela en trouvant toutes les façons possibles d'écrire des valeurs sur les  $6n$  faces de  $n$  dés pour que la distribution des sommes soit la même que celle des sommes de  $n$  dés ordinaires (chaque face doit avoir une valeur entière).
- 37 Soit  $p_n$  la probabilité qu'il faille tirer exactement  $n$  fois à pile ou face pour obtenir deux piles à la suite, et soit  $q_n = \sum_{k \geq n} p_k$ . Trouvez des formes closes pour  $p_n$  et  $q_n$  en fonction des nombres de Fibonacci.
- 38 Quelle est la fonction génératrice de probabilité du nombre de lancers nécessaires pour voir les six faces d'un dé (non pipé) ? Généralisez à un dé à  $m$  faces : donnez des formes closes pour la moyenne et la variance du nombre de lancers nécessaires pour voir  $l$  faces sur les  $m$ . Quelle est la probabilité que ce nombre soit exactement  $n$  ?

- 39** On appelle *fonction génératrice de probabilité de Dirichlet* une série de la forme

$$P(z) = \sum_{n \geq 1} \frac{p_n}{n^z}.$$

Ainsi,  $P(0) = 1$ . Soit  $X$  une variable aléatoire telle que  $\Pr(X = n) = p_n$ . Exprimez  $E(X)$ ,  $V(X)$  et  $E(\ln X)$  en fonction de  $P(z)$  et de ses dérivées.

- 40** Le  $m$ ième cumulant  $\kappa_m$  de la loi binomiale (8.57) est de la forme  $nf_m(p)$ , où  $f_m$  est un polynôme de degré  $m$  (par exemple,  $f_1(p) = p$  et  $f_2(p) = p - p^2$ , car la moyenne est la variance sont respectivement égales à  $np$  et  $npq$ ).
- Trouvez une forme close pour le coefficient de  $p^k$  dans  $f_m(p)$ .
  - Prouvez que  $f_m\left(\frac{1}{2}\right) = (2^m - 1)B_m/m + [m = 1]$ , où  $B_m$  est le  $m$ ième nombre de Bernoulli.
- 41** Soit  $X_n$  la variable aléatoire qui représente le nombre de jets nécessaires pour obtenir un total de  $n$  piles avec une pièce non truquée. Montrez que  $E(X_{n+1}^{-1}) = (-1)^n (\ln 2 + H_{\lfloor n/2 \rfloor} - H_n)$ . Avec les méthodes du chapitre 9, donnez une estimation de cette valeur avec une erreur absolue en  $O(n^{-3})$ .
- 42** Voici la triste histoire de quelqu'un qui a des problèmes pour trouver un travail. S'il n'a pas trouvé de travail le matin, il a des chances d'être embauché l'après-midi avec une probabilité constante  $p_h$  (indépendamment de tout ce qui a pu se passer avant) ; mais s'il travaille le matin, il risque d'être mis à la porte le soir avec une probabilité  $p_f$ . En supposant que ce manège dure  $n$  jours et que notre homme a trouvé un travail pour le premier matin, calculez le nombre moyen d'après-midi à la suite durant lesquels il travaille (par exemple, si  $n = 1$  la réponse est  $1 - p_f$ ).
- 43** Trouvez une forme close pour la fgp  $G_n(z) = \sum_{k \geq 0} p_{k,n} z^k$ , où  $p_{k,n}$  désigne la probabilité qu'une permutation aléatoire de  $n$  éléments ait exactement  $k$  cycles. Calculez la moyenne et l'écart-type du nombre de cycles.
- 44** Considérons un tournoi de tennis qui se déroule classiquement de la sorte : pour  $2^n$  joueurs inscrits, on joue les  $2^{1-n}$ ièmes de finale (c'est-à-dire qu'on choisit aléatoirement et équiprobablement  $2^{n-1}$  paires de joueurs, et que les deux joueurs de chaque paire s'affrontent), puis les  $2^{2-n}$ ièmes de finale (où ne participent que les  $2^{n-1}$  vainqueurs du tour précédent), et ainsi de suite jusqu'en finale. Au  $k$ ième tour, il y a donc  $2^{n-k}$  matches qui voient s'affronter  $2^{n-k+1}$  joueurs en tout, et la finale correspond au  $n$ ième tour. En fait, il existe une hiérarchie des joueurs :  $x_1$  est le meilleurs de tous,  $x_2$  est le second, ...,  $x_{2^n}$  est le

*Bizarres, ces joueurs de tennis.*

moins bon de tous. Si  $j < k$ , lorsque  $x_j$  joue contre  $x_k$ ,  $x_j$  gagne avec probabilité  $p$  et perd avec probabilité  $1 - p$ , indépendamment de tous les autres matches. On suppose que c'est la même probabilité  $p$  qui s'applique à tous les  $j$  et  $k$  possibles.

- a Quelle est la probabilité que  $x_1$  gagne le tournoi ?
  - b Quelle est la probabilité que la finale voie s'affronter les deux meilleurs joueurs,  $x_1$  et  $x_2$  ?
  - c Quelle est la probabilité que ce soient les  $2^k$  meilleurs joueurs qui participent au  $k$ ième tour avant la finale ? (Les deux premières questions concernaient les cas  $k = 0$  et  $k = 1$ ).
  - d Soit  $N(n)$  le nombre de déroulements différents possibles du tournoi (on considère que deux déroulements sont identiques si leurs matches s'effectuent entre les mêmes joueurs et ont les mêmes vainqueurs). Montrez que  $N(n) = 2^n!$ .
  - e Quelle est la probabilité que  $x_2$  gagne le tournoi ?
  - f Montrez que si  $\frac{1}{2} < p < 1$ , alors la probabilité que  $x_j$  gagne le tournoi est strictement supérieure à la probabilité que  $x_{j+1}$  gagne, pour tout  $1 \leq j < 2^n$ .
- 45 Le xérès est fabriqué en Espagne selon un procédé particulier, appelé la "Solera". Pour simplifier, nous supposerons que le fabricant ne possède que trois tonneaux A, B et C de même capacité. Chaque année, on met en bouteilles un tiers du vin contenu dans le tonneau C, et on le remplace par du vin de B ; puis on remplit B avec du vin de A et on remplit A avec du vin nouveau. Soient  $A(z)$ ,  $B(z)$  et  $C(z)$  les fonctions génératrices de probabilité telles que le coefficient de  $z^n$  est égal à la fraction de vin âgé de  $n$  ans dans le tonneau correspondant juste après que tous les transvasements ont été faits.
- a En supposant que cette opération est effectuée depuis des temps immémoriaux, de sorte qu'on se trouve dans un état stationnaire dans lequel  $A(z)$ ,  $B(z)$  et  $C(z)$  ne sont pas modifiées au cours des ans, donnez des formes closes pour ces trois fonctions génératrices.
  - b Sous la même hypothèse, trouvez la moyenne et l'écart-type de l'âge du vin dans chaque tonneau. Quel est l'âge moyen du vin lorsqu'il est mis en bouteilles ? Quelle fraction de ce vin est âgée de 25 ans exactement ?
  - c Supposons maintenant que le temps est un paramètre fini et que, à l'année 0, les trois tonneaux ont été remplis de vin nouveau. Quel est l'âge moyen du vin mis en bouteilles au début de l'année  $n$  ?
- 46 Stefan Banach avait toujours sur lui deux boîtes d'allumettes, contenant au départ  $n$  allumettes chacune. Lorsqu'il devait en allumer une, il choisissait une boîte au hasard avec probabilité  $\frac{1}{2}$ , indépendamment de ses choix précédents. Après avoir pris l'allumette, il remettait la

*"Une rapide opération arithmétique montre que, grâce à cette ingénieuse cascade, les xérès ont toujours au moins trois ans. Pousser plus loin le calcul de leur âge donne le vertige."*

— Revue du vin de France (Nov 1984)

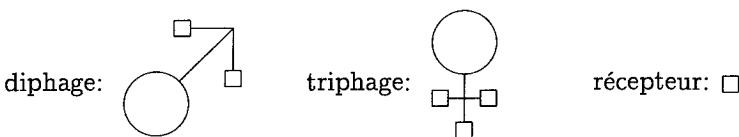
boîte dans sa poche, même si elle était vide (ce n'était pas le seul mathématicien à faire cela). Si la boîte qu'il avait choisie était vide (avant d'y prendre une allumette), il la jetait et utilisait l'autre boîte.

- a Un jour, il s'aperçut que l'autre boîte était vide aussi. Quelle est la probabilité que cela arrive ? (Pour  $n = 1$ , cela arrive la moitié du temps, pour  $n = 2$  cela arrive les  $3/8$ ièmes du temps). Répondre à cette question en trouvant une forme close de la fonction génératrice

$$P(w, z) = \sum_{m,n} p_{m,n} w^m z^n,$$

où  $p_{m,n}$  désigne la probabilité que, en partant de  $m$  allumettes dans une boîte et  $n$  dans l'autre, les deux boîtes soient vides lorsque la première choisie est vide. Trouvez ensuite une forme close pour  $p_{n,n}$ .

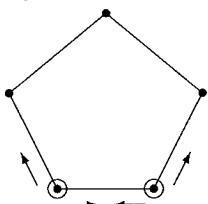
- b Généralisez votre réponse à la question (a) en donnant une forme close de la probabilité que la seconde boîte contienne exactement  $k$  allumettes au moment où on jette la première parce qu'elle est vide.
- c Trouvez une forme close du nombre moyen d'allumettes contenues dans la seconde boîte de la question précédente. ainsi que dans la première.
- 47 Des chercheurs ont récemment découvert une espèce de microbes qui se reproduisent de façon étrange. Le mâle, appelé *diphage*, possède deux récepteurs, et la femelle, appelée *triphage*, en a trois :



Lorsqu'on irradie une culture de diphages et triphages avec une particule psi, cette particule est absorbée par exactement un récepteur de l'un des phages ; tous les récepteurs ont même probabilité d'absorber la particule. Si le phage en question était un diphage, il se transforme en triphage ; si c'était un triphage, il se divise en deux diphages. Ainsi, si on commence l'expérience avec un diphage seulement, la première particule le transforme en triphage, la seconde le divise en deux diphages et la troisième transforme l'un des deux en triphage. La quatrième particule peut être absorbée par le diphage ou par le triphage ; on obtient donc soit deux triphages (avec probabilité  $\frac{2}{5}$ ) soit trois diphages (avec probabilité  $\frac{3}{5}$ ). Trouvez une forme close pour le nombre moyen de diphages obtenus en partant d'un unique diphage et en lançant successivement  $n$  particules psi.

- 48** Cinq personnes, placées sur les sommets d'un pentagone, jouent à s'envoyer des frisbees.

Ou, si ça se passe  
à Arlington, ils  
jouent à s'envoyer  
des missiles.



Ils ont deux frisbees, situés au départ sur des sommets adjacents, comme indiqué sur la figure. A chaque étape du jeu, chacun des frisbees est envoyé vers l'un des deux sommets adjacents avec une égale probabilité. Le jeu s'arrête lorsqu'une même personne reçoit les deux frisbees en même temps.

- a Trouvez la moyenne et la variance de la durée du jeu (en nombre d'étapes).
- b Trouvez une forme close, en fonction des nombres de Fibonacci, pour la probabilité que le jeu dure plus de 100 étapes.
- 49** Luc Snowwalker est en vacances dans son chalet à la montagne. Il possède  $m + n$  paires de bottes, dont  $m$  sont rangées à côté de la porte de devant et  $n$  à côté de la porte de derrière. Chaque fois qu'il veut aller faire un tour dehors, il tire à pile ou face (avec une pièce non truquée) pour choisir la porte qu'il empruntera, il chausse une paire de bottes à côté de cette porte, puis il sort. Lorsqu'il revient, il entre par devant ou par derrière, avec probabilité  $\frac{1}{2}$ , puis il range ses bottes à côté de la porte par laquelle il est entré. Ainsi, après une promenade, il y aura  $m + [-1, 0, \text{ ou } +1]$  paires de bottes devant et  $n - [-1, 0, \text{ ou } +1]$  paires de bottes derrière. Si toutes les bottes se trouvent d'un seul côté et qu'il décide de sortir par l'autre, il sort en chaussettes, se gèle les pieds, et arrête là ses vacances. En supposant que ses vacances ne s'arrêtent qu'à cette condition, soit  $P_N(m, n)$  la probabilité qu'il fasse exactement  $N$  promenades sans se geler les pieds, en partant avec  $m$  paires de bottes devant et  $n$  paires derrière. Ainsi si  $m$  et  $n$  sont strictement positifs,

$$\begin{aligned} P_N(m, n) = & \frac{1}{4}P_{N-1}(m-1, n+1) + \frac{1}{2}P_{N-1}(m, n) \\ & + \frac{1}{4}P_{N-1}(m+1, n-1), \end{aligned}$$

car les possibilités de cette première promenade sont devant/derrière, devant/devant, derrière/derrière, derrière/devant, chacune avec probabilité  $\frac{1}{4}$ , et il reste  $N - 1$  promenades à faire.

- a Complétez la récurrence en trouvant des formules valables pour  $m = 0$  ou  $n = 0$ . Utilisez la pour obtenir des équations concernant

les fonctions génératrices de probabilité

$$g_{m,n}(z) = \sum_{N \geq 0} P_N(m, n) z^N.$$

- b Dérivez ces équations et posez  $z = 1$  pour obtenir des relations concernant les  $g'_{m,n}(1)$ . Résolvez ces équations pour déterminer le nombre moyen de promenades avant les engelures.
- c Montrez que  $g_{m,n}$  admet une forme close si on pose  $z = 1/\cos^2 \theta$  :

$$g_{m,n}\left(\frac{1}{\cos^2 \theta}\right) = \frac{\sin(2m+1)\theta + \sin(2n+1)\theta}{\sin(2m+2n+2)\theta} \cos \theta.$$

- 50 Soit la fonction

$$H(z) = 1 + \frac{1-z}{2z} (z - 3 + \sqrt{(1-z)(9-z)}).$$

Le but de ce problème est de prouver que  $H(z) = \sum_{k \geq 0} h_k z^k$  est une fonction génératrice de probabilité et de trouver quelques-unes de ses propriétés de base.

- a Soit  $(1-z)^{3/2}(9-z)^{1/2} = \sum_{k \geq 0} c_k z^k$ . Prouvez que  $c_0 = 3$ ,  $c_1 = -14/3$ ,  $c_2 = 37/27$ , et  $c_{3+l} = 3 \sum_k \binom{l}{k} \binom{1/2}{3+k} \left(\frac{8}{9}\right)^{k+3}$  pour tout  $l \geq 0$ . *Suggestion* : utilisez l'identité

$$(9-z)^{1/2} = 3(1-z)^{1/2} \left(1 + \frac{8}{9}z/(1-z)\right)^{1/2}$$

et développez le dernier facteur en puissances de  $z/(1-z)$ .

- b Utilisez la question (a) et l'exercice 5.81 pour montrer que les coefficients de  $H(z)$  sont tous strictement positifs.
- c Démontrez la surprenante identité

$$\sqrt{\frac{9-H(z)}{1-H(z)}} = \sqrt{\frac{9-z}{1-z}} + 2.$$

- d Calculez la moyenne et la variance de  $H$ .

- 51 La loterie de l'état d'El Dorado est basée sur la distribution  $H$  définie dans le problème précédent. Chaque ticket coûte 1 doublon, et la probabilité de gagner  $k$  doublons est égale à  $h_k$ . Les chances de gagner avec un ticket donné sont totalement indépendantes des chances de gagner avec d'autres tickets ; en d'autres termes, le fait de gagner ou de perdre avec un ticket n'affecte aucunement les probabilités de gagner ou de perdre avec d'autres tickets achetés pour le même tirage.
- a Supposez que vous commencez à jouer avec un doublon. Si vous gagnez  $k$  doublons, vous achetez  $k$  tickets au second tirage ; puis vous misez de nouveau tout ce que vous avez gagné au second tirage

sur le troisième, et ainsi de suite. Si, à un moment donné, aucun de vos tickets d'est gagnant, vous êtes ruiné et vous ne pouvez plus jouer. Montrez que la fgp de la somme que vous possédez après  $n$  tours de jeu est égale à

$$1 - \frac{4}{\sqrt{(9-z)/(1-z) + 2n-1}} + \frac{4}{\sqrt{(9-z)/(1-z) + 2n+1}}.$$

- b Soit  $g_n$  la probabilité de perdre tout votre argent au  $n$ ième tour, et soit  $G(z) = g_1z + g_2z^2 + \dots$ . Montrez que  $G(1) = 1$  (cela signifie que vous êtes sûr de perdre tôt ou tard. Calculez la moyenne et la variance de  $G$ .
- c Quelle est le nombre moyen de tickets que vous aurez achetés avant d'être ruiné ?
- d Combien de tours pourriez-vous jouer en moyenne si vous partiez avec deux doublons au lieu d'un seul ?

*Un double doublon.*

### Questions subsidiaires

- 52 Montrez que, lorsque l'espace de probabilité considéré est fini, les définitions de la médiane et du mode d'une variable aléatoire données dans le chapitre sont très étroitement liées aux définitions de la médiane et du mode d'une suite de nombres.
- 53 Prouvez ou réfutez : si  $X$ ,  $Y$  et  $Z$  sont des variables aléatoires telles que  $X$  et  $Y$  sont indépendantes,  $X$  et  $Z$  sont indépendantes, et  $Y$  et  $Z$  sont indépendantes, alors  $X + Y$  et  $Z$  sont indépendantes.
- 54 L'équation (8.20) prouve que la valeur moyenne de  $\hat{V}X$  est  $VX$ . Que vaut la *variance* de  $\hat{V}X$  ?
- 55 Un jeu de cartes ordinaire contient 52 cartes, quatre par valeur de l'ensemble  $\{A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}$ . Soient  $X$  et  $Y$  les valeur respectives de la première et de la dernière carte du paquet. Considérons l'algorithme suivant pour battre les cartes :
- S1 Mélanger les cartes au hasard pour que chaque permutation possible apparaisse avec la probabilité  $1/52!$ .
  - S2 Si  $X \neq Y$ , lancer une pièce qui donne pile avec probabilité  $p$  (éventuellement différente de  $\frac{1}{2}$ ) et revenir en S1 si on obtient pile. Sinon, arrêter le processus.
- Pour quelle valeur de  $p$  les variables aléatoires  $X$  et  $Y$  sont-elles indépendantes à la fin du processus ?
- 56 Généralisez le problème du frisbee de l'exercice 48, en passant du pentagone au polygone à  $n$  côtés. Que valent la moyenne et la variance du nombre de lancers sans collision si les deux frisbees se trouvent au départ sur deux sommets adjacents ? Montrez que, si  $m$  est impair, la fgp

du nombre de lancers peut s'écrire comme un produit de distributions "pile ou face" :

$$G_m(z) = \prod_{k=1}^{(m-1)/2} \frac{p_k z}{1 - q_k z},$$

avec  $p_k = \sin^2 \frac{(2k-1)\pi}{2m}$ ,  $q_k = \cos^2 \frac{(2k-1)\pi}{2m}$ .

*Suggestion* : essayez de poser  $z = 1/\cos^2 \theta$ .

- 57 Montrez que si on utilise une pièce non truquée et si  $l \geq 3$ , alors le motif  $\tau_1\tau_2\dots\tau_{l-1}\tau_l$  du jeu de Penney est toujours inférieur au motif  $\bar{\tau}_2\tau_1\tau_2\dots\tau_{l-1}$ .
- 58 Existe-t-il une suite  $A = \tau_1\tau_2\dots\tau_{l-1}\tau_l$  de  $l \geq 3$  piles et faces telle que les suites  $P\tau_1\tau_2\dots\tau_{l-1}$  et  $F\tau_1\tau_2\dots\tau_{l-1}$  se comportent aussi bien si on les oppose à  $A$  dans le jeu de Penney ?
- 59 Existe-t-il des suites  $A$  et  $B$  de piles et faces telles que  $A$  soit plus longue que  $B$ , mais que  $A$  apparaisse avant  $B$  plus de la moitié du temps lorsqu'on tire à pile ou face avec une pièce juste ?
- 60 Soient  $k$  et  $n$  deux entiers strictement positifs tels que  $k < n$ .
- a Trouvez une forme close pour la fonction génératrice de probabilité

$$G(w, z) = \frac{1}{m^n} \sum_{h_1=1}^m \dots \sum_{h_n=1}^m w^{P(h_1, \dots, h_n; k)} z^{P(h_1, \dots, h_n; n)}$$

de la distribution du nombre de tests nécessaires pour trouver les  $k$ ième et  $n$ ième enregistrements qui ont été insérés dans une table de hachage composée de  $m$  listes.

- b Montrez que, bien que les variables aléatoires  $P(h_1, \dots, h_n; k)$  et  $P(h_1, \dots, h_n; n)$  ne soient pas indépendantes, elles satisfont

$$\begin{aligned} E(P(h_1, \dots, h_n; k)P(h_1, \dots, h_n; n)) \\ = (EP(h_1, \dots, h_n; k))(EP(h_1, \dots, h_n; n)). \end{aligned}$$

- 61 Utilisez le résultat de l'exercice précédent pour prouver (8.104).
- 62 Pour continuer l'exercice 47, calculez la *variance* du nombre de di-phages après  $n$  irradiations.

#### *Sujet de recherche*

- 63 La *loi normale* est une distribution de probabilité non discrète, caractérisée par le fait que tous ses cumulants sont nuls sauf la moyenne et la variance. Y a-t-il une façon simple de voir si une suite de cumulants  $(\kappa_1, \kappa_2, \kappa_3, \dots)$  provient d'une distribution *discrète* ?

# 9

## Calcul asymptotique

Trouver une réponse exacte à un problème est une expérience particulièrement réjouissante ; ce n'est cependant pas une raison pour négliger l'utilité de l'approximation. Lorsqu'on aboutit à une somme ou à une récurrence dont la solution n'admet apparemment pas de forme close, on peut raisonnablement avoir envie d'approcher autant que possible cette solution. De plus, même si on connaît une forme close, celle-ci ne nous dit pas tout ; en général, on ne sait même pas la comparer à d'autres formes closes. Par exemple, il n'existe (apparemment) pas de forme close pour la somme

$$S_n = \sum_{k=0}^n \binom{3n}{k}.$$

Il est toutefois bon de savoir que

$$S_n \sim 2 \binom{3n}{n}, \quad \text{lorsque } n \rightarrow \infty.$$

*Voilà, le mot est lâché.*

On dit que la somme “tend asymptotiquement vers” ou “est asymptotiquement égale à”  $2 \binom{3n}{n}$ . On peut obtenir plus d'information, comme la formule

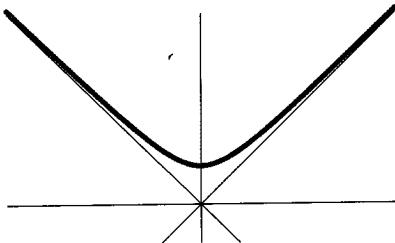
$$S_n = \binom{3n}{n} \left( 2 - \frac{4}{n} + O\left(\frac{1}{n^2}\right) \right), \tag{9.1}$$

qui nous fournit une “erreur relative d'ordre  $1/n^2$ ”. Cependant, ceci ne nous permet toujours pas d'effectuer des comparaisons. Notre somme  $S_n$  est-elle plus grande ou plus petite que le nombre de Fibonacci  $F_{4n}$  ? Voici la réponse : lorsque  $n = 2$ , on a  $S_2 = 22 > F_8 = 21$ , mais  $F_{4n}$  finit par être le plus grand des deux nombres, car  $F_{4n} \sim \phi^{4n}/\sqrt{5}$  et  $\phi^4 \approx 6,8541$ , alors que

$$S_n = \sqrt{\frac{3}{\pi n}} (6.75)^n \left( 1 - \frac{151}{72n} + O\left(\frac{1}{n^2}\right) \right). \tag{9.2}$$

Notre but, dans ce chapitre, sera d'apprendre à comprendre et à obtenir, sans grandes difficultés, des résultats de ce genre.

Etymologiquement, le mot *asymptotique* est d'origine grecque ; il signifie littéralement "qui ne tombe pas avec". Dans leur étude des coniques, les mathématiciens grecs de l'Antiquité considéraient notamment les hyperboles. Celle de la figure qui suit, d'équation  $y = \sqrt{1 + x^2}$ ,



admet deux "asymptotes", les droites d'équations  $y = x$  et  $y = -x$ . Cela signifie que, lorsque  $x \rightarrow \infty$ , la courbe se rapproche de ces droites sans jamais les toucher. Aujourd'hui, on utilise le terme "asymptotique" dans un sens plus large, pour qualifier une valeur approchée qui s'approche de plus en plus de la valeur véritable lorsqu'un paramètre donné tend vers une certaine limite.

Notre but n'est pas d'apprendre à calculer des formules asymptotiques extrêmement difficiles ; nous nous contenterons d'acquérir les bases nécessaires pour bien comprendre le sujet et être capables d'aller plus loin ultérieurement. Nous nous attacherons particulièrement à bien comprendre la signification des symboles comme " $\sim$ " et " $O$ ", et nous étudierons les techniques de base pour manipuler les valeurs asymptotiques.

## 9.1 UNE HIÉRARCHIE DE FONCTIONS

On peut comparer deux fonctions d'une variable  $n$  en regardant le "rapport asymptotique de leurs croissances". Il n'est pas rare que l'une des deux s'approche plus vite de l'infini que l'autre. On formalise cela en posant

$$f(n) \prec g(n) \iff \lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0. \quad (9.3)$$

Cette relation est transitive : si  $f(n) \prec g(n)$  et  $g(n) \prec h(n)$ , alors  $f(n) \prec h(n)$ . On peut aussi écrire  $g(n) \succ f(n)$  au lieu de  $f(n) \prec g(n)$ . cette notation fut introduite en 1871 par Paul du Bois-Reymond [85].

Par exemple, il est facile de voir que  $n \prec n^2$  ; ainsi, on peut dire que "n croît plus lentement que  $n^2$ ". En fait, pour tous réels  $\alpha$  et  $\beta$ , on a

$$n^\alpha \prec n^\beta \iff \alpha < \beta. \quad (9.4)$$

*Le mot "sympotome" a la même étymologie.*

Bien évidemment, il existe d'autres fonctions que les puissances de  $n$ . La relation  $\prec$  permet de les ordonner selon leur comportement asymptotique. On peut ainsi obtenir des suites du genre de

$$1 \prec \log \log n \prec \log n \prec n^\epsilon \prec n^c \prec n^{\log n} \prec c^n \prec n^n \prec c^{c^n},$$

où  $\epsilon$  et  $c$  sont des constantes telles que  $0 < \epsilon < 1 < c$ .

Toutes les fonctions listées ci-dessus, sauf 1, tendent vers l'infini lorsque  $n$  tend vers l'infini. Pour placer une nouvelle fonction dans cette hiérarchie, il ne faut donc pas se demander si elle croît vers l'infini, mais à quelle vitesse elle croît vers l'infini.

Dans le domaine de l'analyse asymptotique, il ne faut pas avoir peur de "penser grand" lorsqu'on se représente une variable qui tend vers l'infini. Par exemple, d'après notre hiérarchie,  $\log n \prec n^{0.0001}$ ; ceci semble inexact si on limite notre horizon à des "petits nombres" comme un gogol (un gogol vaut  $10^{100}$ ); si on prend  $n = 10^{100}$ , alors  $\log n = 100$ , tandis que  $n^{0.0001}$  ne vaut que  $10^{0.01} \approx 1,0233$ . Par contre, si on va jusqu'au gogolplex,  $n = 10^{10^{100}}$ , alors  $\log n = 10^{100}$  semble ridiculement petit par rapport à  $n^{0.0001} = 10^{10^{96}}$ .

Même si on choisit un nombre  $\epsilon$  extrêmement petit (par exemple plus petit que  $1/10^{10^{100}}$ ), on pourra toujours trouver un nombre  $n$  à partir duquel  $\log n$  est bien plus petit que  $n^\epsilon$ . Il suffit pour cela de poser  $n = 10^{10^{2k}}$ , avec  $k$  assez grand pour que  $\epsilon \geq 10^{-k}$ ; alors  $\log n = 10^{2k}$  et  $n^\epsilon \geq 10^{10^k}$ , donc le rapport  $(\log n)/n^\epsilon$  tend vers zéro lorsque  $n \rightarrow \infty$ .

La hiérarchie ci-dessus ne concerne que les fonctions qui croissent vers l'infini. Il est bon de disposer aussi d'une hiérarchie similaire pour les fonctions qui décroissent vers zéro. Il suffit pour cela de considérer les inverses : si  $f(n)$  et  $g(n)$  ne s'annulent jamais, alors

$$f(n) \prec g(n) \iff \frac{1}{g(n)} \prec \frac{1}{f(n)}. \quad (9.5)$$

Ainsi, par exemple, les fonctions suivantes (à part 1) décroissent toutes vers zéro :

$$\frac{1}{c^{c^n}} \prec \frac{1}{n^n} \prec \frac{1}{c^n} \prec \frac{1}{n^{\log n}} \prec \frac{1}{n^c} \prec \frac{1}{n^\epsilon} \prec \frac{1}{\log n} \prec \frac{1}{\log \log n} \prec 1.$$

Maintenant, examinons quelques fonctions que nous connaissons et voyons où nous pouvons les placer. Nous savons que le nombre  $\pi(n)$  de nombres premiers inférieurs ou égaux à  $n$  est à peu près égal à  $n/\ln n$ . Comme  $1/n^\epsilon \prec 1/\ln n \prec 1$ , il suffit de multiplier par  $n$  pour trouver

$$n^{1-\epsilon} \prec \pi(n) \prec n.$$

On peut même généraliser (9.4) si on remarque que, par exemple,

$$\begin{aligned} n^{\alpha_1} (\log n)^{\alpha_2} (\log \log n)^{\alpha_3} &\prec n^{\beta_1} (\log n)^{\beta_2} (\log \log n)^{\beta_3} \\ \iff (\alpha_1, \alpha_2, \alpha_3) &< (\beta_1, \beta_2, \beta_3). \end{aligned} \quad (9.6)$$

Ici, le signe “ $<$ ” désigne l’ordre lexicographique (celui du dictionnaire) : on dit que  $(\alpha_1, \alpha_2, \alpha_3) < (\beta_1, \beta_2, \beta_3)$  si  $\alpha_1 < \beta_1$ , ou  $\alpha_1 = \beta_1$  et  $\alpha_2 < \beta_2$ , ou  $\alpha_1 = \beta_1$  et  $\alpha_2 = \beta_2$  et  $\alpha_3 < \beta_3$ .

Passons à la fonction  $e^{\sqrt{\log n}}$  ; comment se place-t-elle dans cette hiérarchie ? On peut répondre aux questions de ce genre avec la règle

$$e^{f(n)} \prec e^{g(n)} \iff \lim_{n \rightarrow \infty} (f(n) - g(n)) = -\infty, \quad (9.7)$$

qui se déduit de la définition (9.3) en appliquant le logarithme. Par conséquent,

$$1 \prec f(n) \prec g(n) \implies e^{|f(n)|} \prec e^{|g(n)|}.$$

Comme  $1 \prec \log \log n \prec \sqrt{\log n} \prec e \log n$ , nous trouvons que  $\log n \prec e^{\sqrt{\log n}} \prec n^e$ .

Lorsque deux fonctions  $f(n)$  et  $g(n)$  croissent à la même vitesse, on écrit “ $f(n) \asymp g(n)$ ”. Voici la définition formelle de cette notation :

$$\begin{aligned} f(n) \asymp g(n) &\iff |f(n)| \leq C|g(n)| \text{ et } |g(n)| \leq C|f(n)|, \\ &\text{pour un certain } C \\ &\text{et pour tout } n \text{ assez grand.} \end{aligned} \quad (9.8)$$

Cette relation est vraie, par exemple, si  $f(n)$  est constante et  $g(n) = \cos n + \arctan n$ . Nous montrerons bientôt qu’elle est vraie aussi si  $f(n)$  et  $g(n)$  sont deux polynômes de même degré. Il existe aussi une relation plus forte que l’on définit ainsi :

$$f(n) \sim g(n) \iff \lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1. \quad (9.9)$$

Dans ce cas, on dit que “ $f(n)$  est asymptotiquement équivalente à  $g(n)$ ”.

G. H. Hardy [179] a introduit la classe des *fonctions logarithmico-exponentielles*, que l’on définir récursivement comme la plus petite famille  $\mathcal{L}$  de fonctions qui satisfont les propriétés suivantes :

- Pour tout réel  $\alpha$ , la fonction constante  $f(n) = \alpha$  est dans  $\mathcal{L}$ .
- La fonction identité  $f(n) = n$  est dans  $\mathcal{L}$ .
- Si  $f(n)$  et  $g(n)$  sont dans  $\mathcal{L}$ , alors  $f(n) - g(n)$  y est aussi.
- Si  $f(n)$  est dans  $\mathcal{L}$ , alors  $e^{f(n)}$  y est aussi.
- Si  $f(n)$  est dans  $\mathcal{L}$  et si  $f(n)$  est “ultimement positive”, alors  $\ln f(n)$  est dans  $\mathcal{L}$ .

Une fonction  $f(n)$  est dite “ultimement positive” s'il existe un entier  $n_0$  tel que  $f(n) > 0$  pour tout  $n \geq n_0$ .

On peut déduire de ces règles que, par exemple, si  $f(n)$  et  $g(n)$  sont dans  $\mathcal{L}$ , alors  $f(n) + g(n)$  y est aussi, car  $f(n) + g(n) = f(n) - (0 - g(n))$ . Si  $f(n)$  et  $g(n)$  sont ultimement positives et appartiennent à  $\mathcal{L}$ , alors leur produit  $f(n)g(n) = e^{\ln f(n)+\ln g(n)}$  et leur quotient  $f(n)/g(n) = e^{\ln f(n)-\ln g(n)}$  sont aussi dans  $\mathcal{L}$ ; c'est aussi le cas de fonctions comme  $\sqrt{f(n)} = e^{\frac{1}{2}\ln f(n)}$ , etc. Hardy a démontré que toute fonction logarithmico-exponentielle est soit ultimement positive, soit ultimement négative, soit identiquement nulle. Par conséquent, le produit et le quotient de deux fonctions de  $\mathcal{L}$  est toujours dans  $\mathcal{L}$ , sauf si on divise par une fonction identiquement nulle.

Le résultat principal de Hardy concernant ces fonctions logarithmico-exponentielles est le fait qu'elles forment une hiérarchie asymptotique : *si  $f(n)$  et  $g(n)$  sont deux fonctions de  $\mathcal{L}$ , alors soit  $f(n) \prec g(n)$ , soit  $f(n) \succ g(n)$ , soit  $f(n) \asymp g(n)$ . Dans ce dernier cas, il existe une constante  $\alpha$  telle que*

$$f(n) \sim \alpha g(n).$$

Nous ne démontrerons pas ce théorème, mais il est bon de le connaître, car, en pratique, presque toutes les fonctions que nous manipulons appartiennent à  $\mathcal{L}$ . Il n'est donc pas difficile en général de trouver la place d'une fonction donnée dans une hiérarchie.

## 9.2 LA NOTATION O

C'est Paul Bachmann qui a, en 1894, introduit cette extraordinaire notation ; elle a été popularisée par Edmund Landau entre autres. Nous l'avons déjà rencontrée dans des formules du genre de

$$H_n = \ln n + \gamma + O(1/n), \quad (9.10)$$

qui nous dit que le  $n$ ième nombre harmonique est égal au logarithme népérien de  $n$  plus la constante d'Euler, plus quelque chose en “grand O de 1 sur  $n$ ”. Bien que ce quelque chose ne soit pas précisément spécifié, cette notation indique que sa valeur absolue ne dépasse pas un nombre constant de fois  $1/n$ .

Cette notation O est élégante parce qu'elle fait abstraction des détails inutiles pour nous permettre de nous concentrer sur ce qui est réellement important : la quantité  $O(1/n)$  est négligeable si on considère que les multiples constants de  $1/n$  le sont.

De plus, nous pouvons utiliser le O à l'intérieur d'une formule, ce qui n'était pas le cas pour les notations de la section 9.1. Si on veut exprimer

“... wir durch das Zeichen O( $n$ ) eine Größe ausdrücken, deren Ordnung in Bezug auf  $n$  die Ordnung von  $n$  nicht überschreitet; ob sie wirklich Glieder von der Ordnung  $n$  in sich enthält, bleibt bei dem bisherigen Schlußverfahren dahingestellt.”

—P. Bachmann [17]

(9.10) avec ces notations, on doit faire passer “ $\ln n + \gamma$ ” dans le membre gauche pour donner un résultat plus faible comme

$$H_n - \ln n - \gamma \prec \frac{\log \log n}{n}$$

ou un résultat plus fort comme

$$H_n - \ln n - \gamma \asymp \frac{1}{n}.$$

Voici un exemple qui peut aider à clarifier cette idée de quantités non précisément spécifiées : il nous est déjà arrivé d'utiliser la notation “ $\pm 1$ ” ; nous ne savons pas si cela représente  $+1$  ou  $-1$  (peut-être même que cela n'a pas d'importance) mais cela ne nous empêche pas de manipuler sans problème les formules qui contiennent cette expression.

N.G. de Bruijn introduit, au début de son livre *Asymptotic Methods in Analysis* [74] la notation grand L qui aide à comprendre le grand O. Si on écrit  $L(5)$  pour désigner un nombre de valeur absolue inférieure à 5 (sans connaître la valeur exacte de ce nombre), alors on peut effectuer certains calculs. Par exemple, on peut écrire des formules comme  $1 + L(5) = L(6)$ ,  $L(2) + L(3) = L(5)$ ,  $L(2)L(3) = L(6)$ ,  $e^{L(5)} = L(e^5)$ . Par contre, on ne peut pas écrire  $L(5) - L(3) = L(2)$ , car le membre gauche pourrait être égal à  $4 - 0$  ; nous avons seulement le droit d'écrire  $L(5) - L(3) = L(8)$ .

La notation O de Bachmann est similaire à la notation L, mais elle est moins précise encore :  $O(\alpha)$  désigne un nombre dont la valeur absolue est égale à au plus une constante multipliée par  $|\alpha|$ . Nous ne connaissons ni la valeur du nombre, ni même celle de la constante. Bien entendu, la notion de “constante” n'a aucun sens s'il n'y a rien de variable en comparaison ; c'est pourquoi la notation O n'est utilisée que lorsqu'il y a au moins une quantité (par exemple  $n$ ) qui varie. La formule

$$f(n) = O(g(n)) \quad \text{pour tout } n \tag{9.11}$$

signifie donc qu'il existe une constante C telle que

$$|f(n)| \leq C|g(n)| \quad \text{pour tout } n. \tag{9.12}$$

Lorsque  $O(g(n))$  figure à l'intérieur d'une formule, il représente une fonction  $f(n)$  qui satisfait (9.12). Bien qu'on ne connaisse pas les valeurs de  $f(n)$ , on sait qu'elles ne sont pas trop grandes. De même, le “ $L(n)$ ” de de Bruijn représente une fonction non précisée  $f(n)$  dont les valeurs satisfont  $|f(n)| < |n|$ . La principale différence entre L et O est la constante C qui est mise en œuvre dans le cas de la notation O. Chaque occurrence de O peut donner lieu à un C différent, et chaque C est indépendant de n.

Ça peut avoir un sens, mais ça ne sert à rien.

Par exemple, nous savons que la somme des  $n$  premiers carrés est

$$\square_n = \frac{1}{3}n(n + \frac{1}{2})(n + 1) = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n.$$

On peut écrire

$$\square_n = O(n^3)$$

car  $|\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n| \leq \frac{1}{3}|n|^3 + \frac{1}{2}|n|^2 + \frac{1}{6}|n| \leq \frac{1}{3}|n^3| + \frac{1}{2}|n^3| + \frac{1}{6}|n^3| = |n^3|$   
pour tout entier  $n$ . On peut aussi écrire une formule plus précise,

$$\square_n = \frac{1}{3}n^3 + O(n^2);$$

ou, au contraire, un résultat totalement imprécis, comme

$$\square_n = O(n^{10}).$$

Rien dans la définition de  $O$  ne nous oblige à donner la meilleure approximation.

Que se passe-t-il si  $n$  n'est pas un entier, par exemple si on a une formule du genre de  $S(x) = \frac{1}{3}x^3 + \frac{1}{2}x^2 + \frac{1}{6}x$ , où  $x$  est un nombre réel ? Dans ce cas, on ne peut pas dire que  $S(x) = O(x^3)$ , car le rapport  $S(x)/x^3 = \frac{1}{3} + \frac{1}{2}x^{-1} + \frac{1}{6}x^{-2}$  n'est pas borné lorsque  $x \rightarrow 0$ . On ne peut pas dire non plus que  $S(x) = O(x)$ , puisque le rapport  $S(x)/x = \frac{1}{3}x^2 + \frac{1}{2}x + \frac{1}{6}$  n'est pas borné lorsque  $x \rightarrow \infty$ . Il semblerait donc que nous ne puissions pas utiliser la notation  $O$  avec  $S(x)$ .

Il y a pourtant une solution à ce dilemme. En général, les variables sur lesquelles on applique  $O$  sont sujettes à des conditions qui limitent leur portée. Par exemple, si on stipule que  $|x| \geq 1$ , ou que  $x \geq \epsilon$  où  $\epsilon$  est une constante positive, ou encore que  $x$  est un entier, alors on peut écrire  $S(x) = O(x^3)$ ; et si on stipule que  $|x| \leq 1$ , ou que  $|x| \leq c$  où  $c$  est une constante positive, alors on peut écrire  $S(x) = O(x)$ . Ainsi, la notation  $O$  dépend étroitement de son environnement, c'est-à-dire des contraintes sur les variables mises en cause.

Ces contraintes consistent souvent en un passage à la limite. Par exemple, on peut dire que

$$f(n) = O(g(n)) \quad \text{lorsque } n \rightarrow \infty. \tag{9.13}$$

Cela signifie que l'égalité est valable lorsque  $n$  est "proche" de l'infini, sans que l'on ait besoin de préciser à quel point il en est "proche". Dans ce cas, on considère implicitement qu'il existe deux constantes  $C$  et  $n_0$ , telles que

$$|f(n)| \leq C|g(n)| \quad \text{pour tout } n \geq n_0. \tag{9.14}$$

Les valeurs de  $C$  et  $n_0$  peuvent être différentes pour chaque  $O$  ; par contre, elles sont indépendantes de  $n$ . De même, la notation

$$f(x) = O(g(x)) \quad \text{lorsque } x \rightarrow 0$$

signifie qu'il existe deux constantes  $C$  et  $\epsilon$  telles que

$$|f(x)| \leq C|g(x)| \quad \text{pour tout } |x| \leq \epsilon. \quad (9.15)$$

Les valeurs limites n'ont pas besoin d'être forcément égales à  $\infty$  ou  $0$  ; on peut aussi écrire

$$\ln z = z - 1 + O((z-1)^2) \quad \text{lorsque } z \rightarrow 1$$

car

$$|\ln z - z + 1| \leq |z-1|^2 \quad \text{lorsque } |z-1| \leq \frac{1}{2}.$$

Au fil de ces quelques pages, notre définition de  $O$ , qui semblait assez évidente au début, s'est transformée en quelque chose de bien plus compliqué. Maintenant,  $O$  représente une fonction indéfinie et, selon son environnement, une ou deux constantes, indéfinies elles aussi ; et ce n'est pas fini ! Il nous reste encore une subtilité à prendre en compte : on a tout à fait le droit d'écrire

$$\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n = O(n^3),$$

mais il ne faut *jamais* écrire la même égalité en intervertissant les membres droit et gauche. Sinon, on pourrait en déduire des résultats absolument ridicules : par exemple,  $n = n^2$  car  $n = O(n^2)$  et  $n^2 = O(n^2)$ . Lorsqu'on travaille avec la notation  $O$ , comme avec d'autres formules qui représentent des quantités non définies précisément, les égalités que l'on manipule *ne sont pas symétriques*. Le membre droit est en quelque sorte une "approximation" du membre gauche.

A strictement parler, la notation  $O(g(n))$  ne désigne pas *une* fonction  $f(n)$ , mais *l'ensemble* de toutes les fonctions  $f(n)$  telles que  $|f(n)| \leq C|g(n)|$  pour une certaine constante  $C$ . Une formule ordinaire  $g(n)$ , sans notation  $O$ , peut être vue comme l'ensemble contenant une seule fonction  $f(n) = g(n)$ . Si  $S$  et  $T$  sont deux ensembles de fonctions de  $n$ , alors  $S + T$  représente l'ensemble de toutes les fonctions de la forme  $f(n) + g(n)$ , où  $f(n) \in S$  et  $g(n) \in T$  ; on définit de même  $S - T$ ,  $ST$ ,  $S/T$ ,  $\sqrt{S}$ ,  $e^S$  et  $\ln S$ . Toute "équation" entre de tels ensembles de fonctions est en fait une *inclusion* ; dans ce contexte, " $=$ " signifie " $\subseteq$ ". Avec ces définitions, toutes nos manipulations de  $O$  reposent maintenant sur des bases solides.

*Si votre x  
s'approche  
de zéro,  
vous êtes  
le phénix  
des hôtes  
de ce O.*

*"And to avoide the  
tediouuse repetition  
of these woordes:  
is equalle to: I will  
sette as I doe often  
in woorke use, a  
paire of parallels,  
or Gemowe lines of  
one lengthe, thus:  
=====, bicaus  
noe .2. thynges, can  
be moare equalle."  
—R. Recorde [305]*

Par exemple, l’“équation”

$$\frac{1}{3}n^3 + O(n^2) = O(n^3)$$

signifie que  $S_1 \subseteq S_2$ , où  $S_1$  est l’ensemble des fonctions de la forme  $\frac{1}{3}n^3 + f_1(n)$  pour lesquelles il existe une constante  $C_1$  telle que  $|f_1(n)| \leq C_1|n^2|$ , et où  $S_2$  est l’ensemble des fonctions  $f_2(n)$  pour lesquelles il existe une constante  $C_2$  telle que  $|f_2(n)| \leq C_2|n^3|$ . Pour prouver formellement cette “équation”, on peut considérer un élément quelconque du membre gauche et prouver qu’il appartient au membre droit : étant donnée l’expression  $\frac{1}{3}n^3 + f_1(n)$  telle que  $|f_1(n)| \leq C_1|n^2|$ , il nous faut montrer qu’il existe une constante  $C_2$  telle que  $|\frac{1}{3}n^3 + f_1(n)| \leq C_2|n^3|$ . La valeur  $C_2 = \frac{1}{3} + C_1$  convient car  $n^2 \leq |n^3|$  pour tout entier  $n$ .

Si “=” signifie en réalité “ $\subseteq$ ”, pourquoi n’écrivons-nous pas “ $\subseteq$ ” au lieu d’abuser du signe d’égalité ? Il y a quatre raisons à cela.

La première est la tradition. Au départ, les théoriciens des nombres ont utilisé le signe d’égalité avec la notation O, et la pratique est restée. Elle est tellement répandue aujourd’hui qu’on ne peut pas espérer la faire abandonner par la communauté mathématique.

La seconde est la tradition. Les informaticiens ont l’habitude d’abuser de la notation “=”. Depuis des années, les programmeurs en FORTRAN, en BASIC et en langage C écrivent sans sourciller des instructions comme “ $N = N + 1$ ”. Ce n’est pas un abus de plus qui va les déranger.

La troisième est la tradition. Lorsqu’on lit le signe “=”, on dit souvent “est”. Par exemple, la formule  $H_n = O(\log n)$  se lit “ $H_n$  est un grand O de  $\log n$ ”. En français, ce “est” est à sens unique ; on dit qu’un oiseau est un animal, mais pas qu’un animal est un oiseau.

La quatrième est que cette convention est tout à fait naturelle pour notre propos. Si nous limitions l’utilisation de la notation O aux cas où le O occupe tout le membre droit d’une formule, comme pour l’approximation du nombre harmonique  $H_n = O(\log n)$  ou pour le temps d’exécution d’un algorithme de tri  $T(n) = O(n \log n)$ , le fait d’utiliser “=” ou quoi que ce soit d’autre n’aurait aucune importance. Par contre, lorsqu’on utilise la notation O à l’intérieur d’une expression, il est bien agréable de s’imaginer intuitivement le signe “=” comme une égalité, tout comme il est agréable de considérer  $O(1/n)$  comme une très petite quantité.

C’est pourquoi nous continuerons à écrire “=” et à considérer  $O(g(n))$  comme une fonction incomplètement spécifiée, tout nous réservant le droit de revenir si nécessaire aux définitions liées à la théorie des ensembles.

Il nous faut encore mentionner un petit point technique : si l’environnement est composé de plusieurs variables, la notation O représente des ensembles de fonctions à deux ou plusieurs variables. Chaque fonction porte sur l’ensemble des variables “libres”, celles qui ont la possibilité de changer

*“It is obvious that the sign = is really the wrong sign for such relations, because it suggests symmetry, and there is no such symmetry. . . Once this warning has been given, there is, however, not much harm in using the sign =, and we shall maintain it, for no other reason than that it is customary.”*

— N. G. de Bruijn [74]

de valeur. Ce concept est un peu subtil, car il peut arriver qu'une variable ne soit définie que dans certaines parties d'une expression, par exemple dans une somme. Pour illustrer cela, examinons de près l'équation

$$\sum_{k=0}^n (k^2 + O(k)) = \frac{1}{3}n^3 + O(n^2), \quad n \geq 0 \text{ entier.} \quad (9.16)$$

L'expression  $k^2 + O(k)$  du membre gauche désigne l'ensemble de toutes les fonctions de deux variables de la forme  $k^2 + f(k, n)$  pour lesquelles il existe une constante  $C$  telle que  $|f(k, n)| \leq Ck$  pour  $0 \leq k \leq n$ . La somme de cet ensemble de fonctions, pour  $0 \leq k \leq n$ , est l'ensemble des fonctions  $g(n)$  de la forme

$$\sum_{k=0}^n (k^2 + f(k, n)) = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n + f(0, n) + f(1, n) + \cdots + f(n, n),$$

où  $f$  satisfait la propriété considérée. Comme

$$\begin{aligned} & \left| \frac{1}{2}n^2 + \frac{1}{6}n + f(0, n) + f(1, n) + \cdots + f(n, n) \right| \\ & \leq \frac{1}{2}n^2 + \frac{1}{6}n^2 + C \cdot 0 + C \cdot 1 + \cdots + C \cdot n \\ & < n^2 + C(n^2 + n)/2 < (C + 1)n^2, \end{aligned}$$

toutes ces fonctions  $g(n)$  appartiennent au membre droit de (9.16) ; donc (9.16) est vraie.

Certaines personnes abusent de la notation  $O$  en supposant, à tort, qu'elle donne un ordre de grandeur exact ; elles l'utilisent pour décrire à la fois un majorant et un minorant. Elles considèrent par exemple qu'un algorithme de tri de  $n$  nombres en temps  $O(n^2)$  est inefficace. Pourtant, le fait que le temps soit en  $O(n^2)$  de l'empêche pas de pouvoir être aussi en  $O(n)$ . Pour minorer, nous avons à notre disposition une autre notation, grand Oméga :

$$f(n) = \Omega(g(n)) \iff |f(n)| \geq C|g(n)| \text{ pour un certain } C > 0. \quad (9.17)$$

On a  $f(n) = \Omega(g(n))$  si et seulement si  $g(n) = O(f(n))$ . Un algorithme de tri qui s'exécute en temps  $\Omega(n^2)$  est moins efficace qu'un algorithme qui s'exécute en temps  $O(n \log n)$ , pour  $n$  suffisamment grand.

La notation grand Thêta permet de spécifier un ordre de croissance exact :

$$f(n) = \Theta(g(n)) \iff \begin{matrix} f(n) = O(g(n)) \\ \text{et} \\ f(n) = \Omega(g(n)) \end{matrix} \quad (9.18)$$

On a  $f(n) = \Theta(g(n))$  si et seulement si  $f(n) \asymp g(n)$ , selon la notation que nous avons vue précédemment en (9.8).

*(C'est le moment de s'échauffer avec les exercices 3 et 4).*

*Comme  $\Omega$  et  $\Theta$  sont des lettres grecques majuscules, le  $O$  de la notation  $O$  est certainement un Omicron majuscule. Après tout, ce sont les grecs qui ont inventé le calcul asymptotique.*

Il existe aussi la notation “petit o”, qui a été introduite par Edmund Landau [238] :

$$\begin{aligned} f(n) &= o(g(n)) \\ \iff |f(n)| &\leq \epsilon |g(n)| \quad \text{pour tout } n \geq n_0(\epsilon) \text{ et} \\ &\quad \text{pour toute constante } \epsilon > 0. \end{aligned} \quad (9.19)$$

C'est une autre façon d'exprimer la relation  $f(n) \prec g(n)$  de (9.3). On a aussi

$$f(n) \sim g(n) \iff f(n) = g(n) + o(g(n)). \quad (9.20)$$

Le “o” est utilisé par beaucoup d'auteurs, bien qu'une expression “O” soit presque toujours préférable. Par exemple, le temps moyen d'exécution d'une certaine méthode de tri, appelée “tri à bulles”, dépend de la valeur asymptotique de la somme  $P(n) = \sum_{k=0}^n k^{n-k} k! / n!$ . Il est facile de prouver que  $P(n) \sim \sqrt{\pi n}/2$ , ce qui signifie que le rapport  $P(n)/\sqrt{\pi n}/2$  tend vers 1 lorsque  $n \rightarrow \infty$ . Cependant, pour bien comprendre le comportement de  $P(n)$  il vaut mieux considérer la *différence*  $P(n) - \sqrt{\pi n}/2$  au lieu du rapport :

$n$	$P(n)/\sqrt{\pi n}/2$	$P(n) - \sqrt{\pi n}/2$
1	0,798	-0,253
10	0,878	-0,484
20	0,904	-0,538
30	0,918	-0,561
40	0,927	-0,575
50	0,934	-0,585

Les nombres de la deuxième colonne ne sont pas très convaincants : nous sommes loin d'une démonstration spectaculaire du fait que  $P(n)/\sqrt{\pi n}/2$  tend rapidement vers 1, à supposer qu'il le fasse. En revanche, la colonne de droite montre que  $P(n)$  est très proche de  $\sqrt{\pi n}/2$ . Ainsi, le comportement de  $P(n)$  est décrit bien plus précisément avec une formule du genre de

$$P(n) = \sqrt{\pi n}/2 + O(1),$$

ou, mieux encore,

$$P(n) = \sqrt{\pi n}/2 - \frac{2}{3} + O(1/\sqrt{n}).$$

Les formules de ce genre, avec la notation O, sont plus difficiles à obtenir que les précédentes ; mais cela est largement compensé par les informations supplémentaires qu'on en retire.

Par exemple, il se trouve que beaucoup d'algorithmes de tri ont une complexité en temps de la forme  $T(n) = A n \lg n + B n + O(\log n)$ , où  $A$  et  $B$  sont des constantes qui dépendent de l'algorithme considéré. Si on arrête l'analyse au fait que  $T(n) \sim A n \lg n$ , on n'a qu'une information très partielle. De plus, choisir un algorithme en fonction de la valeur de la constante  $A$  s'avère être une mauvaise stratégie. En effet, les algorithmes pour lesquels  $A$  est petit sont souvent les même que ceux pour lesquels  $B$  est grand. Comme  $n \lg n$  ne croît qu'un tout petit peu plus vite que  $n$ , l'algorithme le plus rapide rapide asymptotiquement parlant (celui pour lequel  $A$  est le plus petit) peut très bien ne l'être en fait que pour des valeurs de  $n$  tellement grandes qu'elles n'apparaissent jamais en pratique. C'est pourquoi il faut s'intéresser à la valeur de  $B$  pour faire le bon choix.

Avant d'aller plus loin dans notre étude du  $O$ , arrêtons-nous un peu pour parler d'un petit détail de style mathématique. Nous avons utilisé dans ce chapitre trois notations différentes pour le logarithme :  $\lg$ ,  $\ln$  et  $\log$ . On utilise souvent " $\lg$ " dans des calculs liés à des méthodes informatiques, car c'est le logarithme à base 2 qui s'applique naturellement dans ces cas. En revanche, c'est " $\ln$ ", le logarithme népérien (appelé aussi logarithme naturel) qui apparaît le plus souvent dans les calculs purement mathématiques, du fait qu'il est plus simple à manipuler. Que dire alors de " $\log$ " ? N'est-ce pas le vulgaire logarithme en base 10 qu'on apprend au lycée, celui qui est justement appelé "logarithme vulgaire" ? Eh bien si. D'ailleurs, beaucoup de mathématiciens brouillent les pistes en écrivant " $\log$ " pour le logarithme népérien ou le logarithme à base deux. En fait, il n'y a pas de convention universelle à ce sujet. Toutefois, cela n'a aucune importance lorsqu'un logarithme apparaît à l'intérieur d'une notation  $O$ , car les constantes multiplicatives ne comptent pas dans ce cas. Il n'y a donc aucune différence entre  $O(\lg n)$ ,  $O(\ln n)$  et  $O(\log n)$  lorsque  $n \rightarrow \infty$  ; il en est de même pour  $O(\lg \lg n)$ ,  $O(\ln \ln n)$  et  $O(\log \log n)$ . Chacun peut choisir celui qu'il préfère. En ce qui nous concerne, nous utiliserons " $\log$ ", dans tous les cas où c'est possible sans ambiguïté, car il se prononce plus facilement.

*Remarquez que  
 $\log \log \log n$   
est indéfini  
si  $n \leq 10$ .*

### 9.3 MANIPULATION DE O

Comme tous les formalismes mathématiques, la notation  $O$  admet des règles de manipulation qui permettent de nous libérer des détails de sa définition. Une fois que ces règles sont établies et prouvées, on peut travailler à un niveau plus haut.

Par exemple, si on prouve une fois pour toutes les deux assertions

$$n^m = O(n^{m'}), \quad \text{lorsque } m \leq m', \tag{9.21}$$

$$O(f(n)) + O(g(n)) = O(|f(n)| + |g(n)|), \tag{9.22}$$

alors on peut dire immédiatement que  $\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n = O(n^3) + O(n^3) + O(n^3) = O(n^3)$ , sans en passer par les calculs laborieux de la section précédente.

Voici quelques autres règles qui découlent de la définition :

$$f(n) = O(f(n)); \quad (9.23)$$

$$c \cdot O(f(n)) = O(f(n)), \quad \text{si } c \text{ est constant}; \quad (9.24)$$

$$O(O(f(n))) = O(f(n)); \quad (9.25)$$

$$O(f(n))O(g(n)) = O(f(n)g(n)); \quad (9.26)$$

$$O(f(n)g(n)) = f(n)O(g(n)). \quad (9.27)$$

On démontre (9.22) dans l'exercice 9, et les preuves des autres formules sont similaires. Elles sont toutes indépendantes des conditions éventuelles sur la variable  $n$ .

Nous pouvons déduire des équations (9.27) et (9.23) que  $O(f(n)^2) = O(f(n))^2$ . Cette identité est parfois utile pour supprimer des parenthèses ; on peut ainsi écrire

$$O(\log n)^2 \quad \text{au lieu de} \quad O((\log n)^2).$$

Ces expressions sont toutes deux préférables à " $O(\log^2 n)$ ", notation ambiguë car certains auteurs l'utilisent pour signifier " $O(\log \log n)$ ".

A-t-on aussi le droit d'écrire

$$O(\log n)^{-1} \quad \text{au lieu de} \quad O((\log n)^{-1}) ?$$

Non ! C'est interdit car l'ensemble de fonctions  $1/O(\log n)$  ne contient pas l'ensemble  $O(1/\log n)$ , et n'en est pas un sous-ensemble non plus. Par contre, on peut remplacer  $O((\log n)^{-1})$  par  $\Omega(\log n)^{-1}$ , mais cela ne nous avance pas à grand chose. Nous imposerons donc aux exposants "à l'extérieur du O" d'être des entiers constants et strictement positifs.

Quelques unes des opérations les plus utiles nous sont fournies par les séries entières. Si la somme

$$S(z) = \sum_{n \geq 0} a_n z^n$$

converge absolument pour un certain nombre complexe  $z = z_0$ , alors

$$S(z) = O(1), \quad \text{pour tout } |z| \leq |z_0|.$$

C'est évident, parce que

$$|S(z)| \leq \sum_{n \geq 0} |a_n| |z|^n \leq \sum_{n \geq 0} |a_n| |z_0|^n = C < \infty.$$

*(Note: la formule  $O(f(n))^2$  ne désigne pas l'ensemble des fonctions  $g(n)^2$  telles que  $g(n)$  appartient à  $O(f(n))$  ; ces fonctions  $g(n)^2$  ne peuvent pas être strictement négatives, alors que l'ensemble  $O(f(n))^2$  contient des fonctions strictement négatives. Si  $S$  est un ensemble,  $S^2$  désigne l'ensemble de tous les produits  $s_1 s_2$  tels que  $s_1$  et  $s_2$  appartiennent à  $S$ , et non l'ensemble des carrés  $s^2$  tels que  $s \in S$ ).*

En particulier,  $S(z) = O(1)$  lorsque  $z \rightarrow 0$  et  $S(1/n) = O(1)$  lorsque  $n \rightarrow \infty$ , à la seule condition que  $S(z)$  converge pour au moins une valeur non nulle de  $z$ . On peut appliquer ce principe pour tronquer une série à un endroit désiré et estimer le reste à l'aide d'un  $O$ . Par exemple, on a non seulement  $S(z) = O(1)$ , mais aussi

$$\begin{aligned} S(z) &= a_0 + O(z), \\ S(z) &= a_0 + a_1 z + O(z^2), \end{aligned}$$

et ainsi de suite, du fait que

$$S(z) = \sum_{0 \leq k < m} a_k z^k + z^m \sum_{n \geq m} a_n z^{n-m}$$

et que la seconde somme, comme  $S(z)$  elle-même, converge absolument pour  $z = z_0$  et est un  $O(1)$ . La table 479 contient quelques-unes des formules asymptotiques les plus utiles ; la moitié d'entre elles sont obtenues par troncation d'une série au moyen de cette règle.

Les séries de Dirichlet, qui sont des sommes de la forme  $\sum_{k \geq 1} a_k/k^z$ , peuvent être tronquées de manière similaire : si une série de Dirichlet converge absolument lorsque  $z = z_0$ , alors on peut la tronquer n'importe où pour obtenir l'approximation

$$\sum_{1 \leq k < m} a_k/k^z + O(m^{-z}),$$

valable pour tout  $z$  tel que  $\Re z \geq \Re z_0$ . Ce principe est illustré par la formule concernant les nombres de Bernoulli dans la table 479.

*Souvenez-vous :  $\Re$  est la "partie réelle"*

Par contre, les formules asymptotiques concernant  $H_n$ ,  $n!$  et  $\pi(n)$  dans la table 479 n'ont pas été obtenues par troncation de séries ; si on les prolongeait à l'infini, elles divergeraient pour toute valeur de  $n$ . C'est particulièrement facile à voir dans le cas de  $\pi(n)$ , parce que nous avons déjà vu dans l'exemple 5 de la section 7.3 que la série  $\sum_{k \geq 0} k!/(ln n)^k$  diverge pour tout  $n$ . Cela n'empêche pas ces approximations par des séries divergentes tronquées d'être bien utiles, même si on les tronque après le premier ou le second terme.

On dit qu'une approximation asymptotique admet une *erreur absolue* en  $O(g(n))$  si elle est de la forme  $f(n) + O(g(n))$ , où  $f(n)$  ne contient pas de  $O$ . L'approximation admet une *erreur relative* en  $O(g(n))$  si elle est de la forme  $f(n)(1 + O(g(n)))$ , où  $f(n)$  ne contient pas  $O$ . Par exemple, l'approximation de  $H_n$  dans la table 479 admet une erreur absolue en  $O(n^{-6})$ , tandis que l'approximation de  $n!$  admet une erreur relative en  $O(n^{-4})$ . En fait, le membre droit de (9.29) n'est pas exactement sous la

**Table 479** Approximations asymptotiques, pour  $n \rightarrow \infty$  et  $z \rightarrow 0$ .

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} + O\left(\frac{1}{n^6}\right). \quad (9.28)$$

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} + O\left(\frac{1}{n^4}\right)\right). \quad (9.29)$$

$$B_n = 2[n \text{ even}](-1)^{n/2-1} \frac{n!}{(2\pi)^n} (1 + 2^{-n} + 3^{-n} + O(4^{-n})). \quad (9.30)$$

$$\pi(n) = \frac{n}{\ln n} + \frac{n}{(\ln n)^2} + \frac{2! n}{(\ln n)^3} + \frac{3! n}{(\ln n)^4} + O\left(\frac{n}{(\log n)^5}\right). \quad (9.31)$$

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + O(z^5). \quad (9.32)$$

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + O(z^5). \quad (9.33)$$

$$\frac{1}{1-z} = 1 + z + z^2 + z^3 + z^4 + O(z^5). \quad (9.34)$$

$$(1+z)^\alpha = 1 + \alpha z + \binom{\alpha}{2} z^2 + \binom{\alpha}{3} z^3 + \binom{\alpha}{4} z^4 + O(z^5). \quad (9.35)$$

forme requise  $f(n)(1 + O(n^{-4}))$ , mais on peut le réécrire en

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3}\right) (1 + O(n^{-4})).$$

(L'erreur relative est particulièrement adaptée au calcul d'inverses, car  $1/(1+O(\epsilon)) = 1+O(\epsilon)$ ).

L'exercice 12 est consacré à un calcul similaire. L'erreur absolue de cette approximation de  $n!$  est en  $O(n^{n-3.5} e^{-n})$ . De façon générale, l'erreur absolue d'une approximation est liée au nombre de chiffres corrects après la virgule si on ignore le terme en  $O$ , tandis que l'erreur relative correspond au nombre de "chiffres significatifs".

Les règles suivantes peuvent être démontrées par troncature de séries :

$$\ln(1+O(f(n))) = O(f(n)), \quad \text{si } f(n) \prec 1; \quad (9.36)$$

$$e^{O(f(n))} = 1 + O(f(n)), \quad \text{si } f(n) = O(1). \quad (9.37)$$

(On suppose ici que  $n \rightarrow \infty$ ; il y a aussi des formules similaires pour  $\ln(1+O(f(x)))$  et  $e^{O(f(x))}$  lorsque  $x \rightarrow 0$ ). Par exemple, soit  $\ln(1+g(n))$  une fonction appartenant au membre gauche de (9.36). Alors il existe des constantes  $C$ ,  $n_0$  et  $c$  telles que

$$|g(n)| \leq C|f(n)| \leq c < 1, \quad \text{pour tout } n \geq n_0.$$

Il s'ensuit que la somme infinie

$$\ln(1 + g(n)) = g(n) \cdot (1 - \frac{1}{2}g(n) + \frac{1}{3}g(n)^2 - \dots)$$

converge pour tout  $n \geq n_0$  et que la série entre parenthèses est majorée par la constante  $1 + \frac{1}{2}c + \frac{1}{3}c^2 + \dots$ . Nous avons ainsi prouvé (9.36), et (9.37) se démontre de la même manière. On peut combiner les équations (9.36) et (9.37) pour obtenir la formule

$$(1 + O(f(n)))^{O(g(n))} = 1 + O(f(n)g(n)), \quad \begin{array}{l} \text{si } f(n) \prec 1 \text{ et} \\ f(n)g(n) = O(1). \end{array} \quad (9.38)$$

### **Problème 1 : encore un tour de roue de la fortune.**

Tentons notre chance sur quelques problèmes de calcul asymptotique. Au chapitre 3, nous avons étudié un jeu de hasard, et nous avons déterminé le nombre de numéros gagnants ; c'était la formule (3.13) que voici :

$$W = \lfloor N/K \rfloor + \frac{1}{2}K^2 + \frac{5}{2}K - 3, \quad K = \lfloor \sqrt[3]{N} \rfloor.$$

Nous avions promis, à ce moment-là de donner une version asymptotique de  $W$  au chapitre 9. Eh bien nous y sommes. Essayons donc d'estimer  $W$  lorsque  $N \rightarrow \infty$ .

Commençons par supprimer les crochets de partie entière en remplaçant  $K$  par  $N^{1/3} + O(1)$ . Nous pouvons même aller plus loin et écrire

$$K = N^{1/3}(1 + O(N^{-1/3}));$$

c'est ce qu'on appelle "retirer la plus grande part" (nous le ferons souvent). Maintenant, d'après (9.38) et (9.26),

$$\begin{aligned} K^2 &= N^{2/3}(1 + O(N^{-1/3}))^2 \\ &= N^{2/3}(1 + O(N^{-1/3})) = N^{2/3} + O(N^{1/3}). \end{aligned}$$

De même,

$$\begin{aligned} \lfloor N/K \rfloor &= N^{1-1/3}(1 + O(N^{-1/3}))^{-1} + O(1) \\ &= N^{2/3}(1 + O(N^{-1/3})) + O(1) = N^{2/3} + O(N^{1/3}). \end{aligned}$$

Il s'ensuit que le nombre de numéros gagnants est

$$\begin{aligned} W &= N^{2/3} + O(N^{1/3}) + \frac{1}{2}(N^{2/3} + O(N^{1/3})) + O(N^{1/3}) + O(1) \\ &= \frac{3}{2}N^{2/3} + O(N^{1/3}). \end{aligned} \quad (9.39)$$

Remarquez comment l'un des termes en  $O$  a absorbé les autres ; c'est tout à fait typique, et c'est cette capacité qui rend la notation  $O$  si utile dans les formules.

**Problème 2 : perturbons la formule de Stirling.**

La formule de Stirling, qui donne une approximation de  $n!$ , est certainement la plus célèbre des formules asymptotiques. Nous la prouverons dans quelque temps ; pour l'instant, nous allons juste nous familiariser avec ses propriétés. On peut écrire une version de cette approximation sous la forme suivante :

$$n! = \sqrt{2\pi n} \left( \frac{n}{e} \right)^n \left( 1 + \frac{a}{n} + \frac{b}{n^2} + O(n^{-3}) \right), \quad \text{lorsque } n \rightarrow \infty, \quad (9.40)$$

pour certaines constantes  $a$  et  $b$ . Comme c'est vrai pour tout  $n$  assez grand, cela doit être vrai aussi lorsqu'on remplace  $n$  par  $n - 1$  :

$$\begin{aligned} (n-1)! &= \sqrt{2\pi(n-1)} \left( \frac{n-1}{e} \right)^{n-1} \\ &\times \left( 1 + \frac{a}{n-1} + \frac{b}{(n-1)^2} + O((n-1)^{-3}) \right). \end{aligned} \quad (9.41)$$

Nous savons bien sûr que  $(n-1)! = n!/n$ . Par conséquent, on doit pouvoir simplifier le membre droit de cette formule pour obtenir celui de (9.40) divisé par  $n$ .

Essayons donc de simplifier (9.41). Nous pouvons commencer par retirer la plus grande part du premier facteur :

$$\begin{aligned} \sqrt{2\pi(n-1)} &= \sqrt{2\pi n} (1 - n^{-1})^{1/2} \\ &= \sqrt{2\pi n} \left( 1 - \frac{1}{2n} - \frac{1}{8n^2} + O(n^{-3}) \right). \end{aligned}$$

Nous avons appliqué (9.35) pour trouver la seconde ligne.

De même, on a

$$\begin{aligned} \frac{a}{n-1} &= \frac{a}{n} (1 - n^{-1})^{-1} = \frac{a}{n} + \frac{a}{n^2} + O(n^{-3}); \\ \frac{b}{(n-1)^2} &= \frac{b}{n^2} (1 - n^{-1})^{-2} = \frac{b}{n^2} + O(n^{-3}); \\ O((n-1)^{-3}) &= O(n^{-3}(1 - n^{-1})^{-3}) = O(n^{-3}). \end{aligned}$$

La seule partie de (9.41) qui demande un peu de doigté, c'est le facteur  $(n-1)^{n-1}$ , qui est égal à

$$n^{n-1} (1 - n^{-1})^{n-1} = n^{n-1} (1 - n^{-1})^n (1 + n^{-1} + n^{-2} + O(n^{-3})).$$

Remarquez que nous faisons en sorte que tous nos développements aient une erreur relative en  $O(n^{-3})$  ; comme l'erreur relative d'un produit est la

somme des erreurs relatives des facteurs, tous les termes  $O(n^{-3})$  se rassembleront pour n'en faire qu'un.

Pour développer  $(1 - n^{-1})^n$ , nous allons d'abord calculer  $\ln(1 - n^{-1})$ , puis en prendre l'exponentielle,  $e^{n \ln(1 - n^{-1})}$  :

$$\begin{aligned}
 (1 - n^{-1})^n &= \exp(n \ln(1 - n^{-1})) \\
 &= \exp(n(-n^{-1} - \frac{1}{2}n^{-2} - \frac{1}{3}n^{-3} + O(n^{-4}))) \\
 &\stackrel{\approx}{=} \exp(-1 - \frac{1}{2}n^{-1} - \frac{1}{3}n^{-2} + O(n^{-3})) \\
 &= \exp(-1) \cdot \exp(-\frac{1}{2}n^{-1}) \cdot \exp(-\frac{1}{3}n^{-2}) \cdot \exp(O(n^{-3})) \\
 &= \exp(-1) \cdot (1 - \frac{1}{2}n^{-1} + \frac{1}{8}n^{-2} + O(n^{-3})) \\
 &\quad \cdot (1 - \frac{1}{3}n^{-2} + O(n^{-4})) \cdot (1 + O(n^{-3})) \\
 &= e^{-1} \left(1 - \frac{1}{2}n^{-1} - \frac{5}{24}n^{-2} + O(n^{-3})\right).
 \end{aligned}$$

Nous avons préféré écrire  $\exp z$  au lieu de  $e^z$  dans ce calcul pour que l'exposant compliqué soit sur la ligne principale de la formule au lieu d'être en exposant. Il nous a fallu développer  $\ln(1 - n^{-1})$  avec une erreur absolue en  $O(n^{-4})$  pour obtenir finalement une erreur relative en  $O(n^{-3})$ , parce que le logarithme était multiplié par  $n$ .

Le membre droit de (9.41) est maintenant réduit à  $\sqrt{2\pi n}$  fois  $n^{n-1}/e^n$  fois un produit de plusieurs facteurs :

$$\begin{aligned}
 &(1 - \frac{1}{2}n^{-1} - \frac{1}{8}n^{-2} + O(n^{-3})) \\
 &\quad \cdot (1 + n^{-1} + n^{-2} + O(n^{-3})) \\
 &\quad \cdot (1 - \frac{1}{2}n^{-1} - \frac{5}{24}n^{-2} + O(n^{-3})) \\
 &\quad \cdot (1 + an^{-1} + (a+b)n^{-2} + O(n^{-3})). 
 \end{aligned}$$

En multipliant tout cela et en rassemblant tous les termes asymptotiques dans un seul  $O(n^{-3})$ , on obtient

$$1 + an^{-1} + (a + b - \frac{1}{12})n^{-2} + O(n^{-3}).$$

En fait, ce qu'il nous faudrait, c'est  $1 + an^{-1} + bn^{-2} + O(n^{-3})$ , pour que cela concorde avec le membre droit de (9.40). Y a-t-il quelque chose qui ne va pas ? Eh bien non, tout va bien pourvu que  $a + b - \frac{1}{12} = b$ .

Ce que nous venons de faire ne prouve pas la formule de Stirling. Nous avons quand même démontré quelque chose : la formule (9.40) ne peut être vraie que si  $a = \frac{1}{12}$ . Si nous avions remplacé le  $O(n^{-3})$  de (9.40) par  $cn^{-3} + O(n^{-4})$  et effectué nos calculs avec une erreur relative en  $O(n^{-4})$ , nous aurions trouvé que  $b$  doit être égal à  $\frac{1}{288}$ , comme indiqué dans la table 479 (ce n'est pas le moyen le plus simple de trouver les valeurs de  $a$  et  $b$ , mais ça marche).

**Problème 3 : le *nième nombre premier*.**

L'équation (9.31) est une approximation asymptotique de  $\pi(n)$ , le nombre de nombres premiers inférieurs ou égaux à  $n$ . Si on remplace  $n$  par  $p = P_n$ , le *nième nombre premier*, on obtient  $\pi(p) = n$  ; par conséquent,

$$n = \frac{p}{\ln p} + O\left(\frac{p}{(\log p)^2}\right) \quad (9.42)$$

lorsque  $n \rightarrow \infty$ . Nous allons essayer de “résoudre” cette équation en  $p$  pour connaître la valeur approchée du *nième nombre premier*.

Commençons par simplifier le terme en  $O$ . Si on divise les deux membres par  $p/\ln p$ , on trouve que  $n \ln p/p \rightarrow 1$  ; donc  $p/\ln p = O(n)$  et

$$O\left(\frac{p}{(\log p)^2}\right) = O\left(\frac{n}{\log p}\right) = O\left(\frac{n}{\log n}\right).$$

(C'est parce que  $p \geq n$  que  $(\log p)^{-1} \leq (\log n)^{-1}$ ).

Continuons en intervertissant les deux membres de (9.42), sauf le terme en  $O$  qui reste à sa place. C'est tout à fait légal en vertu de la règle

$$a_n = b_n + O(f(n)) \iff b_n = a_n + O(f(n)), \quad (9.43)$$

que l'on peut prouver en multipliant les deux membres de l'une des deux équations par  $-1$ , puis en leur ajoutant  $a_n + b_n$ . Par conséquent,

$$\frac{p}{\ln p} = n + O\left(\frac{n}{\log n}\right) = n(1 + O(1/\log n)),$$

et on obtient

$$p = n \ln p (1 + O(1/\log n)), \quad (9.44)$$

une “réurrence approchée” de  $p = P_n$  en fonction de lui-même. Nous allons nous efforcer de la transformer en une “forme close approchée”. Pour cela, nous allons “déplier asymptotiquement la récurrence”.

En prenant les logarithmes des deux membres, on trouve que

$$\ln p = \ln n + \ln \ln p + O(1/\log n). \quad (9.45)$$

Nous pourrions remplacer le  $\ln p$  de (9.44) par cette valeur ; mais auparavant, nous devons nous débarrasser du  $p$  du membre droit. Or, nous ne pouvons pas le faire comme nous le faisons d'habitude pour une récurrence car aucune condition initiale n'est spécifiée dans (9.44).

Nous allons nous en sortir en commençant par prouver un résultat plus faible, à savoir que  $p = O(n^2)$ . Il suffit pour cela d'élever (9.44) au carré et de diviser par  $pn^2$ ,

$$\frac{p}{n^2} = \frac{(\ln p)^2}{p} (1 + O(1/\log n)),$$

car le membre droit tend vers zéro lorsque  $n \rightarrow \infty$ . Bien, nous savons que  $p = O(n^2)$ ; alors  $\log p = O(\log n)$  et  $\log \log p = O(\log \log n)$ . Nous pouvons donc déduire de (9.45) que

$$\ln p = \ln n + O(\log \log n).$$

De ce fait,  $\ln \ln p = \ln \ln n + O(\log \log n / \log n)$ , et (9.45) implique que

$$\ln p = \ln n + \ln \ln n + O(\log \log n / \log n).$$

Nous pouvons maintenant injecter ce résultat dans le membre droit de (9.44) pour obtenir

$$p = n \ln n + n \ln \ln n + O(n).$$

C'est une valeur approchée du  $n$ ième nombre premier.

On peut la raffiner en utilisant une meilleure approximation de  $\pi(n)$  que celle de (9.42). D'après (9.31),

$$n = \frac{p}{\ln p} + \frac{p}{(\ln p)^2} + O\left(\frac{p}{(\log p)^3}\right); \quad (9.46)$$

en procédant comme ci-dessus, on obtient la récurrence

$$p = n \ln p \left(1 + (\ln p)^{-1}\right)^{-1} \left(1 + O(1/\log n)^2\right), \quad (9.47)$$

qui admet une erreur relative en  $O(1/\log n)^2$  au lieu de  $O(1/\log n)$ . En prenant le logarithme et en travaillant avec la précision adéquate (juste ce qu'il faut), on trouve

$$\begin{aligned} \ln p &= \ln n + \ln \ln p + O(1/\log n) \\ &= \ln n \left(1 + \frac{\ln \ln p}{\ln n} + O(1/\log n)^2\right); \\ \ln \ln p &= \ln \ln n + \frac{\ln \ln n}{\ln n} + O\left(\frac{\log \log n}{\log n}\right)^2. \end{aligned}$$

Pour finir, injectons ces résultats dans (9.47). Voici la réponse que nous cherchions :

$$p_n = n \ln n + n \ln \ln n - n + n \frac{\ln \ln n}{\ln n} + O\left(\frac{n}{\log n}\right). \quad (9.48)$$

*Retournez chercher  
le papier brouillon,  
les gars.*

Par exemple, pour  $n = 10^6$ , cette approximation donne  $15631363,6 + O(n/\log n)$ , et le millionième nombre premier est 15485863. L'exercice 21 montre qu'on peut obtenir une approximation encore meilleure de  $P_n$  en partant d'une meilleure approximation de  $\pi(n)$  que celle de (9.46).

**Problème 4 : une somme tirée d'un vieil examen.**

La première année où les Mathématiques Concrites ont été enseignées à l'Université de Stanford, en 1970–1971, une partie de l'examen consistait à trouver la valeur asymptotique de la somme

$$S_n = \frac{1}{n^2+1} + \frac{1}{n^2+2} + \cdots + \frac{1}{n^2+n}, \quad (9.49)$$

avec une erreur absolue en  $O(n^{-7})$ . Supposons qu'on vienne juste de nous poser ce problème pour un examen (à la maison). Quelle est notre première réaction ?

Non, nous ne paniquons pas. Notre première réaction est de PENSER GRAND. Si on pose par exemple  $n = 10^{100}$ , on voit que la somme est composée de  $n$  termes, chacun d'entre eux étant un peu plus petit que  $1/n^2$ ; par conséquent, le résultat est légèrement inférieur à  $1/n$ . Il n'est pas inutile de faire ainsi une première approximation “à vue de nez” lorsqu'on attaque un problème de calcul asymptotique.

Essayons d'améliorer cette première approximation en retirant la plus grande part de chaque terme. Ainsi,

$$\frac{1}{n^2+k} = \frac{1}{n^2(1+k/n^2)} = \frac{1}{n^2} \left( 1 - \frac{k}{n^2} + \frac{k^2}{n^4} - \frac{k^3}{n^6} + O\left(\frac{k^4}{n^8}\right) \right),$$

et, en sommant toutes ces approximations, on a

$$\frac{1}{n^2+1} = \frac{1}{n^2} - \frac{1}{n^4} + \frac{1^2}{n^6} - \frac{1^3}{n^8} + O\left(\frac{1^4}{n^{10}}\right)$$

$$\frac{1}{n^2+2} = \frac{1}{n^2} - \frac{2}{n^4} + \frac{2^2}{n^6} - \frac{2^3}{n^8} + O\left(\frac{2^4}{n^{10}}\right)$$

⋮

$$\frac{1}{n^2+n} = \frac{1}{n^2} - \frac{n}{n^4} + \frac{n^2}{n^6} - \frac{n^3}{n^8} + O\left(\frac{n^4}{n^{10}}\right)$$

$$S_n = \frac{n}{n^2} - \frac{n(n+1)}{2n^4} + \cdots.$$

Si on se base sur les sommes des deux premières colonnes, il semblerait qu'on obtienne  $S_n = n^{-1} - \frac{1}{2}n^{-2} + O(n^{-3})$ . Cependant, le calcul devient de plus en plus difficile à chaque colonne.

Si nous persévérons dans cette approche, nous finirons par atteindre notre but. Cependant, nous n'allons pas continuer dans ce sens, et cela pour deux raisons. D'abord la dernière colonne nous donnerait des termes en  $O(n^{-6})$  lorsque  $n/2 \leq k \leq n$ , donc nous aurions finalement une erreur en  $O(n^{-5})$ ; ce serait trop, et il nous faudrait ajouter encore une colonne à notre développement. Aucun enseignant ne peut être assez sadique pour nous imposer cela. Il doit donc y avoir une meilleure méthode. La seconde raison est qu'en effet il y a une meilleure méthode, et de plus elle nous saute aux yeux.

En effet, nous connaissons une forme close de  $S_n$ : c'est  $H_{n^2+n} - H_{n^2}$ . Nous connaissons aussi une bonne approximation des nombres harmoniques. Nous n'avons qu'à l'appliquer deux fois :

$$\begin{aligned} H_{n^2+n} &= \ln(n^2+n) + \gamma + \frac{1}{2(n^2+n)} - \frac{1}{12(n^2+n)^2} + O\left(\frac{1}{n^8}\right); \\ H_{n^2} &= \ln n^2 + \gamma + \frac{1}{2n^2} - \frac{1}{12n^4} + O\left(\frac{1}{n^8}\right). \end{aligned}$$

Maintenant, nous pouvons retirer la plus grande part des termes et simplifier, exactement comme ce que nous avons fait pour la formule de Stirling. Ainsi,

$$\begin{aligned} \ln(n^2+n) &= \ln n^2 + \ln\left(1 + \frac{1}{n}\right) = \ln n^2 + \frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} - \dots; \\ \frac{1}{n^2+n} &= \frac{1}{n^2} - \frac{1}{n^3} + \frac{1}{n^4} - \dots; \\ \frac{1}{(n^2+n)^2} &= \frac{1}{n^4} - \frac{2}{n^5} + \frac{3}{n^6} - \dots. \end{aligned}$$

Beaucoup de choses se simplifient et on trouve

$$\begin{aligned} S_n &= n^{-1} - \frac{1}{2}n^{-2} + \frac{1}{3}n^{-3} - \frac{1}{4}n^{-4} + \frac{1}{5}n^{-5} - \frac{1}{6}n^{-6} \\ &\quad - \frac{1}{2}n^{-3} + \frac{1}{2}n^{-4} - \frac{1}{2}n^{-5} + \frac{1}{2}n^{-6} \\ &\quad + \frac{1}{6}n^{-5} - \frac{1}{4}n^{-6} \end{aligned}$$

plus des termes en  $O(n^{-7})$ . Un petit peu d'arithmétique, et c'est terminé :

$$S_n = n^{-1} - \frac{1}{2}n^{-2} - \frac{1}{6}n^{-3} + \frac{1}{4}n^{-4} - \frac{2}{15}n^{-5} + \frac{1}{12}n^{-6} + O(n^{-7}). \quad (9.50)$$

Il serait intéressant de pouvoir vérifier numériquement ce résultat, comme nous le faisions pour des résultats exacts dans les chapitres précédents. Les formules asymptotiques sont plus difficiles à vérifier que les formules exactes car le terme en  $O$  peut cacher des constantes très grandes ; dans ce cas, on ne peut rien conclure d'une vérification numérique. Cependant, en pratique, nous n'avons aucune raison de penser que quelqu'un

On parie ?

essaie de nous piéger, et nous pouvons donc supposer a priori que ces constantes inconnues sont raisonnablement petites. Avec une calculatrice, on trouve que  $S_4 = \frac{1}{17} + \frac{1}{18} + \frac{1}{19} + \frac{1}{20} = 0,2170107$ , et notre estimation asymptotique pour  $n = 4$  donne

$$\frac{1}{4}\left(1 + \frac{1}{4}\left(-\frac{1}{2} + \frac{1}{4}\left(-\frac{1}{6} + \frac{1}{4}\left(\frac{1}{4} + \frac{1}{4}\left(-\frac{2}{15} + \frac{1}{4}\cdot\frac{1}{12}\right)\right)\right)\right)\right) = 0,2170125.$$

Si nous avions fait un erreur de, disons  $\frac{1}{12}$  dans le terme en  $n^{-6}$ , nous aurions trouvé une différence de  $\frac{1}{12}\cdot\frac{1}{4096}$  dans la cinquième décimale. Notre estimation est donc probablement correcte.

### *Problème 5 : une somme infinie.*

Intéressons-nous maintenant à une question posée par Solomon Golomb [152] : quelle est la valeur asymptotique de

$$S_n = \sum_{k \geq 1} \frac{1}{kN_n(k)^2}, \quad (9.51)$$

où  $N_n(k)$  désigne le nombre de chiffres nécessaires pour écrire  $k$  en base  $n$  ?

Essayons d'abord, comme tout à l'heure, d'approximer "à vue de nez".

Le nombre  $N_n(k)$  de chiffres vaut à peu près  $\log_n k = \log k / \log n$  ; chaque terme de la somme est donc à peu près égal à  $(\log n)^2 / k(\log k)^2$ . En sommant sur  $k$ , on trouve  $\approx (\log n)^2 \sum_{k \geq 2} 1/k(\log k)^2$ , et cette somme converge vers une valeur constante car on peut la comparer à l'intégrale

$$\int_2^\infty \frac{dx}{x(\ln x)^2} = -\frac{1}{\ln x} \Big|_2^\infty = \frac{1}{\ln 2}.$$

Par conséquent, nous pouvons nous attendre à ce que  $S_n$  vaille à peu près  $C(\log n)^2$ , pour une certaine constante  $C$ .

Bien que ce genre de raisonnement soit bien utile pour nous donner une idée de la solution, il nous faut une meilleure estimation pour résoudre le problème. On peut voir ce que cela donne si on exprime  $N_n(k)$  exactement :

$$N_n(k) = \lfloor \log_n k \rfloor + 1. \quad (9.52)$$

Ainsi, par exemple,  $k$  s'écrit avec trois chiffres en base  $n$  si  $n^2 \leq k < n^3$ , et cela arrive précisément lorsque  $\lfloor \log_n k \rfloor = 2$ . Il s'ensuit que  $N_n(k) > \log_n k$ , donc  $S_n = \sum_{k \geq 1} 1/kN_n(k)^2 < 1 + (\log n)^2 \sum_{k \geq 2} 1/k(\log k)^2$ .

On peut aussi essayer de procéder comme au problème 1, en écrivant  $N_n(k) = \log_n k + O(1)$  et en injectant cette expression dans la formule donnée pour  $S_n$ . Le terme représenté par  $O(1)$  se comporte bien, car il est toujours entre 0 et 1, et il vaut à peu près  $\frac{1}{2}$  en moyenne. Cependant, on n'obtient toujours pas de bonne approximation pour  $S_n$  ; on a zéro chiffre significatif (donc une erreur relative très importante) lorsque  $k$  est petit,

justement pour les termes qui contribuent le plus à la somme. Il nous faut une autre idée.

Ce qu'il faut faire, c'est, comme pour le problème 4, d'abord transformer la somme en quelque chose de plus sympathique, avant d'essayer de l'approximer. Pour ce faire, introduisons une nouvelle variable de sommation,  $m = N_n(k)$  :

$$\begin{aligned} S_n &= \sum_{k,m \geq 1} \frac{[m = N_n(k)]}{km^2} \\ &= \sum_{k,m \geq 1} \frac{[n^{m-1} \leq k < n^m]}{km^2} \\ &= \sum_{m \geq 1} \frac{1}{m^2} (H_{n^m-1} - H_{n^{m-1}-1}). \end{aligned}$$

Bien que cela puisse sembler pire que la somme de départ, c'est en réalité un progrès parce que nous savons très bien approximer les nombres harmoniques.

Inutile cependant de nous précipiter tout de suite sur le calcul asymptotique. Retenons-nous et essayons de simplifier encore un peu notre somme. En sommant par parties, nous pouvons regrouper les termes de chaque valeur de  $H_{n^m-1}$  à approximer :

$$S_n = \sum_{k \geq 1} H_{n^k-1} \left( \frac{1}{k^2} - \frac{1}{(k+1)^2} \right).$$

Par exemple,  $H_{n^2-1}$  est multiplié par  $1/2^2$  puis par  $-1/3^2$  (nous avons utilisé le fait que  $H_{n^0-1} = H_0 = 0$ ).

Nous sommes maintenant prêts à développer les nombres harmoniques. Nous avons compris, depuis que nous avons travaillé sur  $(n-1)!$ , qu'il sera certainement plus facile d'estimer  $H_{n^k}$  que  $H_{n^k-1}$ . C'est pourquoi nous écrivons

$$\begin{aligned} H_{n^k-1} &= H_{n^k} - \frac{1}{n^k} = \ln n^k + \gamma + \frac{1}{2n^k} + O\left(\frac{1}{n^{2k}}\right) - \frac{1}{n^k} \\ &= k \ln n + \gamma - \frac{1}{2n^k} + O\left(\frac{1}{n^{2k}}\right). \end{aligned}$$

Maintenant, notre somme se trouve réduite à

$$\begin{aligned} S_n &= \sum_{k \geq 1} \left( k \ln n + \gamma - \frac{1}{2n^k} + O\left(\frac{1}{n^{2k}}\right) \right) \left( \frac{1}{k^2} - \frac{1}{(k+1)^2} \right) \\ &= (\ln n) \Sigma_1 + \gamma \Sigma_2 - \frac{1}{2} \Sigma_3(n) + O(\Sigma_3(n^2)). \end{aligned} \tag{9.53}$$

Il ne reste que quatre sommes,  $\Sigma_1$ ,  $\Sigma_2$ ,  $\Sigma_3(n)$  et  $\Sigma_3(n^2)$ , qui ne présentent pas de difficulté.

Commençons par les  $\Sigma_3$ , car  $\Sigma_3(n^2)$  constitue le terme en  $O$  ; nous allons voir ainsi l'erreur que nous obtiendrons (il n'y a pas de raison d'effectuer les autres calculs avec une grande précision si en définitive ils doivent être absorbés par un grand  $O$ ). Cette somme est simplement la série

$$\Sigma_3(x) = \sum_{k \geq 1} \left( \frac{1}{k^2} - \frac{1}{(k+1)^2} \right) x^{-k},$$

qui converge lorsque  $x \geq 1$  ; nous pouvons donc la tronquer en n'importe quel terme. Si nous la tronquons au terme tel que  $k = 1$ , nous obtenons  $\Sigma_3(n^2) = O(n^{-2})$  ; donc (9.53) aura une erreur absolue en  $O(n^{-2})$  (pour la diminuer, nous pourrions utiliser une meilleure approximation de  $H_{n^k}$ , mais  $O(n^{-2})$  nous suffira). Si nous tronquons  $\Sigma_3(n)$  au terme tel que  $k = 2$ , nous obtenons

$$\Sigma_3(n) = \frac{3}{4}n^{-1} + O(n^{-2}).$$

C'est exactement la précision qu'il nous faut.

Tant que nous y sommes, occupons-nous tout de suite de  $\Sigma_2$ , cela ne nous prendra pas longtemps :

$$\Sigma_2 = \sum_{k \geq 1} \left( \frac{1}{k^2} - \frac{1}{(k+1)^2} \right).$$

C'est la série télescopique  $(1 - \frac{1}{4}) + (\frac{1}{4} - \frac{1}{9}) + (\frac{1}{9} - \frac{1}{16}) + \dots = 1$ .

Pour finir, voici  $\Sigma_1$ , le coefficient de  $\ln n$  dans (9.53) :

$$\Sigma_1 = \sum_{k \geq 1} k \left( \frac{1}{k^2} - \frac{1}{(k+1)^2} \right).$$

Cela donne  $(1 - \frac{1}{4}) + (\frac{2}{4} - \frac{2}{9}) + (\frac{3}{9} - \frac{3}{16}) + \dots = \frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \dots = H_\infty^{(2)} = \pi^2/6$ . Remarquons que si nous n'avions pas fait une sommation par parties tout à l'heure, nous aurions pu voir directement que  $S_n \sim \sum_{k \geq 1} (\ln n)/k^2$ , car  $H_{n^k-1} - H_{n^k-1-1} \sim \ln n$ . Ainsi, la sommation par parties ne nous a été d'aucune utilité pour calculer le terme principal de (9.53) ; par contre, elle a facilité le reste du travail.

Maintenant, il ne nous reste plus qu'à rassembler tout cela pour trouver la réponse au problème de Golomb :

$$S_n = \frac{\pi^2}{6} \ln n + \gamma - \frac{3}{8n} + O\left(\frac{1}{n^2}\right). \quad (9.54)$$

Notez que cette expression croît plus lentement que notre première estimation “à vue de nez”, qui était de  $C(\log n)^2$ . Les sommes discrètes n’obéissent pas forcément à des approximations “continues”.

**Problème 6 : grand Phi.**

Vers la fin du chapitre 4, nous avons remarqué que le nombre de fractions de la suite de Farey  $\mathcal{F}_n$  est égal à  $1 + \Phi(n)$ , où  $\Phi(n) = \varphi(1) + \varphi(2) + \dots + \varphi(n)$ , et nous avons établi, en (4.62), que

$$\Phi(n) = \frac{1}{2} \sum_{k \geq 1}^{\lfloor n/k \rfloor} \mu(k) \lfloor n/k \rfloor \lfloor 1 + n/k \rfloor. \quad (9.55)$$

Nous allons tenter d'estimer  $\Phi(n)$  lorsque  $n$  est grand (ce sont des sommes de ce genre qui ont incité Bachmann à inventer la notation  $O$ ).

Si on n'oublie pas de penser GRAND, on peut voir tout de suite que  $\Phi(n)$  sera très probablement proportionnel à  $n^2$ . En effet, si le dernier facteur était  $\lfloor n/k \rfloor$  au lieu de  $\lfloor 1 + n/k \rfloor$ , on aurait  $|\Phi(n)| \leq \frac{1}{2} \sum_{k \geq 1} \lfloor n/k \rfloor^2 \leq \frac{1}{2} \sum_{k \geq 1} (n/k)^2 = \frac{\pi^2}{12} n^2$ , car la fonction de Möbius  $\mu(k)$  vaut soit  $-1$ , soit  $0$ , soit  $+1$ . Le “ $1 +$ ” de ce dernier facteur a pour effet d'ajouter  $\sum_{k \geq 1} \mu(k) \lfloor n/k \rfloor$  à la somme ; comme c'est nul pour  $k > n$ , cela ne peut pas être plus grand que  $nH_n = O(n \log n)$  en valeur absolue.

A la lumière de cette analyse préliminaire, nous écrivons

$$\begin{aligned} \Phi(n) &= \frac{1}{2} \sum_{k=1}^n \mu(k) \left( \left( \frac{n}{k} \right) + O(1) \right)^2 \\ &= \frac{1}{2} \sum_{k=1}^n \mu(k) \left( \left( \frac{n}{k} \right)^2 + O\left( \frac{n}{k} \right) \right) \\ &= \frac{1}{2} \sum_{k=1}^n \mu(k) \left( \frac{n}{k} \right)^2 + \sum_{k=1}^n O\left( \frac{n}{k} \right) \\ &= \frac{1}{2} \sum_{k=1}^n \mu(k) \left( \frac{n}{k} \right)^2 + O(n \log n). \end{aligned}$$

Nous avons ainsi supprimé les parties entières. Il nous reste à évaluer la somme  $\frac{1}{2} \sum_{k=1}^n \mu(k) n^2/k^2$  avec une précision en  $O(n \log n)$ . Autrement dit, nous devons calculer  $\sum_{k=1}^n \mu(k) 1/k^2$  avec une précision en  $O(n^{-1} \log n)$ . Facile : il suffit de calculer cette somme en autorisant  $k$  à aller jusqu'à  $\infty$ , puisque ce qu'on y ajoute ainsi est

$$\begin{aligned} \sum_{k>n} \frac{\mu(k)}{k^2} &= O\left(\sum_{k>n} \frac{1}{k^2}\right) = O\left(\sum_{k>n} \frac{1}{k(k-1)}\right) \\ &= O\left(\sum_{k>n} \left( \frac{1}{k-1} - \frac{1}{k} \right)\right) = O\left(\frac{1}{n}\right). \end{aligned}$$

Nous avons démontré en (7.89) que  $\sum_{k \geq 1} \mu(k)/k^z = 1/\zeta(z)$ . Par conséquent,  $\sum_{k \geq 1} \mu(k)/k^2 = 1/(\sum_{k \geq 1} 1/k^2) = 6/\pi^2$ , et voici donc notre réponse :

$$\Phi(n) = \frac{3}{\pi^2} n^2 + O(n \log n). \quad (9.56)$$

(Saltykov a montré en 1960 [316] que l'erreur est au plus en  $O(n(\log n)^{2/3} \times (\log \log n)^{1+\epsilon})$  ; on sait aussi, grâce à Montgomery [275], qu'elle est plus grande que  $o(n(\log \log n)^{1/2})$ ).

## 9.4 DEUX TRUCS ASYMPTOTIQUES

Maintenant que nous commençons à bien savoir manipuler les O, prenons un peu de recul pour observer ce que nous venons de faire. Cela va nous fournir encore des armes à ajouter à notre arsenal asymptotique. Elles nous seront bien utiles pour attaquer des problèmes plus coriaces.

### *Truc numéro 1 : l'amorçage.*

Lorsque nous avons approximé le  $n$ ème nombre premier  $P_n$ , dans le problème 3 de la section 9.3, nous avons résolu une récurrence asymptotique de la forme

$$P_n = n \ln P_n (1 + O(1/\log n)).$$

Pour démontrer que  $P_n = n \ln n + O(n)$ , nous avons commencé par prouver, en utilisant la récurrence, que  $P_n = O(n^2)$ . Nous avons ainsi appliqué une méthode générale, appelée *amorçage*, qui consiste à résoudre asymptotiquement une récurrence en partant d'une estimation grossière et en l'injectant dans la récurrence (pour "amorcer la pompe" en quelque sorte). De cette façon, on peut trouver des estimations de plus en plus précises.

Voici un autre problème qui illustre bien cette méthode : quelle est la valeur asymptotique du coefficient  $g_n = [z^n] G(z)$  dans la fonction génératrice

$$G(z) = \exp\left(\sum_{k \geq 1} \frac{z^k}{k^2}\right), \quad (9.57)$$

lorsque  $n \rightarrow \infty$ ? En dérivant cette équation par rapport à  $z$ , on trouve

$$G'(z) = \sum_{n=0}^{\infty} n g_n z^{n-1} = \left(\sum_{k \geq 1} \frac{z^{k-1}}{k}\right) G(z);$$

puis, en mettant en équations les coefficients de  $z^{n-1}$  des deux membres, on obtient la récurrence

$$n g_n = \sum_{0 \leq k < n} \frac{g_k}{n-k}. \quad (9.58)$$

Notre problème consiste à trouver un équivalent asymptotique de la solution de (9.58), avec la condition initiale  $g_0 = 1$ . Les premières valeurs

$n$	0	1	2	3	4	5	6
$g_n$	1	1	$\frac{3}{4}$	$\frac{19}{36}$	$\frac{107}{288}$	$\frac{641}{2400}$	$\frac{51103}{259200}$

ne révèlent pas de motif évident, et la suite d'entiers  $\langle n!^2 g_n \rangle$  ne se trouve pas dans l'Encyclopédie de Sloane et Plouffe [330]. Par conséquent, il semble hors de question de trouver une forme close pour  $g_n$  ; nous ne pourrons très probablement pas faire mieux qu'une formule asymptotique.

Commençons par remarquer que  $0 < g_n \leq 1$  pour tout  $n \geq 0$  ; cela se prouve facilement par induction. Nous avons donc un point de départ :

$$g_n = O(1).$$

Nous allons l'utiliser pour "amorcer la pompe". Injectons donc cette équation dans le membre droit de (9.58) :

$$ng_n = \sum_{0 \leq k < n} \frac{O(1)}{n-k} = H_n O(1) = O(\log n).$$

Par conséquent,

$$g_n = O\left(\frac{\log n}{n}\right), \quad \text{pour } n > 1.$$

Continuons notre amorçage :

$$\begin{aligned} ng_n &= \frac{1}{n} + \sum_{0 < k < n} \frac{O((1 + \log k)/k)}{n-k} \\ &= \frac{1}{n} + \sum_{0 < k < n} \frac{O(\log n)}{k(n-k)} \\ &= \frac{1}{n} + \sum_{0 < k < n} \left(\frac{1}{k} + \frac{1}{n-k}\right) \frac{O(\log n)}{n} \\ &= \frac{1}{n} + \frac{2}{n} H_{n-1} O(\log n) = \frac{1}{n} O(\log n)^2, \end{aligned}$$

ce qui nous donne

$$g_n = O\left(\frac{\log n}{n}\right)^2. \tag{9.59}$$

Peut-on continuer ainsi à l'infini, pour obtenir peut-être quelque chose comme  $g_n = O(n^{-1} \log n)^m$  pour tout  $m$  ?

Eh bien non. Si on effectue encore une étape d'amorçage, on trouve la somme

$$\begin{aligned} \sum_{0 < k < n} \frac{1}{k^2(n-k)} &= \sum_{0 < k < n} \left( \frac{1}{nk^2} + \frac{1}{n^2k} + \frac{1}{n^2(n-k)} \right) \\ &= \frac{1}{n} H_{n-1}^{(2)} + \frac{2}{n^2} H_{n-1}, \end{aligned}$$

qui est en  $\Omega(n^{-1})$ . Il n'est donc pas possible d'obtenir une estimation de  $g_n$  en dessous de  $\Omega(n^{-2})$ .

Nous en savons cependant assez sur  $g_n$  pour appliquer notre vieux truc consistant à retirer la plus grande part :

$$\begin{aligned} ng_n &= \sum_{0 \leq k \leq n} \frac{g_k}{n} + \sum_{0 \leq k < n} g_k \left( \frac{1}{n-k} - \frac{1}{n} \right) \\ &= \frac{1}{n} \sum_{k \geq 0} g_k - \frac{1}{n} \sum_{k \geq n} g_k + \frac{1}{n} \sum_{0 \leq k < n} \frac{kg_k}{n-k}. \end{aligned} \quad (9.60)$$

La première somme est  $G(1) = \exp(\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \dots) = e^{\pi^2/6}$ , car  $G(z)$  converge pour tout  $|z| \leq 1$ . La seconde somme est la queue de la première ; nous pouvons en obtenir un majorant en appliquant (9.59) :

$$\sum_{k \geq n} g_k = O\left(\sum_{k \geq n} \frac{(\log k)^2}{k^2}\right) = O\left(\frac{(\log n)^2}{n}\right).$$

La dernière égalité de la formule ci-dessus peut se démontrer ainsi :

$$\sum_{k > n} \frac{(\log k)^2}{k^2} < \sum_{m \geq 1} \sum_{n^m < k \leq n^{m+1}} \frac{(\log n^{m+1})^2}{k(k-1)} < \sum_{m \geq 1} \frac{(m+1)^2 (\log n)^2}{n^m}.$$

(On trouvera dans l'exercice 54 une méthode plus générale pour estimer ce genre de queues de séries).

En utilisant un argument que nous commençons à bien connaître, on trouve que la troisième somme de (9.60) est en

$$O\left(\sum_{0 < k < n} \frac{(\log n)^2}{k(n-k)}\right) = O\left(\frac{(\log n)^3}{n}\right).$$

Par conséquent, (9.60) entraîne que

$$g_n = \frac{e^{\pi^2/6}}{n^2} + O\left(\frac{(\log n)^3}{n}\right). \quad (9.61)$$

Terminons par un dernier amorçage en injectant cette formule dans la récurrence. Nous obtenons

$$g_n = \frac{e^{\pi^2/6}}{n^2} + O(\log n/n^3). \quad (9.62)$$

(L'exercice 23 donne un aperçu de ce qu'il y a dans le terme en  $O$ ).

### **Truc numéro 2 : couper la somme**

Pour arriver à (9.62), nous avons utilisé une méthode similaire à celle qui nous a permis de calculer la valeur asymptotique (9.56) de  $\Phi(n)$  : dans les deux cas, nous sommes partis d'une somme finie, et nous avons résolu notre problème en la prolongeant en une somme infinie. Pour ce faire, il nous a fallu faire bien attention et considérer deux approches différentes selon que  $k$  était petit ou grand.

En fait, nous avons appliqué une importante méthode de sommation asymptotique en trois étapes, que nous allons voir maintenant dans toute sa généralité. Voici ce qu'on peut essayer de faire lorsqu'on doit estimer la valeur d'une somme  $\sum_k a_k(n)$  :

(Cette méthode est due à Laplace [240].)

- 1 Découper l'ensemble des valeurs possibles de l'indice de sommation en deux parties disjointes  $D_n$  et  $T_n$ . La sommation sur  $D_n$  doit constituer la part "dominante", en ce sens qu'elle doit comporter assez de termes pour déterminer les chiffres significatifs de la somme lorsque  $n$  est grand. La sommation sur  $T_n$ , la "queue" de la somme, ne doit contribuer que très peu au résultat.
- 2 Trouver une estimation asymptotique

$$a_k(n) = b_k(n) + O(c_k(n))$$

correcte lorsque  $k \in D_n$ . Le terme en  $O$  n'a pas besoin d'être valable pour  $k \in T_n$ .

- 3 Prouver que les trois sommes suivantes sont petites :

$$\begin{aligned} \Sigma_a(n) &= \sum_{k \in T_n} a_k(n); & \Sigma_b(n) &= \sum_{k \in T_n} b_k(n); \\ \Sigma_c(n) &= \sum_{k \in D_n} |c_k(n)|. \end{aligned} \quad (9.63)$$

Si on a pu effectuer ces trois étapes, on obtient une bonne estimation de la somme :

$$\sum_{k \in D_n \cup T_n} a_k(n) = \sum_{k \in D_n \cup T_n} b_k(n) + O(\Sigma_a(n)) + O(\Sigma_b(n)) + O(\Sigma_c(n)).$$

Voici en fait ce qui se passe lorsqu'on suit ces trois étapes. On "coupe" la queue de la somme de départ pour obtenir une bonne estimation dans  $D_n$  :

$$\sum_{k \in D_n} a_k(n) = \sum_{k \in D_n} (b_k(n) + O(c_k(n))) = \sum_{k \in D_n} b_k(n) + O(\Sigma_c(n)).$$

Puis on remplace l'ancienne queue par une autre. La nouvelle queue peut même être une très mauvaise approximation de l'ancienne ; cela n'a pas d'importance car la queue compte très peu dans la somme totale. Cela donne

$$\begin{aligned} \sum_{k \in T_n} a_k(n) &= \sum_{k \in T_n} (b_k(n) - b_k(n) + a_k(n)) \\ &= \sum_{k \in T_n} b_k(n) + O(\Sigma_b(n)) + O(\Sigma_a(n)). \end{aligned}$$

*Le calcul asymptotique est l'art de savoir où on peut se permettre d'être imprécis et où il faut absolument être précis.*

Si on regarde sous cet angle le calcul de la somme (9.60), par exemple, on voit que

$$\begin{aligned} a_k(n) &= [0 \leq k < n] g_k / (n - k), \\ b_k(n) &= g_k / n, \\ c_k(n) &= k g_k / n(n - k); \end{aligned}$$

les deux ensembles de sommation sont

$$D_n = \{0, 1, \dots, n-1\}, \quad T_n = \{n, n+1, \dots\},$$

et on trouve

$$\Sigma_a(n) = 0, \quad \Sigma_b(n) = O((\log n)^2/n^2), \quad \Sigma_c(n) = O((\log n)^3/n^2),$$

ce qui mène à (9.61).

En ce qui concerne l'estimation de  $\Phi(n)$  en (9.55), on a

$$\begin{aligned} a_k(n) &= \mu(k) \lfloor n/k \rfloor \lfloor 1+n/k \rfloor, \quad b_k(n) = \mu(k) n^2/k^2, \quad c_k(n) = n/k; \\ D_n &= \{1, 2, \dots, n\}, \quad T_n = \{n+1, n+2, \dots\}. \end{aligned}$$

On trouve (9.56) en remarquant que  $\Sigma_a(n) = 0$ ,  $\Sigma_b(n) = O(n)$  et  $\Sigma_c(n) = O(n \log n)$ .

Voici un autre exemple où cette méthode marche bien (contrairement à nos exemples précédents, celui-ci illustre la méthode dans toute sa généralité, lorsque  $\Sigma_a(n) \neq 0$ ). Nous devons trouver la valeur asymptotique de

$$L_n = \sum_{k \geq 0} \frac{\ln(n+2^k)}{k!}.$$

A cause du  $k!$  au dénominateur, la plus grande part de cette somme se trouve dans les termes où  $k$  est petit. Si  $k$  est petit, donc, on a

$$\ln(n + 2^k) = \ln n + \frac{2^k}{n} - \frac{2^{2k}}{2n^2} + O\left(\frac{2^{3k}}{n^3}\right). \quad (9.64)$$

On peut montrer que cette estimation est valable pour  $0 \leq k < \lfloor \lg n \rfloor$ , car les termes qui ont été absorbés par le  $O$  sont majorés par la série convergente

$$\sum_{m \geq 3} \frac{2^{km}}{mn^m} \leq \frac{2^{3k}}{n^3} \sum_{m \geq 3} \frac{2^{k(m-3)}}{n^{m-3}} \leq \frac{2^{3k}}{n^3} \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) = \frac{2^{3k}}{n^3} \cdot 2.$$

(Dans cet intervalle,  $2^k/n \leq 2^{\lfloor \lg n \rfloor - 1}/n \leq \frac{1}{2}$ ).

Nous pouvons donc appliquer la méthode que nous venons d'apprendre, en prenant

$$\begin{aligned} a_k(n) &= \ln(n + 2^k)/k!, \\ b_k(n) &= (\ln n + 2^k/n - 4^k/2n^2)/k!, \\ c_k(n) &= 8^k/n^3 k!; \end{aligned}$$

$$\begin{aligned} D_n &= \{0, 1, \dots, \lfloor \lg n \rfloor - 1\}, \\ T_n &= \{\lfloor \lg n \rfloor, \lfloor \lg n \rfloor + 1, \dots\}. \end{aligned}$$

Il ne nous reste plus qu'à trouver de bons majorants pour les trois  $\Sigma$  de (9.63). Nous saurons alors que  $\sum_{k \geq 0} a_k(n) \approx \sum_{k \geq 0} b_k(n)$ .

L'erreur que nous avons commise dans la partie dominante de la somme,  $\Sigma_c(n) = \sum_{k \in D_n} 8^k/n^3 k!$ , est trivialement majorée par  $\sum_{k \geq 0} 8^k/n^3 k! = e^8/n^3$ ; nous pouvons donc la remplacer par  $O(n^{-3})$ . Passons à  $\Sigma_b(n)$ :

$$\begin{aligned} |\Sigma_b(n)| &= \left| \sum_{k \geq \lfloor \lg n \rfloor} b_k(n) \right| \\ &< \sum_{k \geq \lfloor \lg n \rfloor} \frac{\ln n + 2^k + 4^k}{k!} \\ &< \frac{\ln n + 2^{\lfloor \lg n \rfloor} + 4^{\lfloor \lg n \rfloor}}{\lfloor \lg n \rfloor!} \sum_{k \geq 0} \frac{4^k}{k!} = O\left(\frac{n^2}{\lfloor \lg n \rfloor!}\right). \end{aligned}$$

Comme  $\lfloor \lg n \rfloor!$  croît plus vite que  $n$ 'importe quelle puissance de  $n$ , cette erreur minuscule est absorbée par  $\Sigma_c(n) = O(n^{-3})$ . L'erreur qui provient de la queue d'origine,

$$\Sigma_a(n) = \sum_{k \geq \lfloor \lg n \rfloor} a_k(n) < \sum_{k \geq \lfloor \lg n \rfloor} \frac{k + \ln n}{k!},$$

est même encore plus petite.

Pour finir, on trouve sans difficulté une forme close de  $\sum_{k \geq 0} b_k(n)$ , ce qui nous donne la formule asymptotique recherchée :

$$\sum_{k \geq 0} \frac{\ln(n + 2^k)}{k!} = e \ln n + \frac{e^2}{n} - \frac{e^4}{2n^2} + O\left(\frac{1}{n^3}\right). \quad (9.65)$$

En fait, on peut même écrire que

$$\sum_{k \geq 0} \frac{\ln(n + 2^k)}{k!} = e \ln n + \sum_{k=1}^{m-1} (-1)^{k+1} \frac{e^{2^k}}{kn^k} + O\left(\frac{1}{n^m}\right), \quad (9.66)$$

pour tout  $m > 0$  fixé (c'est une série tronquée, qui diverge pour tout  $n$  si  $m \rightarrow \infty$ ).

Il y a quand même un défaut dans notre solution : nous avons été trop prudents. Nous avons supposé que  $k < \lfloor \lg n \rfloor$  pour trouver (9.64) ; or, l'exercice 53 montre que la formule est valable en fait pour tout  $k$ . Si nous avions su cela, nous aurions pu trouver directement la solution ! Nous verrons cependant des cas où la méthode en trois étapes est la seule approche décemment possible.

## 9.5 LA FORMULE DE SOMMATION D'EULER

Voici maintenant la dernière technique importante que nous allons présenter dans ce livre. C'est une méthode générale d'approximation qui fut découverte et publiée par Leonhard Euler [101] en 1732 (on l'associe parfois à Colin Maclaurin, un professeur de mathématiques à Edinburgh, qui la découvrit indépendamment un peu plus tard [263, page 305]).

Voici la formule :

$$\sum_{a \leq k < b} f(k) = \int_a^b f(x) dx + \sum_{k=1}^m \frac{B_k}{k!} f^{(k-1)}(x) \Big|_a^b + R_m, \quad (9.67)$$

$$\text{avec } R_m = (-1)^{m+1} \int_a^b \frac{B_m(f(x))}{m!} f^{(m)}(x) dx, \quad \begin{array}{l} a \leq b \text{ entiers;} \\ m \geq 1 \text{ entier.} \end{array} \quad (9.68)$$

A notre gauche, une somme à évaluer. A notre droite, une expression de cette somme, à base d'intégrales et de dérivées. Si  $f(x)$  est une fonction suffisamment "lisse" pour admettre  $m$  dérivées  $f'(x), \dots, f^{(m)}(x)$ , cette formule est une identité. Le membre droit donne souvent une excellente approximation de la somme de gauche car le reste  $R_m$  est souvent petit. Par exemple, nous verrons que la formule de Stirling et notre approximation du nombre harmonique  $H_n$  sont des conséquences de la formule de sommation d'Euler.

Les nombres  $B_k$  de (9.67) sont les nombres de Bernoulli que nous avons rencontrés au chapitre 6. La fonction  $B_m(\{x\})$  de (9.68) est le polynôme de Bernoulli que nous avons rencontré au chapitre 7. La notation  $\{x\}$  désigne la partie réelle  $x - \lfloor x \rfloor$  de  $x$ , comme au chapitre 3. La formule de sommation d'Euler rassemble tout cela.

Rappelons les premiers nombres de Bernoulli, qui pourront certainement nous être utiles :

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \quad B_8 = -\frac{1}{30}; \\ B_3 = B_5 = B_7 = B_9 = B_{11} = \cdots = 0.$$

Jakob Bernoulli découvrit ces nombres alors qu'il étudiait les sommes de puissances d'entiers, et la formule d'Euler explique pourquoi : si on pose  $f(x) = x^{m-1}$ , alors  $f^{(m)}(x) = 0$ ; donc  $R_m = 0$ , et (9.67) se réduit à

$$\sum_{a \leq k < b} k^{m-1} = \left. \frac{x^m}{m} \right|_a^b + \left. \sum_{k=1}^m \frac{B_k}{k!} (m-1)^{k-1} x^{m-k} \right|_a^b \\ = \frac{1}{m} \sum_{k=0}^m \binom{m}{k} B_k \cdot (b^{m-k} - a^{m-k}).$$

Par exemple, pour  $m = 3$ , on retrouve notre somme préférée :

$$\sum_{0 \leq k < n} k^2 = \frac{1}{3} \left( \binom{3}{0} B_0 n^3 + \binom{3}{1} B_1 n^2 + \binom{3}{2} B_2 n \right) = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}.$$

(C'est la dernière fois que nous calculons cette somme dans ce livre).

Avant de prouver la formule d'Euler, nous allons voir une bonne raison (due à Lagrange [234]) pour laquelle il faudrait l'inventer si elle n'existe pas déjà. Au chapitre 2, nous avons défini l'opérateur de différence  $\Delta$  et expliqué que  $\sum$  est l'inverse de  $\Delta$ , exactement comme  $\int$  est l'inverse de l'opérateur de dérivation  $D$ . On peut exprimer  $\Delta$  en fonction de  $D$  avec la formule de Taylor :

$$f(x + \epsilon) = f(x) + \frac{f'(x)}{1!} \epsilon + \frac{f''(x)}{2!} \epsilon^2 + \cdots$$

En posant  $\epsilon = 1$ , on obtient

$$\begin{aligned} \Delta f(x) &= f(x+1) - f(x) \\ &= f'(x)/1! + f''(x)/2! + f'''(x)/3! + \cdots \\ &= (D/1! + D^2/2! + D^3/3! + \cdots) f(x) = (e^D - 1) f(x). \end{aligned} \quad (9.69)$$

*Toutes les bonnes choses ont une fin.*

Ici,  $e^D$  désigne l'opération  $1+D/1!+D^2/2!+D^3/3!+\dots$ . Comme  $\Delta = e^D - 1$ , l'opérateur inverse  $\Sigma = 1/\Delta$  doit être  $1/(e^D - 1)$ . Or, nous savons, d'après la table 373, que  $z/(e^z - 1)$  est une série dans laquelle apparaissent les nombres de Bernoulli :  $z/(e^z - 1) = \sum_{k \geq 0} B_k z^k/k!$ . Par conséquent,

$$\sum = \frac{B_0}{D} + \frac{B_1}{1!} + \frac{B_2}{2!} D + \frac{B_3}{3!} D^2 + \dots = \int + \sum_{k \geq 1} \frac{B_k}{k!} D^{k-1}. \quad (9.70)$$

En appliquant cette équation d'opérateurs à  $f(x)$  et en y ajoutant des bornes, on trouve

$$\sum_a^b f(x) dx = \int_a^b f(x) dx + \sum_{k \geq 1} \frac{B_k}{k!} f^{(k-1)}(x) \Big|_a^b, \quad (9.71)$$

c'est-à-dire exactement la formule de sommation d'Euler (9.67) sans le terme résiduel. (En fait, ni Euler ni personne ne pensa à considérer ce terme résiduel jusqu'à ce que S. D. Poisson [295] publiait un important mémoire sur l'approximation en 1823. Ce terme résiduel est important car il arrive souvent que la somme infinie  $\sum_{k \geq 1} (B_k/k!) f^{(k-1)}(x) \Big|_a^b$  diverge. Notre calcul menant à (9.71) est purement formel, sans aucune considération liée à la convergence).

Maintenant, nous allons prouver (9.67) en y incluant le terme résiduel. Il suffit de montrer qu'elle est vraie dans le cas où  $a = 0$  et  $b = 1$ , c'est-à-dire que

$$f(0) = \int_0^1 f(x) dx + \sum_{k=1}^m \frac{B_k}{k!} f^{(k-1)}(x) \Big|_0^1 - (-1)^m \int_0^1 \frac{B_m(x)}{m!} f^{(m)}(x) dx.$$

En effet, on peut alors remplacer  $f(x)$  par  $f(x+l)$  pour tout entier  $l$  et obtenir

$$f(l) = \int_l^{l+1} f(x) dx + \sum_{k=1}^m \frac{B_k}{k!} f^{(k-1)}(x) \Big|_l^{l+1} - (-1)^m \int_l^{l+1} \frac{B_m(x)}{m!} f^{(m)}(x) dx.$$

La formule (9.67) n'est rien d'autre que la somme de cette identité pour  $a \leq l < b$  ; les termes intermédiaires ont le bon goût de se télescopier.

La preuve pour le cas  $a = 0$  et  $b = 1$  se fait par induction sur  $m$ , en partant de  $m = 1$  :

$$f(0) = \int_0^1 f(x) dx - \frac{1}{2}(f(1) - f(0)) + \int_0^1 (x - \frac{1}{2}) f'(x) dx.$$

(Le polynôme de Bernoulli  $B_m(x)$  est défini par l'équation

$$B_m(x) = \binom{m}{0} B_0 x^m + \binom{m}{1} B_1 x^{m-1} + \dots + \binom{m}{m} B_m x^0, \quad (9.72)$$

donc  $B_1(x) = x - \frac{1}{2}$  en particulier). En d'autres termes, nous devons prouver que

$$\frac{f(0) + f(1)}{2} = \int_0^1 f(x) dx + \int_0^1 (x - \frac{1}{2}) f'(x) dx.$$

C'est tout simplement un cas particulier de la formule

$$u(x)v(x)\Big|_0^1 = \int_0^1 u(x) dv(x) + \int_0^1 v(x) du(x) \quad (9.73)$$

d'intégration par parties, avec  $u(x) = f(x)$  et  $v(x) = x - \frac{1}{2}$ . Le cas  $m = 1$  ne pose donc aucun problème.

Pour passer de  $m - 1$  à  $m$ , et donc pour terminer notre preuve par induction, il nous faut montrer que  $R_{m-1} = (B_m/m!)f^{(m-1)}(x)\Big|_0^1 + R_m$ , c'est-à-dire que

$$\begin{aligned} & (-1)^m \int_0^1 \frac{B_{m-1}(x)}{(m-1)!} f^{(m-1)}(x) dx \\ &= \frac{B_m}{m!} f^{(m-1)}(x)\Big|_0^1 - (-1)^m \int_0^1 \frac{B_m(x)}{m!} f^{(m)}(x) dx, \end{aligned}$$

ce qui se réduit à l'équation

$$\begin{aligned} & (-1)^m B_m f^{(m-1)}(x)\Big|_0^1 \\ &= m \int_0^1 B_{m-1}(x) f^{(m-1)}(x) dx + \int_0^1 B_m(x) f^{(m)}(x) dx. \end{aligned}$$

On peut encore appliquer (9.73) à ces deux intégrales, avec  $u(x) = f^{(m-1)}(x)$  et  $v(x) = B_m(x)$ , car la dérivée du polynôme de Bernoulli (9.72) est

$$\begin{aligned} \frac{d}{dx} \sum_k \binom{m}{k} B_k x^{m-k} &= \sum_k \binom{m}{k} (m-k) B_k x^{m-k-1} \\ &= m \sum_k \binom{m-1}{k} B_k x^{m-1-k} \\ &= m B_{m-1}(x). \end{aligned} \quad (9.74)$$

(Nous avons utilisé l'identité d'absorption (5.7) dans ce calcul). Par conséquent, la formule que nous cherchons à prouver sera vraie si et seulement si

$$(-1)^m B_m f^{(m-1)}(x)\Big|_0^1 = B_m(x) f^{(m-1)}(x)\Big|_0^1.$$

En d'autres termes, il faut que

$$(-1)^m B_m = B_m(1) = B_m(0), \quad \text{pour } m > 1. \quad (9.75)$$

Cela peut paraître gênant, car  $B_m(0)$  est égal à  $B_m$ , et pas à  $(-1)^m B_m$  a priori ; mais en réalité il n'y a pas de problème car  $m > 1$  et nous savons que  $B_m$  est nul lorsque  $m$  est impair.

Pour finir notre preuve de la formule de sommation d'Euler, il nous reste à montrer que  $B_m(1) = B_m(0)$ , ce qui revient à dire que

$$\sum_k \binom{m}{k} B_k = B_m, \quad \text{pour } m > 1.$$

C'est exactement la définition des nombres de Bernoulli, (6.79). Nous en avons donc terminé.

L'identité  $B'_m(x) = mB_{m-1}(x)$  implique que

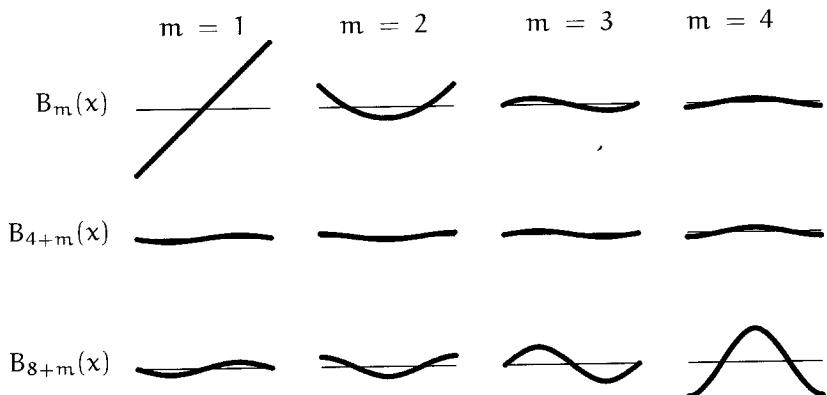
$$\int_0^1 B_m(x) dx = \frac{B_{m+1}(1) - B_{m+1}(0)}{m+1},$$

et nous savons maintenant que cette intégrale est nulle lorsque  $m \geq 1$ . Par conséquent, dans le terme résiduel de la formule d'Euler,

$$R_m = \frac{(-1)^{m+1}}{m!} \int_a^b B_m(\{x\}) f^{(m)}(x) dx,$$

$f^{(m)}(x)$  est multiplié par une fonction  $B_m(\{x\})$  dont la valeur moyenne est égale à zéro. Il y a donc une bonne chance pour que  $R_m$  soit raisonnablement petit.

Puisque c'est  $B_m(x)$  qui détermine le comportement de  $R_m$ , observons ce polynôme de plus près, pour  $0 \leq x \leq 1$ . Voici donc les graphes de  $B_m(x)$  pour les douze premières valeurs de  $m$  :



Les valeurs de  $B_m(x)$ , pour  $m$  allant de 1 à 9, sont plutôt petites. En revanche, elles peuvent devenir extrêmement grandes dès que  $m$  dépasse 9. Heureusement,  $R_m$  contient le facteur  $1/m!$  qui compense cette augmentation.

Dès que  $m \geq 3$ , le graphe de  $B_m(x)$  commence à ressembler de très près à une sinusoïde. On montre dans l'exercice 58 qu'on peut trouver une bonne approximation de  $B_m(x)$  par un multiple négatif de  $\cos(2\pi x - \frac{1}{2}\pi m)$ , avec une erreur en  $O(2^{-m} \max_x B_m(\{x\}))$ .

Le polynôme  $B_{4k+1}(x)$  est négatif si  $0 < x < \frac{1}{2}$  et positif si  $\frac{1}{2} < x < 1$ . Par conséquent, sa primitive  $B_{4k+2}(x)/(4k+2)$  décroît lorsque  $0 < x < \frac{1}{2}$  et croît lorsque  $\frac{1}{2} < x < 1$ . De plus, on a

$$B_{4k+1}(1-x) = -B_{4k+1}(x), \quad \text{pour } 0 \leq x \leq 1,$$

et il s'ensuit que

$$B_{4k+2}(1-x) = B_{4k+2}(x), \quad \text{pour } 0 \leq x \leq 1.$$

En raison du terme constant  $B_{4k+2}$ , l'intégrale  $\int_0^1 B_{4k+2}(x) dx$  est égale à zéro ; donc  $B_{4k+2} > 0$ . Si on intègre  $B_{4k+2}(x)$ , on trouve  $B_{4k+3}(x)/(4k+3)$ , qui doit donc être positif lorsque  $0 < x < \frac{1}{2}$  et négatif lorsque  $\frac{1}{2} < x < 1$  ; de plus,  $B_{4k+3}(1-x) = -B_{4k+3}(x)$ , donc  $B_{4k+3}(x)$  a les mêmes propriétés que  $B_{4k+1}(x)$ , mais avec un changement de signe. Par conséquent,  $B_{4k+4}(x)$  a les mêmes propriétés que  $B_{4k+2}(x)$ , mais avec un changement de signe. Pour les mêmes raisons,  $B_{4k+5}(x)$  a les mêmes propriétés que  $B_{4k+1}(x)$ . Ainsi, ces propriétés sont vraies pour tout  $k$ , par induction.

Selon cette analyse, le maximum de  $B_{2m}(x)$  doit être soit en  $x = 0$ , soit en  $x = \frac{1}{2}$ . On montre dans l'exercice 17 que

$$B_{2m}\left(\frac{1}{2}\right) = (2^{1-2m} - 1)B_{2m}; \tag{9.76}$$

donc

$$|B_{2m}(\{x\})| \leq |B_{2m}|. \tag{9.77}$$

On peut utiliser ce résultat pour établir un majorant du terme résiduel de la formule de sommation d'Euler, ce qui peut être bien utile. D'après (6.89), nous savons que

$$\begin{aligned} \frac{|B_{2m}|}{(2m)!} &= \frac{2}{(2\pi)^{2m}} \sum_{k \geq 1} \frac{1}{k^{2m}} \\ &= O((2\pi)^{-2m}), \quad \text{lorsque } m > 0. \end{aligned}$$

Nous pouvons donc réécrire la formule d'Euler (9.67) comme suit :

$$\begin{aligned} \sum_{a \leq k < b} f(k) &= \int_a^b f(x) dx - \frac{1}{2} f(x)|_a^b + \sum_{k=1}^m \frac{B_{2k}}{(2k)!} f^{(2k-1)}(x)|_a^b \\ &\quad + O((2\pi)^{-2m}) \int_a^b |f^{(2m)}(x)| dx. \end{aligned} \quad (9.78)$$

Par exemple, si  $f(x) = e^x$ , toutes les dérivées sont égales et cette formule nous dit que  $\sum_{a \leq k < b} e^k = (e^b - e^a)(1 - \frac{1}{2} + B_2/2! + B_4/4! + \dots + B_{2m}/(2m)!) + O((2\pi)^{-2m})$ . Bien sûr, nous savons que cette somme est en fait une série géométrique, égale à  $(e^b - e^a)/(e-1) = (e^b - e^a) \sum_{k \geq 0} B_k/k!$ .

Si  $f^{(2m)}(x) \geq 0$  pour  $a \leq x \leq b$ , l'intégrale  $\int_a^b |f^{(2m)}(x)| dx$  est égale à  $|f^{(2m-1)}(x)|_a^b$ , donc on a

$$|R_{2m}| \leq \left| \frac{B_{2m}}{(2m)!} f^{(2m-1)}(x)|_a^b \right|;$$

en d'autres termes, dans ce cas, le terme résiduel est majoré par la valeur absolue du *dernier terme* (celui qui se trouve juste avant le terme résiduel). On peut même donner une meilleure estimation si on sait que

$$f^{(2m+2)}(x) \geq 0 \quad \text{et} \quad f^{(2m+4)}(x) \geq 0, \quad \text{pour } a \leq x \leq b. \quad (9.79)$$

En effet, cela implique la relation

$$R_{2m} = \theta_m \frac{B_{2m+2}}{(2m+2)!} f^{(2m+1)}(x)|_a^b, \quad \text{pour un certain } 0 \leq \theta_m \leq 1; \quad (9.80)$$

autrement dit, le terme résiduel se situe alors entre 0 et le *premier terme supprimé* dans (9.78), c'est-à-dire celui qui suivrait le dernier terme si on incrémentait  $m$ .

En voici la preuve : la formule de sommation d'Euler est vraie pour tout  $m$ , et  $B_{2m+1} = 0$  lorsque  $m > 0$ ; donc  $R_{2m} = R_{2m+1}$ , et le premier terme supprimé est

$$R_{2m} - R_{2m+2}.$$

Par conséquent, il nous faut prouver que  $R_{2m}$  se trouve entre 0 et  $R_{2m} - R_{2m+2}$ ; c'est vrai si et seulement si  $R_{2m}$  et  $R_{2m+2}$  sont de signes opposés. Nous allons montrer que

$$f^{(2m+2)}(x) \geq 0 \quad \text{pour } a \leq x \leq b \quad \text{implique} \quad (-1)^m R_{2m} \geq 0. \quad (9.81)$$

Cette propriété, associée à (9.79), démontrera que  $R_{2m}$  et  $R_{2m+2}$  sont de signes opposés, et nous aurons terminé la preuve de (9.80).

Il n'est pas difficile de prouver (9.81) si on se souvient de la définition de  $R_{2m+1}$  et des faits que nous avons établis concernant le graphe de  $B_{2m+1}(x)$ . On a

$$R_{2m} = R_{2m+1} = \int_a^b \frac{B_{2m+1}(\{x\})}{(2m+1)!} f^{(2m+1)}(x) dx,$$

et  $f^{(2m+1)}(x)$  est croissante car sa dérivée  $f^{(2m+2)}(x)$  est positive. Le graphe de  $B_{2m+1}(\{x\})$  est très proche d'une sinusoïde multipliée par  $(-1)^{m+1}$ ; il est géométriquement évident que, si on multiplie cette courbe par une fonction croissante, la seconde moitié de la sinusoïde a plus d'importance que la première moitié. C'est pourquoi  $(-1)^m R_{2m+1} \geq 0$ , comme nous le voulions. L'exercice 16 est consacré à une preuve formelle de ce résultat.

## 9.6 DERNIÈRES SOMMATIONS

Avant de conclure ce livre, nous allons appliquer la formule de sommation d'Euler à quelques exemples intéressants et importants.

### *Sommation 1 : trop facile.*

Nous allons cependant commencer par un exemple intéressant mais *pas important du tout*, c'est-à-dire une somme que nous connaissons déjà. Voyons ce que donne la formule de sommation d'Euler si on l'applique à la somme télescopique

$$S_n = \sum_{1 \leq k < n} \frac{1}{k(k+1)} = \sum_{1 \leq k < n} \left( \frac{1}{k} - \frac{1}{k+1} \right) = 1 - \frac{1}{n}.$$

De toutes façons, cet exemple ne pourra pas nous faire de mal.

Commençons par développer la fonction  $f(x) = 1/x(x+1)$  en éléments simples,

$$f(x) = \frac{1}{x} - \frac{1}{x+1},$$

pour faciliter les opérations d'intégration et de dérivation. On trouve ainsi  $f'(x) = -1/x^2 + 1/(x+1)^2$ ,  $f''(x) = 2/x^3 - 2/(x+1)^3$  et, plus généralement,

$$f^{(k)}(x) = (-1)^k k! \left( \frac{1}{x^{k+1}} - \frac{1}{(x+1)^{k+1}} \right), \quad \text{pour } k \geq 0.$$

De plus,

$$\int_1^n f(x) dx = \ln x - \ln(x+1) \Big|_1^n = \ln \frac{2n}{n+1}.$$

Injectons tout cela dans la formule de sommation (9.67) :

$$\begin{aligned} S_n &= \ln \frac{2n}{n+1} - \sum_{k=1}^m (-1)^k \frac{B_k}{k} \left( \frac{1}{n^k} - \frac{1}{(n+1)^k} - 1 + \frac{1}{2^k} \right) + R_m(n), \\ \text{où } R_m(n) &= - \int_1^n B_m(x) \left( \frac{1}{x^{m+1}} - \frac{1}{(x+1)^{m+1}} \right) dx. \end{aligned}$$

Par exemple, si  $m = 4$ , le membre droit est égal à

$$\begin{aligned} \ln \frac{2n}{n+1} - \frac{1}{2} \left( \frac{1}{n} - \frac{1}{n+1} - \frac{1}{2} \right) - \frac{1}{12} \left( \frac{1}{n^2} - \frac{1}{(n+1)^2} - \frac{3}{4} \right) \\ + \frac{1}{120} \left( \frac{1}{n^4} - \frac{1}{(n+1)^4} - \frac{15}{16} \right) + R_4(n). \end{aligned}$$

Quelle pagaille ! Reconnaissons que c'est assez éloigné de la bonne réponse,  $1 - n^{-1}$ . Continuons quand même, juste pour voir. Nous pouvons développer les termes du membre droit en puissances négatives de  $n$ , disons jusqu'à  $O(n^{-5})$  :

$$\begin{aligned} \ln \frac{n}{n+1} &= -n^{-1} + \frac{1}{2}n^{-2} - \frac{1}{3}n^{-3} + \frac{1}{4}n^{-4} + O(n^{-5}); \\ \frac{1}{n+1} &= n^{-1} - n^{-2} + n^{-3} - n^{-4} + O(n^{-5}); \\ \frac{1}{(n+1)^2} &= n^{-2} - 2n^{-3} + 3n^{-4} + O(n^{-5}); \\ \frac{1}{(n+1)^4} &= n^{-4} + O(n^{-5}). \end{aligned}$$

Par conséquent, le membre droit de notre approximation s'écrit

$$\begin{aligned} \ln 2 + \frac{1}{4} + \frac{1}{16} - \frac{1}{128} + (-1 - \frac{1}{2} + \frac{1}{2})n^{-1} + (\frac{1}{2} - \frac{1}{2} - \frac{1}{12} + \frac{1}{12})n^{-2} \\ + (-\frac{1}{3} + \frac{1}{2} - \frac{2}{12})n^{-3} + (\frac{1}{4} - \frac{1}{2} + \frac{3}{12} + \frac{1}{120} - \frac{1}{120})n^{-4} + R_4(n) \\ = \ln 2 + \frac{39}{128} - n^{-1} + R_4(n) + O(n^{-5}). \end{aligned}$$

Les coefficients de  $n^{-2}$ ,  $n^{-3}$  et  $n^{-4}$  se sont mutuellement neutralisés, exactement comme il fallait.

Si nous vivions dans un monde parfait, nous pourrions montrer que  $R_4(n)$  est asymptotiquement très petit, peut-être en  $O(n^{-5})$ , et nous aurions ainsi une approximation de la somme. Mais c'est impossible, car il se trouve que nous savons que le terme constant est égal à 1 et non à  $\ln 2 + \frac{39}{128}$  (qui vaut à peu près 0,9978). Par conséquent,  $R_4(n)$  est en fait égal à  $\frac{89}{128} - \ln 2 + O(n^{-5})$ , et la formule de sommation d'Euler est incapable de nous le dire. Autrement dit, nous avons échoué.

Voici peut être un moyen d'arranger les choses. Remarquons que les termes constants de l'approximation forment un motif :

$$\ln 2 - \frac{1}{2}B_1 + \frac{1}{2} \cdot \frac{3}{4}B_2 - \frac{1}{3} \cdot \frac{7}{8}B_3 + \frac{1}{4} \cdot \frac{15}{16}B_4 - \frac{1}{5} \cdot \frac{31}{32}B_5 + \dots$$

Si on arrive à montrer que cette série tend vers 1 lorsque le nombre de termes tend vers l'infini, ce sera gagné. Hélas, ce n'est pas possible : les nombres de Bernoulli deviennent extrêmement grands. Par exemple,  $B_{22} = \frac{854513}{138} > 6192$  ; par conséquent,  $|R_{22}(n)|$  sera bien plus grand que  $|R_4(n)|$ . Nous avons encore échoué.

Nous allons quand même pouvoir nous en sortir, grâce à une méthode qui pourra être utile aussi en d'autres occasions. Remarquons que  $R_4(n)$  tend vers une limite bien définie lorsque  $n \rightarrow \infty$  :

$$\lim_{n \rightarrow \infty} R_4(n) = - \int_1^{\infty} B_4(\{x\}) \left( \frac{1}{x^5} - \frac{1}{(x+1)^5} \right) dx = R_4(\infty).$$

L'intégrale  $\int_1^{\infty} B_m(\{x\}) f^{(m)}(x) dx$  existe à condition que  $f^{(m)}(x) = O(x^{-2})$  lorsque  $x \rightarrow \infty$ . Dans notre cas,  $f^{(4)}(x)$  satisfait cette condition. De plus,

$$\begin{aligned} R_4(n) &= R_4(\infty) + \int_n^{\infty} B_4(\{x\}) \left( \frac{1}{x^5} - \frac{1}{(x+1)^5} \right) dx \\ &= R_4(\infty) + O\left(\int_n^{\infty} x^{-6} dx\right) = R_4(\infty) + O(n^{-5}). \end{aligned}$$

Ainsi, la formule de sommation d'Euler nous a permis de prouver que

$$\begin{aligned} \sum_{1 \leq k < n} \frac{1}{k(k+1)} &= \ln 2 + \frac{39}{128} - n^{-1} + R_4(\infty) + O(n^{-5}) \\ &= C - n^{-1} + O(n^{-5}) \end{aligned}$$

pour une certaine constante  $C$ . Nous ne connaissons pas cette constante, mais nous savons qu'elle existe (pour trouver sa valeur, il faut appliquer une autre méthode).

Si nous avions choisi un  $m$  plus grand, le même raisonnement aurait donné

$$R_m(n) = R_m(\infty) + O(n^{-m-1}),$$

et nous aurions trouvé la formule

$$\sum_{1 \leq k < n} \frac{1}{k(k+1)} = C - n^{-1} + c_2 n^{-2} + c_3 n^{-3} + \dots + c_m n^{-m} + O(n^{-m-1})$$

pour certaines constantes  $c_2, c_3, \dots$ . Nous savons que, dans notre cas, les  $c_k$  sont tous nuls ; nous allons quand même le prouver, juste pour

reprendre un peu confiance en la formule d'Euler. Le terme  $\ln \frac{n}{n+1}$  contribue pour  $(-1)^m/m$  à  $c_m$ ; le terme  $(-1)^{m+1}(B_m/m)n^{-m}$  y contribue pour  $(-1)^{m+1}B_m/m$ ; enfin, le terme  $(-1)^k(B_k/k)(n+1)^{-k}$  y contribue pour  $(-1)^m \binom{m-1}{k-1} B_k/k$ . Par conséquent,

$$\begin{aligned} (-1)^m c_m &= \frac{1}{m} - \frac{B_m}{m} + \sum_{k=1}^m \binom{m-1}{k-1} \frac{B_k}{k} \\ &= \frac{1}{m} - \frac{B_m}{m} + \frac{1}{m} \sum_{k=1}^m \binom{m}{k} B_k \\ &= \frac{1}{m} (1 - B_m + B_m(1) - 1). \end{aligned}$$

C'est évidemment nul lorsque  $m > 1$ . Nous venons donc de prouver que

$$\sum_{1 \leq k < n} \frac{1}{k(k+1)} = C - n^{-1} + O(n^{-m-1}), \quad \text{pour tout } m \geq 1. \quad (9.82)$$

Cela ne veut pas dire que la somme est exactement égale à  $C - n^{-1}$ ; elle pourrait valoir, par exemple, quelque chose comme  $C - n^{-1} + 2^{-n}$ . Toutefois, la formule de sommation d'Euler nous donne en quelque sorte un majorant de l'erreur, en indiquant qu'elle est en  $O(n^{-m-1})$ .

### *Sommation 1, bis : récapitulation et généralisation.*

Regardons avec un peu plus de recul ce que nous venons de faire. Nous sommes partis d'une somme  $S_n = \sum_{1 \leq k < n} f(k)$  et nous avons appliqué la formule de sommation d'Euler pour écrire

$$S_n = F(n) - F(1) + \sum_{k=1}^m (T_k(n) - T_k(1)) + R_m(n), \quad (9.83)$$

où  $F(x)$  était égal à  $\int f(x) dx$  et  $T_k(x)$  était un terme dépendant de  $B_k$  et de  $f^{(k-1)}(x)$ . Nous avons aussi remarqué qu'il existait une constante  $c$  telle que  $f^{(m)}(x) = O(x^{c-m})$  lorsque  $x \rightarrow \infty$ , pour tout  $m$  assez grand. (Plus précisément, on avait  $f(k) = 1/k(k+1)$ ;  $F(x) = \ln(x/(x+1))$ ;  $c = -2$ ;  $T_k(x) = (-1)^{k+1}(B_k/k)(x^{-k} - (x+1)^{-k})$ ). Cela impliquait que, pour toute valeur de  $m$  assez grande, les queues des restes étaient petites :

$$\begin{aligned} R'_m(n) &= R_m(\infty) - R_m(n) \\ &= (-1)^{m+1} \int_n^\infty \frac{B_m(\{x\})}{m!} f^{(m)}(x) dx = O(n^{c+1-m}). \end{aligned} \quad (9.84)$$

Nous avons donc pu en déduire qu'il existait une constante  $C$  telle que

$$S_n = F(n) + C + \sum_{k=1}^m T_k(n) - R'_m(n). \quad (9.85)$$

## 508 CALCUL ASYMPTOTIQUE

(De plus, C avait la bonne idée d'absorber les termes  $T_k(1)$  qui nous ennuiaient). Pour nous économiser du travail par la suite, déclarons tout de suite que, pour que C existe, il suffit que  $R_m(\infty)$  existe.

Supposons maintenant que  $f^{(2m+2)}(x) \geq 0$  et  $f^{(2m+4)}(x) \geq 0$  pour tout  $1 \leq x \leq n$ . Nous avons montré en (9.80) que cela entraîne une majoration simple du reste,

$$R_{2m}(n) = \theta_{m,n} (T_{2m+2}(n) - T_{2m+2}(1)),$$

où  $\theta_{m,n}$  est quelque part entre 0 et 1. Cependant, les majorants en fonction de  $R_{2m}(n)$  et  $T_{2m+2}(1)$  ne nous intéressent pas. Ce qu'il nous faut, c'est une majoration du genre

$$-R'_{2m}(n) = \phi_{m,n} T_{2m+2}(n),$$

où  $0 < \phi_{m,n} < 1$ ; cela nous permettra de conclure de (9.85) que

$$S_n = F(n) + C + T_1(n) + \sum_{k=1}^m T_{2k}(n) + \phi_{m,n} T_{2m+2}(n), \quad (9.86)$$

et donc que le reste sera vraiment entre zéro et le premier terme supprimé.

Avec une petite modification de l'argument que nous avons utilisé précédemment, tout va coller parfaitement. Supposons que

$$f^{(2m+2)}(x) \geq 0 \quad \text{et} \quad f^{(2m+4)}(x) \geq 0, \quad \text{lorsque } x \rightarrow \infty. \quad (9.87)$$

Le membre droit de (9.85) est exactement comme l'opposé du membre droit de la formule de sommation d'Euler (9.67) avec  $a = n$  et  $b = \infty$ , du moins jusqu'à ce que les termes résiduels apparaissent; ces termes résiduels peuvent être trouvés par induction sur  $m$ . Nous pouvons donc appliquer notre précédent argument.

### *Sommation 2 : harmonisons les nombres harmoniques.*

Voyons maintenant un exemple non trivial : appliquons la formule de sommation d'Euler pour prouver la formule d'approximation de  $H_n$ , que nous connaissons déjà sans l'avoir encore démontrée. Dans ce cas,  $f(x) = 1/x$ . Nous avons déjà calculé, pour la sommation 1, l'intégrale et les dérivées de  $f$ , et nous savons que  $f^{(m)}(x) = O(x^{-m-1})$  lorsque  $x \rightarrow \infty$ . Nous pouvons donc immédiatement injecter tout cela dans la formule (9.85) :

$$\sum_{1 \leq k < n} \frac{1}{k} = \ln n + C + B_1 n^{-1} - \sum_{k=1}^m \frac{B_{2k}}{2kn^{2k}} - R'_{2m}(n),$$

pour une certaine constante  $C$ . Remarquez que la somme du membre gauche est égale à  $H_{n-1}$ , non à  $H_n$ , car, plutôt que de s'ennuyer avec des  $(n+1)$  dans

le membre droit, il vaut mieux travailler avec  $H_{n-1}$  et ajouter  $1/n$  plus tard. Le  $B_1 n^{-1}$  deviendra alors  $(B_1 + 1)n^{-1} = 1/(2n)$ . Convenons d'appeler notre constante  $\gamma$  au lieu de  $C$ , puisque, par définition, la constante d'Euler  $\gamma$  est égale à  $\lim_{n \rightarrow \infty} (H_n - \ln n)$ .

On peut estimer le terme résiduel avec la méthode que nous avons vue récemment, car  $f^{(2m)}(x) = (2m)!/x^{2m+1} \geq 0$  pour tout  $x > 0$ . Par conséquent, d'après (9.86),

$$H_n = \ln n + \gamma + \frac{1}{2n} - \sum_{k=1}^m \frac{B_{2k}}{2kn^{2k}} - \theta_{m,n} \frac{B_{2m+2}}{(2m+2)n^{2m+2}}, \quad (9.88)$$

où  $\theta_{m,n}$  est une certaine fraction entre 0 et 1. Nous retrouvons ainsi la formule générale dont les premiers termes sont écrits dans la table 479. Par exemple, pour  $m = 2$ , on obtient

$$H_n = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \frac{\theta_{2,n}}{252n^6}. \quad (9.89)$$

Notons en passant que cette équation nous donne une bonne approximation de  $\gamma$ , même si  $n = 2$  :

$$\gamma = H_2 - \ln 2 - \frac{1}{4} + \frac{1}{48} - \frac{1}{1920} + \epsilon = 0,577165\dots + \epsilon,$$

où  $\epsilon$  est compris entre zéro et  $\frac{1}{16128}$ . Si on prend  $n = 10^4$  et  $m = 250$ , on trouve une valeur dont les 1271 premières décimales sont correctes :

$$\gamma = 0,57721\,56649\,01532\,86060\,65120\,90082\,40243\dots. \quad (9.90)$$

Il existe d'autres formules qui permettent d'évaluer la constante d'Euler plus efficacement encore [345].

### **Sommation 3 : la formule de Stirling.**

Si  $f(x) = \ln x$ , alors  $f'(x) = 1/x$ ; le calcul de la somme des logarithmes peut donc s'effectuer presque comme celui du nombre harmonique. D'après la formule de sommation d'Euler, on a

$$\begin{aligned} \sum_{1 \leq k \leq n} \ln k &= n \ln n - n + \sigma - \frac{\ln n}{2} \\ &\quad + \sum_{k=1}^m \frac{B_{2k}}{2k(2k-1)n^{2k-1}} + \varphi_{m,n} \frac{B_{2m+2}}{(2m+2)(2m+1)n^{2m+1}} \end{aligned}$$

où  $\sigma$  est une certaine constante, la "constante de Stirling", et  $0 < \varphi_{m,n} < 1$ . (Dans ce cas,  $f^{(2m)}(x)$  est négatif au lieu d'être positif, mais on peut toujours dire que le terme résiduel est borné par le premier terme supprimé, car nous

## 510 CALCUL ASYMPTOTIQUE

aurions pu partir de  $f(x) = -\ln x$  au lieu de  $f(x) = \ln x$ ). En ajoutant  $\ln n$  aux deux côtés, on obtient

$$\ln n! = n \ln n - n + \frac{\ln n}{2} + \sigma + \frac{1}{12n} - \frac{1}{360n^3} + \frac{\varphi_{2,n}}{1260n^5} \quad (9.91)$$

lorsque  $m = 2$ . En appliquant l'exponentielle aux deux membres, on retrouve l'approximation donnée dans la table 479. (En fait,  $e^\sigma$  est égal à  $\sqrt{2\pi}$ , mais nous ne sommes pas encore capables de le prouver. Stirling ne trouva la forme close de  $\sigma$  que plusieurs années après que de Moivre [76] eut prouvé que cette constante existait).

Si  $m$  est fixé et si  $n \rightarrow \infty$ , la formule générale donne une approximation de  $\ln n!$  de plus en plus précise en termes d'erreur absolue, donc une approximation de  $n!$  de plus en plus précise en termes d'erreur relative. En revanche, si  $n$  est fixé et si  $m$  augmente, le majorant de l'erreur,  $|B_{2m+2}|/(2m+2)(2m+1)n^{2m+1}$ , diminue jusqu'à un certain point, puis se met à augmenter. Il existe donc un point au-delà duquel, en vertu d'une sorte de "principe d'incertitude", la précision de l'approximation de  $n!$  ne peut pas être améliorée.

Dans l'équation (5.83) du chapitre 5, nous avons généralisé la fonction factorielle à tout nombre réel avec la définition

$$\frac{1}{\alpha!} = \lim_{n \rightarrow \infty} \left( \frac{n+\alpha}{n} \right)^n n^{-\alpha}$$

suggérée par Euler. Supposons que  $\alpha$  est un grand nombre ; alors

$$\ln \alpha! = \lim_{n \rightarrow \infty} \left( \alpha \ln n + \ln n! - \sum_{k=1}^n \ln(\alpha+k) \right),$$

et on peut appliquer la formule de sommation d'Euler, avec  $f(x) = \ln(x+\alpha)$ , pour estimer cette somme :

$$\begin{aligned} \sum_{k=1}^n \ln(k+\alpha) &= F_m(\alpha, n) - F_m(\alpha, 0) + R_{2m}(\alpha, n), \\ F_m(\alpha, x) &= (x+\alpha) \ln(x+\alpha) - x + \frac{\ln(x+\alpha)}{2} \\ &\quad + \sum_{k=1}^m \frac{B_{2k}}{2k(2k-1)(x+\alpha)^{2k-1}}, \\ R_{2m}(\alpha, n) &= \int_0^n \frac{B_{2m}(\{x\})}{2m} \frac{dx}{(x+\alpha)^{2m}}. \end{aligned}$$

(Nous avons utilisé (9.67) avec  $a = 0$  et  $b = n$ , puis ajouté  $\ln(n+\alpha) - \ln \alpha$  dans les deux membres). En soustrayant ce résultat à l'approximation de

*Heisenberg a dû passer par là.*

$\ln n!$  donnée par la formule de Stirling, puis en ajoutant  $\alpha \ln n$  et en prenant la limite lorsque  $n \rightarrow \infty$ , on obtient

$$\begin{aligned}\ln \alpha! &= \alpha \ln \alpha - \alpha + \frac{\ln \alpha}{2} + \sigma \\ &\quad + \sum_{k=1}^m \frac{B_{2k}}{(2k)(2k-1)\alpha^{2k-1}} - \int_0^\infty \frac{B_{2m}(x)}{2m} \frac{dx}{(x+\alpha)^{2m}},\end{aligned}$$

car  $\alpha \ln n + n \ln n - n + \frac{1}{2} \ln n - (n+\alpha) \ln(n+\alpha) + n - \frac{1}{2} \ln(n+\alpha) \rightarrow -\alpha$  et les autres termes, que nous n'avons pas écrits ici, tendent vers zéro. La formule de Stirling est donc valable aussi bien pour la factorielle généralisée (et pour la fonction Gamma  $\Gamma(\alpha+1) = \alpha!$ ) que pour la factorielle ordinaire.

#### Sommation 4 : un terme en cloche.

Passons à une somme d'allure tout à fait différente :

$$\begin{aligned}\Theta_n &= \sum_k e^{-k^2/n} \\ &= \dots + e^{-9/n} + e^{-4/n} + e^{-1/n} + 1 + e^{-1/n} + e^{-4/n} + e^{-9/n} + \dots\end{aligned}\tag{9.92}$$

C'est une somme doublement infinie, dont le terme maximum est  $e^0 = 1$ , celui pour lequel  $k = 0$ . Les séries de ce genre, constituées de termes  $e^{-k^2/n}$  élevés à la puissance  $p(k)$ , où  $p(k)$  est un polynôme de degré 2, sont traditionnellement appelées des "fonctions thêta" ; c'est pourquoi nous notons  $\Theta_n$  cette fonction particulière. Si  $n = 10^{100}$ , on a

$$e^{-k^2/n} = \begin{cases} e^{-0,01} \approx 0,99005, & \text{si } k = 10^{49}; \\ e^{-1} \approx 0,36788, & \text{si } k = 10^{50}; \\ e^{-100} < 10^{-43}, & \text{si } k = 10^{51}. \end{cases}$$

Le terme général reste donc très proche de 1 jusqu'à ce que  $k$  atteigne à peu près  $\sqrt{n}$  ; puis il diminue très vite pour rester proche de zéro. On peut raisonnablement conjecturer que  $\Theta_n$  sera proportionnel à  $\sqrt{n}$ . Voici le graphe de  $e^{-k^2/n}$  lorsque  $n = 10$  :



Le graphe est similaire pour toute valeur de  $n$ .

Pour estimer  $\Theta_n$ , on peut poser  $f(x) = e^{-x^2/n}$  et prendre  $a = -\infty$  et  $b = +\infty$  dans la formule de sommation d'Euler (si l'infini vous effraie, prenez  $a = -A$  et  $b = +B$ , puis considérez les limites lorsque  $A, B \rightarrow \infty$ ). Si on remplace  $x$  par  $u\sqrt{n}$ , l'intégrale de  $f(x)$  est

$$\int_{-\infty}^{+\infty} e^{-x^2/n} dx = \sqrt{n} \int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{n} C.$$

## 512 CALCUL ASYMPTOTIQUE

La valeur de  $\int_{-\infty}^{+\infty} e^{-u^2} du$  est bien connue, mais notons la  $C$  pour l'instant ; nous y reviendrons lorsque nous en aurons fini avec la formule de sommation d'Euler.

Maintenant, il nous faut connaître la suite des dérivées  $f'(x)$ ,  $f''(x)$  etc. Pour cela, posons

$$f(x) = g(x/\sqrt{n}), \quad g(x) = e^{-x^2}.$$

Alors nous pouvons écrire que

$$\frac{df(x)}{dx} = \frac{dg(y)}{dy} \frac{dy}{dx}, \quad y = \frac{x}{\sqrt{n}},$$

ce qui revient à dire que

$$f'(x) = \frac{1}{\sqrt{n}} g'(x/\sqrt{n}).$$

Par induction, on trouve que

$$f^{(k)}(x) = n^{-k/2} g^{(k)}(x/\sqrt{n}).$$

Par exemple,  $g'(x) = -2xe^{-x^2}$  et  $g''(x) = (4x^2 - 2)e^{-x^2}$ , donc

$$f'(x) = \frac{1}{\sqrt{n}} \left( -2 \frac{x}{\sqrt{n}} \right) e^{-x^2/n}, \quad f''(x) = \frac{1}{n} \left( 4 \left( \frac{x}{\sqrt{n}} \right)^2 - 2 \right) e^{-x^2/n}.$$

On voit plus facilement ce qui se passe si on travaille avec la fonction plus simple  $g(x)$ .

Nous n'avons pas besoin de calculer exactement les dérivées de  $g(x)$ , car seules leur limites lorsque  $x = \pm\infty$  nous intéressent. Pour les trouver, il suffit de remarquer que toute dérivée de  $g(x)$  est égale à  $e^{-x^2}$  multiplié par un polynôme en  $x$  :

$$g^{(k)}(x) = P_k(x)e^{-x^2}, \quad \text{où } P_k \text{ est un polynôme de degré } k.$$

Cela se démontre par induction.

Lorsque  $x \rightarrow \pm\infty$ , l'exponentielle tend vers zéro beaucoup plus vite que  $P_k(x)$ . Pour cette raison, on a

$$f^{(k)}(+\infty) = f^{(k)}(-\infty) = 0$$

pour tout  $k \geq 0$ . Par conséquent, tous les termes

$$\sum_{k=1}^m \frac{B_k}{k!} f^{(k-1)}(x) \Big|_{-\infty}^{+\infty}$$

disparaissent, et il ne nous reste plus que le terme qui provient de  $\int f(x) dx$  et le terme résiduel :

$$\begin{aligned}\Theta_n &= C\sqrt{n} + (-1)^{m+1} \int_{-\infty}^{+\infty} \frac{B_m(\{x\})}{m!} f^{(m)}(x) dx \\ &= C\sqrt{n} + \frac{(-1)^{m+1}}{n^{m/2}} \int_{-\infty}^{+\infty} \frac{B_m(\{x\})}{m!} g^{(m)}\left(\frac{x}{\sqrt{n}}\right) dx \\ &= C\sqrt{n} + \frac{(-1)^{m+1}}{n^{(m-1)/2}} \int_{-\infty}^{+\infty} \frac{B_m(\{u\sqrt{n}\})}{m!} P_m(u)e^{-u^2} du \\ &= C\sqrt{n} + O(n^{(1-m)/2}).\end{aligned}$$

Le  $O(n^{(1-m)/2})$  vient du fait que  $|B_m(\{u\sqrt{n}\})|$  est majoré et que l'intégrale  $\int_{-\infty}^{+\infty} |P(u)|e^{-u^2} du$  existe pour tout polynôme  $P$ . Notez que la constante liée à ce  $O$  dépend de  $m$ .

Nous avons démontré que  $\Theta_n = C\sqrt{n} + O(n^{-M})$ , pour un  $M$  aussi grand que l'on veut ; la différence entre  $\Theta_n$  et  $C\sqrt{n}$  est "exponentiellement petite". Nous allons donc déterminer cette constante  $C$  qui joue un si grand rôle dans la valeur de  $\Theta_n$ .

Une façon de la déterminer consisterait à regarder dans une table. Nous préférions bien entendu la calculer nous-mêmes. Pour cela, il suffit de considérer l'intégrale double

$$C^2 = \int_{-\infty}^{+\infty} e^{-x^2} dx \int_{-\infty}^{+\infty} e^{-y^2} dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dx dy.$$

En convertissant en coordonnées polaires, on obtient

$$\begin{aligned}C^2 &= \int_0^{2\pi} \int_0^\infty e^{-r^2} r dr d\theta \\ &= \frac{1}{2} \int_0^{2\pi} d\theta \int_0^\infty e^{-u} du \\ &= \frac{1}{2} \int_0^{2\pi} d\theta = \pi.\end{aligned}$$

Ainsi,  $C = \sqrt{\pi}$ . L'équation  $x^2 + y^2 = r^2$  est l'équation d'un cercle de circonférence  $2\pi r$  ; cela explique l'apparition de  $\pi$  dans notre constante.

Il y a une autre façon de calculer  $C$ , qui consiste à remplacer  $x$  par  $\sqrt{t}$  et  $dx$  par  $\frac{1}{2}t^{-1/2} dt$  :

$$C = \int_{-\infty}^{+\infty} e^{-x^2} dx = 2 \int_0^\infty e^{-x^2} dx = \int_0^\infty t^{-1/2} e^{-t} dt.$$

Cette intégrale est égale à  $\Gamma\left(\frac{1}{2}\right)$  car, selon (5.84),  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ . Nous avons donc démontré que  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ .

Voici donc finalement notre formule :

$$\Theta_n = \sum_k e^{-k^2/n} = \sqrt{\pi n} + O(n^{-M}), \quad \text{pour tout } M \text{ fixé.} \quad (9.93)$$

C'est parce que la constante liée au  $O$  dépend de  $M$  que nous disons que  $M$  est "fixé".

Lorsque  $n = 2$ , par exemple, la somme infinie  $\Theta_2$  est à peu près égale à 2,506628288, ce qui est déjà très proche de  $\sqrt{2\pi} \approx 2,506628275$ . La valeur de  $\Theta_{100}$  a 427 décimales correctes si on la compare à  $10\sqrt{\pi}$ . Dans l'exercice 59, on considère une série qui converge rapidement vers  $\Theta_n$  ; on trouve ainsi que

$$\Theta_n/\sqrt{\pi n} = 1 + 2e^{-n\pi^2} + O(e^{-4n\pi^2}). \quad (9.94)$$

### **Sommation 5 : la der des der.**

Le temps est venu de calculer notre dernière somme, qui va nous donner la valeur de  $\sigma$ , la constante de Stirling. Elle va aussi nous permettre d'appliquer un bon nombre de techniques que nous avons apprises dans ce chapitre, et même dans tout ce livre. C'est donc exactement la somme qui convient pour terminer notre exploration des Mathématiques Concètes.

Notre problème peut sembler trop facile au premier abord : nous allons chercher la valeur de

$$A_n = \sum_k \binom{2n}{k}$$

en appliquant la formule de sommation d'Euler.

C'est encore un cas où nous connaissons la réponse ; mais nous savons aussi qu'il est toujours intéressant d'appliquer des méthodes neuves à des vieux problèmes ; on peut ainsi découvrir des choses nouvelles.

Commençons à PENSER GRAND pour réaliser que ce sont les termes du milieu, là où  $k$  est proche de  $n$ , qui contribuent le plus à  $A_n$ . Il est presque toujours payant de modifier une somme de façon que les termes les plus importants se retrouvent au début ; il est plus facile alors de se débarrasser des termes pour lesquels  $|k|$  est grand. Remplaçons donc  $k$  par  $n + k$  :

$$A_n = \sum_k \binom{2n}{n+k} = \sum_k \frac{(2n)!}{(n+k)!(n-k)!}.$$

Les choses se présentent bien, puisque nous savons approximer  $(n \pm k)!$  lorsque  $n$  est grand et  $k$  petit.

Maintenant, nous allons appliquer notre truc numéro 2, la procédure en trois étapes que nous avons apprise dans ce chapitre. Il faut que nous arrivions à écrire

$$\frac{(2n)!}{(n+k)!(n-k)!} = a_k(n) = b_k(n) + O(c_k(n)), \quad \text{pour } k \in D_n,$$

pour obtenir l'approximation

$$A_n = \sum_k b_k(n) + O\left(\sum_{k \notin D_n} a_k(n)\right) + O\left(\sum_{k \notin D_n} b_k(n)\right) + \sum_{k \in D_n} O(c_k(n)).$$

Nous allons donc essayer d'estimer  $\binom{2n}{n+k}$  lorsque  $|k|$  est petit. Nous pourrions utiliser la formule de Stirling de la table 479, mais il vaut mieux travailler avec son équivalent logarithmique (9.91) :

$$\begin{aligned} \ln a_k(n) &= \ln(2n)! - \ln(n+k)! - \ln(n-k)! \\ &= 2n \ln 2n - 2n + \frac{1}{2} \ln 2n + \sigma + O(n^{-1}) \\ &\quad - (n+k) \ln(n+k) + n+k - \frac{1}{2} \ln(n+k) - \sigma + O((n+k)^{-1}) \\ &\quad - (n-k) \ln(n-k) + n-k - \frac{1}{2} \ln(n-k) - \sigma + O((n-k)^{-1}). \end{aligned} \tag{9.95}$$

Nous voudrions transformer cela en une jolie estimation par un  $O$ .

Avec la méthode en trois étapes, on peut se permettre de travailler avec des estimations qui ne sont valables que lorsque  $k$  se trouve dans l'ensemble "dominant"  $D_n$ . La question est de savoir choisir ce  $D_n$ . Il faut qu'il soit assez petit pour que l'estimation soit bonne ; en particulier, il vaut mieux éviter que  $k$  ne soit trop proche de  $n$ , sinon le terme  $O((n-k)^{-1})$  de (9.95) sera trop gros. Il faut cependant que  $D_n$  soit assez grand pour que les termes de la queue (ceux pour lesquels  $k \notin D_n$ ) soient négligeables par rapport à la somme totale. En général, c'est par essais successifs qu'on trouve l'ensemble  $D_n$  qui convient. Dans le problème qui nous préoccupe, les calculs que nous allons faire confirmeront que le bon choix consiste à poser

$$k \in D_n \iff |k| \leq n^{1/2+\epsilon}. \tag{9.96}$$

Ici,  $\epsilon$  est une petite constante positive que nous préciserons plus tard (nos approximations en  $O$  dépendront de  $\epsilon$ ). L'équation (9.95) se réduit maintenant à

$$\begin{aligned} \ln a_k(n) &= (2n + \frac{1}{2}) \ln 2 - \sigma - \frac{1}{2} \ln n + O(n^{-1}) \\ &\quad - (n+k + \frac{1}{2}) \ln(1+k/n) - (n-k + \frac{1}{2}) \ln(1-k/n). \end{aligned} \tag{9.97}$$

*En fait, je ne domine pas tellement, là.*

## 516 CALCUL ASYMPTOTIQUE

(Nous nous sommes débarrassés d'un certain nombre de logarithmes en écrivant  $\ln(n \pm k) = \ln n + \ln(1 \pm k/n)$ ).

Maintenant, il nous faut faire un développement asymptotique de  $\ln(1 \pm k/n)$ , en s'arrêtant lorsque le terme d'erreur tendra vers zéro lorsque  $n \rightarrow \infty$ . Comme  $\ln(1 \pm k/n)$  est multiplié par  $(n \pm k + \frac{1}{2})$ , nous n'avons qu'à développer le logarithme jusqu'à atteindre  $O(n^{-1})$ , en utilisant le fait que  $|k| \leq n^{1/2+\epsilon}$  :

$$\ln\left(1 \pm \frac{k}{n}\right) = \pm \frac{k}{n} - \frac{k^2}{2n^2} + O(n^{-3/2+3\epsilon}).$$

En multipliant par  $n \pm k + \frac{1}{2}$ , on trouve

$$\pm k - \frac{k^2}{2n} + \frac{k^2}{n} + O(n^{-1/2+3\epsilon}),$$

plus d'autres termes qui sont absorbés par le  $O(n^{-1/2+3\epsilon})$ . Ainsi, (9.97) devient

$$\ln a_k(n) = (2n + \frac{1}{2}) \ln 2 - \sigma - \frac{1}{2} \ln n - k^2/n + O(n^{-1/2+3\epsilon}).$$

Prenons l'exponentielle pour obtenir

$$a_k(n) = \frac{2^{2n+1/2}}{e^\sigma \sqrt{n}} e^{-k^2/n} (1 + O(n^{-1/2+3\epsilon})). \quad (9.98)$$

C'est l'approximation que nous cherchons, avec

$$b_k(n) = \frac{2^{2n+1/2}}{e^\sigma \sqrt{n}} e^{-k^2/n}, \quad c_k(n) = 2^{2n} n^{-1+3\epsilon} e^{-k^2/n}.$$

Remarquez que l'expression de  $k$  dans  $b_k(n)$  et  $c_k(n)$  est particulièrement simple. C'est une chance, car c'est sur  $k$  que nous allons sommer.

Selon notre méthode en trois étapes,  $\sum_k a_k(n)$  doit être à peu près égal à  $\sum_k b_k(n)$ , du moins si nous avons approximé comme il faut. Nous allons donc calculer

$$\begin{aligned} \sum_k b_k(n) &= \frac{2^{2n+1/2}}{e^\sigma \sqrt{n}} \sum_k e^{-k^2/n} \\ &= \frac{2^{2n+1/2}}{e^\sigma \sqrt{n}} \Theta_n = \frac{2^{2n} \sqrt{2\pi}}{e^\sigma} (1 + O(n^{-M})). \end{aligned}$$

(Encore un coup de chance : la somme  $\Theta_n$  de l'exemple précédent nous *Quelle coïncidence !* est bien utile ici). Ce résultat est encourageant car nous savons que notre somme d'origine est en fait

$$A_n = \sum_k \binom{2n}{k} = (1+1)^{2n} = 2^{2n},$$

ce qui semble confirmer que  $e^\sigma = \sqrt{2\pi}$ , comme nous le pensons.

Cependant, rien n'est encore sûr : il nous faut encore prouver que nos estimations sont suffisamment précises. Examinons tout d'abord l'erreur correspondant à  $c_k(n)$  :

$$\sum_{k \in D_n} 2^{2n} n^{-1+3\epsilon} e^{-k^2/n} \leq 2^{2n} n^{-1+3\epsilon} \Theta_n = O(2^{2n} n^{-\frac{1}{2}+3\epsilon}).$$

Parfait. C'est asymptotiquement plus petit que la somme précédente si  $3\epsilon < \frac{1}{2}$ .

Vérifions maintenant les queues des sommes. On a

$$\begin{aligned} \sum_{k > n^{1/2+\epsilon}} e^{-k^2/n} &< \exp(-[n^{1/2+\epsilon}]^2/n) (1 + e^{-1/n} + e^{-2/n} + \dots) \\ &= O(e^{-n^{2\epsilon}}) \cdot O(n), \end{aligned}$$

ce qui est en  $O(n^{-M})$  pour tout  $M$  ; donc  $\sum_{k \notin D_n} b_k(n)$  est asymptotiquement négligeable. (Nous coupons en  $n^{1/2+\epsilon}$  pour que les  $e^{-k^2/n}$  qui sont en dehors de  $D_n$  soient exponentiellement petits. D'autres choix auraient pu aussi convenir, comme  $n^{1/2} \log n$  par exemple ; les approximations auraient été un peu plus précises, mais les formules plus compliquées. N'oublions pas que notre but principal est de trouver la valeur de la constante  $\sigma$  ; nous n'avons donc pas besoin des meilleures approximations possibles). De même, l'autre queue

$$\sum_{k > n^{1/2+\epsilon}} \binom{2n}{n+k}$$

est bornée par  $2n$  fois son plus grand terme, qui apparaît au point de coupure  $k \approx n^{1/2+\epsilon}$ . Ce terme, à peu près égal à  $b_k(n)$ , est exponentiellement petit par rapport à  $A_n$ , et le facteur  $2n$  se trouve proprement anéanti.

Nous avons donc démontré que

$$2^{2n} = \sum_k \binom{2n}{k} = \frac{\sqrt{2\pi}}{e^\sigma} 2^{2n} + O(2^{2n} n^{-\frac{1}{2}+3\epsilon}), \text{ si } 0 < \epsilon < \frac{1}{6}. \quad (9.99)$$

*Merci de nous avoir lus, nous espérons que cela vous sera utile.*

— Les auteurs

CQFD.

## Exercices

### Echauffements

- 1 Prouvez ou réfutez : Si  $f_1(n) \prec g_1(n)$  et  $f_2(n) \prec g_2(n)$ , alors  $f_1(n) + f_2(n) \prec g_1(n) + g_2(n)$ .
- 2 Quelle fonction croît le plus vite :
  - a  $n^{(\ln n)}$  ou  $(\ln n)^n$  ?
  - b  $n^{(\ln \ln \ln n)}$  ou  $(\ln n)!$  ?
  - c  $(n!)!$  ou  $((n-1)!)!(n-1)!^{n!}$  ?
  - d  $F_{[H_n]}^2$  ou  $H_{F_n}$  ?
- 3 Qu'est-ce qui cloche dans l'argument suivant ? "Puisque  $n = O(n)$ ,  $2n = O(n)$  et ainsi de suite, on a  $\sum_{k=1}^n kn = \sum_{k=1}^n O(n) = O(n^2)$ ".
- 4 Donnez un exemple d'équation correcte qui contient une notation  $O$  dans le membre gauche et pas dans le membre droit (pas le droit de multiplier par zéro, ce serait trop facile). *Suggestion* : n'oubliez pas les limites.
- 5 Prouvez ou réfutez : si  $f(n)$  et  $g(n)$  sont positives pour tout  $n$ , alors  $O(f(n) + g(n)) = f(n) + O(g(n))$  (à comparer avec (9.27)).
- 6 Multipliez  $(\ln n + \gamma + O(1/n))$  par  $(n + O(\sqrt{n}))$  et exprimez votre réponse en notation  $O$ .
- 7 Approximez  $\sum_{k \geq 0} e^{-k/n}$  avec une erreur absolue en  $O(n^{-1})$ .

### Exercices de base

- 8 Trouvez deux fonctions  $f(n)$  et  $g(n)$  telles qu'aucune des trois relations  $f(n) \prec g(n)$ ,  $f(n) \succ g(n)$ ,  $f(n) \asymp g(n)$  ne soit satisfaite, bien que  $f(n)$  et  $g(n)$  soient strictement croissantes et tendent vers l'infini.
- 9 Prouvez rigoureusement (9.22) en montrant que le membre gauche est un sous-ensemble du membre droit.
- 10 Prouvez ou réfutez :  $\cos O(x) = 1 + O(x^2)$  pour tout réel  $x$ .
- 11 Prouvez ou réfutez :  $O(x+y)^2 = O(x^2) + O(y^2)$ .
- 12 Prouvez que

$$1 + \frac{2}{n} + O(n^{-2}) = \left(1 + \frac{2}{n}\right)(1 + O(n^{-2})),$$

lorsque  $n \rightarrow \infty$ .

- 13 Approximez  $(n + 2 + O(n^{-1}))^n$  avec une erreur relative en  $O(n^{-1})$ .
- 14 Montrez que  $(n + \alpha)^{n+\beta} = n^{n+\beta} e^\alpha (1 + \alpha(\beta - \frac{1}{2}\alpha)n^{-1} + O(n^{-2}))$ .
- 15 Donnez un équivalent asymptotique du coefficient trinomial "central"  $\binom{3n}{n,n,n}$ , avec une erreur relative en  $O(n^{-3})$ .

16 Montrez que si  $B(1-x) = -B(x) \geq 0$  pour  $0 < x < \frac{1}{2}$ , alors

$$\int_a^b B(\{x\}) f(x) dx \geq 0$$

à condition que  $f'(x) \geq 0$  pour  $a \leq x \leq b$ .

17 Utilisez les fonctions génératrices pour montrer que  $B_m(\frac{1}{2}) = (2^{1-m} - 1)B_m$ , pour tout  $m \geq 0$ .

18 Approximez  $\sum_k \binom{2n}{k}^\alpha$  avec une erreur relative en  $O(n^{-1/4})$ , lorsque  $\alpha > 0$ .

#### *Devoirs à la maison*

19 Avec l'aide d'un ordinateur, comparez les membres gauche et droit de chaque approximation de la table 479, pour  $n = 10$ ,  $z = \alpha = 0, 1$  et  $O(f(n)) = O(f(z)) = 0$ .

20 Prouvez ou réfutez les approximations suivantes, lorsque  $n \rightarrow \infty$  :

a  $O\left(\left(\frac{n^2}{\log \log n}\right)^{1/2}\right) = O(\lfloor \sqrt{n} \rfloor^2)$ .

b  $e^{(1+O(1/n))^2} = e + O(1/n)$ .

c  $n! = O\left((1 - 1/n)^n n^n\right)$ .

21 L'équation (9.48) donne le  $n$ ième nombre premier avec une erreur relative en  $O(\log n)^{-2}$ . Faites la diminuer jusqu'à  $O(\log n)^{-3}$  en partant d'un autre terme de (9.31) dans (9.46).

22 Améliorez (9.54) pour obtenir une erreur en  $O(n^{-3})$ .

23 Améliorez (9.62) pour obtenir une erreur absolue en  $O(n^{-3})$ . *Suggestion* : soit  $g_n = c/(n+1)(n+2) + h_n$  ; quelle récurrence  $h_n$  satisfait-il ?

24 Supposez que  $a_n = O(f(n))$  et  $b_n = O(f(n))$ . Est-il vrai que la convolution  $\sum_{k=0}^n a_k b_{n-k}$  est aussi en  $O(f(n))$  dans les cas suivants ?

a  $f(n) = n^{-\alpha}$ ,  $\alpha > 1$ .

b  $f(n) = \alpha^{-n}$ ,  $\alpha > 1$ .

25 Démontrez (9.1) et (9.2).

26 Avec l'équation (9.91), on peut approximer  $\ln 10!$  avec une erreur absolue  $< \frac{1}{126000000}$ . Par conséquent, si on prend l'exponentielle, on obtient  $10!$  avec une erreur relative inférieure à  $e^{1/126000000} - 1 < 10^{-8}$  (on trouve en fait 3628799,9714). Si on arrondit à l'entier le plus proche, on obtient le résultat exact.

Est-il toujours possible de calculer  $n!$  de cette façon si on dispose d'un nombre suffisant de termes de la formule de Stirling ? Estimez la valeur

## 520 CALCUL ASYMPTOTIQUE

de  $m$  qui donne la meilleure approximation de  $\ln n!$ , lorsque  $n$  est un (grand) entier fixé. Comparez l'erreur absolue de cette approximation avec  $n!$ .

- 27 Utilisez la formule de sommation d'Euler pour trouver la valeur asymptotique de  $H_n^{(-\alpha)} = \sum_{k=1}^n k^{-\alpha}$ , où  $\alpha$  est un nombre réel fixé quelconque. (Vous avez le droit de donner votre réponse en fonction d'une constante dont vous ne connaissez pas de forme close).
- 28 La fonction hyperfactorielle  $Q_n = 1^1 2^2 \dots n^n$  est définie dans l'exercice 5.13. Trouvez la valeur asymptotique de  $Q_n$  avec une erreur relative en  $O(n^{-1})$ . (Vous avez le droit de donner votre réponse en fonction d'une constante dont vous ne connaissez pas de forme close).
- 29 Approximez la fonction  $1^{1/1} 2^{1/2} \dots n^{1/n}$  comme dans l'exercice précédent.
- 30 Trouvez la valeur asymptotique de  $\sum_{k \geq 0} k^l e^{-k^2/n}$  avec une erreur absolue en  $O(n^{-3})$ , lorsque  $l$  est un entier positif ou nul fixé.
- 31 Approximez  $\sum_{k \geq 0} 1/(c^k + c^m)$  avec une erreur absolue en  $O(c^{-3m})$ , lorsque  $c > 1$  et  $m$  est un entier strictement positif.

### Problèmes d'examen

- 32 Approximez  $e^{H_n + H_n^{(2)}}$  avec une erreur absolue en  $O(n^{-1})$ .
- 33 Approximez  $\sum_{k \geq 0} \binom{n}{k} / n^{\bar{k}}$  avec une erreur absolue en  $O(n^{-3})$ .
- 34 Trouvez les valeurs de  $A, B, \dots, F$  pour que  $(1 + 1/n)^{nH_n}$  soit égal à
$$An + B(\ln n)^2 + C \ln n + D + \frac{E(\ln n)^2}{n} + \frac{F \ln n}{n} + O(n^{-1}).$$
- 35 Approximez  $\sum_{k=1}^n 1/k H_k$  avec une erreur absolue en  $O(1)$ .
- 36 Approximez  $\sum_{k=1}^n 1/(n^2 + k^2)$  avec une erreur absolue en  $O(n^{-5})$ .
- 37 Approximez  $\sum_{k=1}^n (n \bmod k)$  avec une erreur absolue en  $O(n \log n)$ .
- 38 Approximez  $\sum_{k \geq 0} k^k \binom{n}{k}$  avec une erreur relative en  $O(n^{-1})$ .
- 39 Approximez  $\sum_{0 \leq k < n} \ln(n - k)(\ln n)^k / k!$  avec une erreur absolue en  $O(n^{-1})$ . *Suggestion :* montrez que les termes pour  $k \geq 10 \ln n$  sont négligeables.
- 40 Soit  $m$  un entier strictement positif fixé. Approximez  $\sum_{k=1}^n (-1)^k H_k^m$  avec une erreur absolue en  $O(1)$ .
- 41 Approximez la "factorielle de Fibonacci"  $\prod_{k=1}^n F_k$  avec une erreur relative en  $O(n^{-1})$  ou mieux. (Vous avez le droit de donner votre réponse en fonction d'une constante dont vous ne connaissez pas de forme close).

- 42 Soit  $\alpha$  une constante telle que  $0 < \alpha < \frac{1}{2}$ . Nous avons vu dans les chapitres précédents qu'il n'existe pas de forme close générale pour la somme  $\sum_{k \leq \alpha n} \binom{n}{k}$ . Montrez qu'il existe néanmoins une formule asymptotique

$$\sum_{k \leq \alpha n} \binom{n}{k} = 2^{nH(\alpha) - \frac{1}{2} \lg n + O(1)},$$

où  $H(\alpha) = \alpha \lg \frac{1}{\alpha} + (1-\alpha) \lg \left(\frac{1}{1-\alpha}\right)$ . *Suggestion* : montrez que  $\binom{n}{k-1} < \frac{\alpha}{1-\alpha} \binom{n}{k}$  pour  $0 < k \leq \alpha n$ .

- 43 Montrez que  $C_n$ , le nombre de façons de payer  $n$  cents que nous avons considéré au chapitre 7, est asymptotiquement égal à  $cn^4 + O(n^3)$ , pour une certaine constante  $c$ . Que vaut cette constante ?

- 44 Montrez que

$$\frac{x^{1/2}}{x} = x^{1/2} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} - x^{-1/2} \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} + x^{-3/2} \begin{bmatrix} 1/2 \\ -3/2 \end{bmatrix} + O(x^{-5/2})$$

lorsque  $x \rightarrow \infty$ . (Rappelez-vous que  $x^{1/2} = x!/(x - \frac{1}{2})!$  d'après la définition (5.88) ; souvenez-vous aussi de la définition des nombres de Stirling généralisés donnée dans la table 289).

- 45 Soit  $\alpha$  un nombre irrationnel entre 0 et 1. Nous avons introduit au chapitre 3 la quantité  $D(\alpha, n)$  qui mesure la discrépance qui sépare les parties fractionnaires  $\{k\alpha\}$ , pour  $0 \leq k < n$ , d'une distribution uniforme. La récurrence

$$D(\alpha, n) \leq D(\{\alpha^{-1}\}, \lfloor \alpha n \rfloor) + \alpha^{-1} + 2$$

a été démontrée en (3.31) ; on a aussi les bornes triviales

$$0 \leq D(\alpha, n) \leq n.$$

Montrez que  $\lim_{n \rightarrow \infty} D(\alpha, n)/n = 0$ . *Suggestion* : dans le chapitre 6, il est question de fractions continues.

- 46 Montrez que le nombre de Bell  $\omega_n = e^{-1} \sum_{k \geq 0} k^n/k!$  de l'exercice 7.15 est asymptotiquement égal à

$$m(n)^n e^{m(n)-n-1/2} / \sqrt{\ln n},$$

où  $m(n) \ln m(n) = n - \frac{1}{2}$ , est estimatez l'erreur relative de cette approximation.

- 47 Soit  $m$  un entier  $\geq 2$ . Analysez les deux sommes

$$\sum_{k=1}^n \lfloor \log_m k \rfloor \quad \text{et} \quad \sum_{k=1}^n \lceil \log_m k \rceil.$$

Laquelle est la plus proche de  $\log_m n!$  ?

- 48 Soit une table qui contient des approximations des nombres harmoniques  $H_k$  pour  $1 \leq k \leq n$ , en notation décimale. Pour tout  $1 \leq k \leq n$ , le  $k$ ème nombre de cette table,  $\hat{H}_k$ , est un arrondi à  $d_k$  chiffres significatifs de  $H_k$ , où  $d_k$  est choisi juste assez grand pour pouvoir distinguer cette valeur de celles de  $H_{k-1}$  et  $H_{k+1}$ . Voici par exemple un extrait de la table :

$k$	$H_k$	$\hat{H}_k$	$d_k$
12364	9,99980041-	9,9998	5
12365	9,99988128+	9,9999	5
12366	9,99996215-	9,99996	6
12367	10,00004301-	10,0000	6
12368	10,00012386+	10,0001	6

Approximez le nombre total de chiffres de la table,  $\sum_{k=1}^n d_k$ , avec une erreur absolue en  $O(n)$ .

- 49 Dans le chapitre 6, nous avons considéré l'histoire du ver qui finit par atteindre l'extrémité d'un élastique au bout de  $n$  secondes, avec  $H_{n-1} < 100 \leq H_n$ . Montrez que si  $n$  est un entier strictement positif tel que  $H_{n-1} \leq \alpha \leq H_n$ , alors

$$\lfloor e^{\alpha-\gamma} \rfloor \leq n \leq \lceil e^{\alpha-\gamma} \rceil.$$

- 50 On propose aux adeptes du capital-risque de la Silicon Valley une possibilité de faire un profit exponentiel par rapport à leur investissement : pour un investissement de  $n$  millions de dollars ( $n \geq 2$ ), le consortium GKP leur promet de leur payer  $N$  millions de dollars au bout d'un an, avec  $N = 10^n$ . Il y a bien entendu un petit risque : en réalité, GKP paie  $k$  millions de dollars avec probabilité  $1/(k^2 H_N^{(2)})$  pour tout entier  $k$  tel que  $1 \leq k \leq N$ . Tous les paiements s'effectuent en mégadollars, c'est-à-dire en multiples entiers d'un million de dollars. Notez que tout investisseur est assuré de récupérer au moins un million de dollars.
- a Si on investit  $n$  millions de dollars, quelle est la valeur asymptotique de la somme qu'on peut espérer gagner au bout d'un an ? En d'autres termes, quelle est la moyenne de la somme payée au bout d'un an ? L'erreur absolue de votre réponse ne doit pas dépasser  $O(10^{-n})$  dollars. *Un jour, j'ai gagné  $O(10^{-n})$  dollars.*

- b Quelle est la valeur asymptotique de la probabilité de gagner de l'argent si on investit  $n$  millions ? L'erreur absolue de votre réponse ne doit pas dépasser  $O(n^{-3})$ .

### Questions subsidiaires

- 51 Prouvez ou réfutez :  $\int_n^\infty O(x^{-2}) dx = O(n^{-1})$  lorsque  $n \rightarrow \infty$ .
- 52 Montrez qu'il existe une série  $A(z) = \sum_{k \geq 0} a_n z^n$ , convergente pour tout nombre complexe  $z$ , telle que

$$A(n) \succ n^{n^{n^{\dots^n}}}$$

- 53 Montrez que si  $f(x)$  est une fonction dont les dérivées satisfont

$$f'(x) \leq 0, \quad -f''(x) \leq 0, \quad f'''(x) \leq 0, \quad \dots, \quad (-1)^m f^{(m+1)}(x) \leq 0$$

pour tout  $x \geq 0$ , alors

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \dots + \frac{f^{(m-1)}(0)}{(m-1)!}x^{m-1} + O(x^m), \text{ pour } x \geq 0.$$

En particulier, le cas  $f(x) = -\ln(1+x)$  permet de prouver (9.64) pour tous  $k, n > 0$ .

- 54 Montrez que, si  $f(x)$  est une fonction strictement positive et dérivable telle que  $xf'(x) \prec f(x)$  lorsque  $x \rightarrow \infty$ , alors

$$\sum_{k \geq n} \frac{f(k)}{k^{1+\alpha}} = O\left(\frac{f(n)}{n^\alpha}\right), \quad \text{si } \alpha > 0.$$

*Suggestion :* Considérez  $f(k - \frac{1}{2})/(k - \frac{1}{2})^\alpha - f(k + \frac{1}{2})/(k + \frac{1}{2})^\alpha$ .

- 55 Améliorez (9.99) pour obtenir une erreur relative en  $O(n^{-3/2+5\epsilon})$ .

- 56 La somme  $Q(n) = 1 + \frac{n-1}{n} + \frac{n-1}{n} \frac{n-2}{n} + \dots = \sum_{k \geq 1} n^k/n^k$  apparaît dans l'analyse de bien des algorithmes. Trouvez sa valeur asymptotique, avec une erreur absolue en  $o(1)$ .

- 57 On donne dans (9.54) un équivalent asymptotique de la somme de Golomb  $\sum_{k \geq 1} 1/k[1 + \log_n k]^2$ . Trouvez un équivalent asymptotique de la somme  $\sum_{k \geq 1} 1/k(1 + \log_n k)^2$ , c'est-à-dire la même sans les parties entières inférieures. *Suggestion :* considérez  $\int_0^\infty ue^{-u}k^{-tu} du = 1/(1 + t \ln k)^2$ .

- 58 Montrez que

$$B_m(\{x\}) = -2 \frac{m!}{(2\pi)^m} \sum_{k \geq 1} \frac{\cos(2\pi kx - \frac{1}{2}\pi m)}{k^m}, \quad \text{pour } m \geq 2,$$

avec la méthode des résidus, en intégrant

$$\frac{1}{2\pi i} \oint \frac{2\pi i e^{2\pi iz\theta}}{e^{2\pi iz} - 1} \frac{dz}{z^m}$$

sur le contour carré  $z = x + iy$ , où  $\max(|x|, |y|) = M + \frac{1}{2}$ , et en faisant tendre l'entier  $M$  vers  $\infty$ .

- 59 Soit  $\Theta_n(t) = \sum_k e^{-(k+t)^2/n}$  une fonction périodique de  $t$ . Montrez que le développement en série de Fourier de  $\Theta_n(t)$  est

$$\begin{aligned}\Theta_n(t) = \sqrt{\pi n} & (1 + 2e^{-\pi^2 n}(\cos 2\pi t) + 2e^{-4\pi^2 n}(\cos 4\pi t) \\ & + 2e^{-9\pi^2 n}(\cos 6\pi t) + \dots).\end{aligned}$$

(Cette formule donne une série rapidement convergente pour la somme  $\Theta_n = \Theta_n(0)$  de l'équation (9.93)).

- 60 Expliquez pourquoi les dénominateurs des coefficients du développement asymptotique

$$\binom{2n}{n} = \frac{4^n}{\sqrt{\pi n}} \left(1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} - \frac{21}{32768n^4} + O(n^{-5})\right)$$

sont tous des puissances de 2.

- 61 On prouve dans l'exercice 45 que la discrépance  $D(\alpha, n)$  est en  $o(n)$  pour tout nombre irrationnel  $\alpha$ . Trouvez un nombre irrationnel  $\alpha$  tel que, pour tout  $\epsilon > 0$ ,  $D(\alpha, n)$  n'est pas en  $O(n^{1-\epsilon})$ .
- 62 Etant donné  $n$ , soit  $\{\frac{n}{m(n)}\} = \max_k \{\frac{n}{k}\}$  le plus grand nombre de la ligne  $n$  du triangle des sous-ensembles de Stirling. Montrez que, pour tout  $n$  suffisamment grand, on a  $m(n) = \lfloor \bar{m}(n) \rfloor$  ou  $m(n) = \lceil \bar{m}(n) \rceil$ , où

$$\bar{m}(n)(\bar{m}(n) + 2) \ln(\bar{m}(n) + 2) = n(\bar{m}(n) + 1).$$

*Suggestion :* ce n'est pas facile.

- 63 Montrez que la suite auto-descriptive de Golomb, qui est définie dans l'exercice 2.36 satisfait  $f(n) = \phi^{2-\Phi} n^{\Phi-1} + O(n^{\Phi-1}/\log n)$ .
- 64 Démontrez l'identité

$$\sum_{n \geq 1} \frac{\cos 2n\pi x}{n^2} = \pi^2(x^2 - x + \frac{1}{6}) \quad \text{pour } 0 \leq x \leq 1,$$

en utilisant exclusivement les mathématiques "eulériennes" (celles du dix-huitième siècle).

- 65 Quels sont les coefficients du développement asymptotique

$$1 + \frac{1}{n-1} + \frac{1}{(n-1)(n-2)} + \cdots + \frac{1}{(n-1)!} = a_0 + \frac{a_1}{n} + \frac{a_2}{n^2} + \cdots ?$$

**Problèmes de recherche**

- 66 Trouvez une preuve “combinatoire” de la formule de Stirling. (Remarquez que  $n^n$  est le nombre d’applications de l’ensemble  $\{1, 2, \dots, n\}$  dans lui-même et que  $n!$  est le nombre de bijections de  $\{1, 2, \dots, n\}$  dans lui-même).
- 67 Considérez un tableau  $n \times n$  de points ( $n \geq 3$ ), dans lequel chaque point a quatre voisins (chaque point d’un côté est voisin du point correspondant sur le côté opposé). Soit  $\chi_n$  le nombre de façons de colorier ces points en bleu, blanc ou rouge en faisant en sorte que deux points voisins n’aient jamais la même couleur (ainsi,  $\chi_3 = 12$  par exemple). montrez que
- $$\chi_n \sim \left(\frac{4}{3}\right)^{3n^2/2} e^{-\pi/6}.$$
- 68 Soit  $Q_n$  le plus petit entier  $m$  tel que  $H_m > n$ . Trouvez le plus petit entier  $n$  tel que  $Q_n \neq \lfloor e^{n-\gamma} + \frac{1}{2} \rfloor$ , ou alors prouvez que ce  $n$  n’existe pas.

*Th-th-th—that’s all,  
folks !*

# A

## Solutions des exercices

ON TROUVERA ICI une réponse (parfois brève) à chaque exercice ; certaines réponses vont même au-delà de ce qui était demandé. Le lecteur profitera mieux des exercices s'il essaie sérieusement de les résoudre AVANT de lorgner sur cette annexe. Toute solution (même partielle) d'un sujet de recherche, ou toute solution plus simple (ou plus juste) pour un exercice donné, sont susceptibles d'intéresser les auteurs.

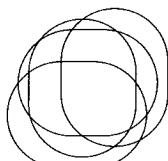
**1.1** La preuve est bonne sauf pour  $n = 2$ . Si tous les ensembles de deux chevaux ont leurs chevaux de la même couleur, alors la proposition est vraie.

**1.2** Soit  $X_n$  le nombre de mouvements. On a  $X_0 = 0$  et  $X_n = X_{n-1} + 1 + X_{n-1} + 1 + X_{n-1}$  lorsque  $n > 0$ . Il s'ensuit (par exemple en ajoutant 1 de chaque côté) que  $X_n = 3^n - 1$ . Note : après  $\frac{1}{2}X_n$ , la tour entière sera sur l'axe du milieu, à mi-chemin de sa destination !

**1.3** Il y a  $3^n$  configurations possibles, car chaque disque peut être posé sur chacun des trois axes. On les obtient forcément toutes car la solution la plus courte nécessite  $3^n - 1$  mouvements. Cette construction est équivalente à un "code de Gray ternaire", qui parcourt tous les nombres de  $(0 \dots 0)_3$  à  $(2 \dots 2)_3$  en modifiant un seul chiffre à chaque étape.

**1.4** Non. Si le plus grand des disques n'a pas à être déplacé, on montre par induction que  $2^{n-1} - 1$  mouvements suffisent ; sinon, encore par induction,  $(2^{n-1} - 1) + 1 + (2^{n-1} - 1)$  mouvements suffisent.

**1.5** Non. Comme deux cercles distincts ne peuvent se couper qu'en deux points au plus, on ne peut délimiter que 14 régions avec quatre cercles. Par contre, c'est possible avec des ovales :



*C'est le nombre d'intersections qui donne la solution ; la convexité était une fausse piste.*

Venn [359] prétendait qu'il n'était pas possible de représenter une configuration de cinq ensembles avec des ellipses, mais Grünbaum [167] a trouvé une solution.

*On suppose ici que  $n > 0$ .*

**1.6** Si la  $n$ ième ligne coupe les autres en  $k > 0$  points distincts, on obtient  $k - 1$  nouvelles régions bornées (en supposant qu'il n'y a pas deux droites parallèles parmi celles tracées précédemment) et deux nouvelles régions infinies. Par conséquent, le nombre maximum de régions bornées est  $(n-2) + (n-3) + \dots = S_{n-2} = (n-1)(n-2)/2 = L_n - 2n$ .

**1.7** On n'a pas prouvé la base. En fait,  $H(1) \neq 2$ .

**1.8**  $Q_2 = (1 + \beta)/\alpha$  ;  $Q_3 = (1 + \alpha + \beta)/\alpha\beta$  ;  $Q_4 = (1 + \alpha)/\beta$  ;  $Q_5 = \alpha$  ;  $Q_6 = \beta$ . La suite est donc périodique !

**1.9** (a) On obtient  $P(n - 1)$  à partir de l'inégalité

$$x_1 \dots x_{n-1} \left( \frac{x_1 + \dots + x_{n-1}}{n-1} \right) \leq \left( \frac{x_1 + \dots + x_{n-1}}{n-1} \right)^n.$$

(b)  $x_1 \dots x_n x_{n+1} \dots x_{2n} \leq ((x_1 + \dots + x_n)/n)((x_{n+1} + \dots + x_{2n})/n)^n$  d'après  $P(n)$  ; et d'après  $P(2)$ , le produit à l'intérieur de l'expression précédente est  $\leq ((x_1 + \dots + x_{2n})/2n)^2$ . (c) Par exemple,  $P(5)$  se déduit de  $P(6)$  qui se déduit de  $P(3)$  qui se déduit de  $P(2)$ .

**1.10** Il faut d'abord montrer que  $R_n = R_{n-1} + 1 + Q_{n-1} + 1 + R_{n-1}$  lorsque  $n > 0$ . Profitons-en pour annoncer que les méthodes que nous verrons au chapitre 7 nous permettront de montrer que  $Q_n = ((1 + \sqrt{3})^{n+1} - (1 - \sqrt{3})^{n+1})/(2\sqrt{3}) - 1$ .

**1.11** (a) Le mieux qu'on puisse faire, c'est déplacer une tour double à  $2n - 2$  disques, puis déplacer les deux disques les plus larges (inversant ainsi leur ordre), enfin déplacer à nouveau la tour de  $n - 2$  disques. Donc  $A_n = 2A_{n-1} + 2$  et  $A_n = 2T_n = 2^{n+1} - 2$ . Cette solution intervertit les deux disques les plus larges, mais laisse les  $2n - 2$  autres dans leur ordre d'origine.

(b) Soit  $B_n$  le nombre minimal de mouvements. Alors  $B_1 = 3$  et on peut montrer qu'aucune stratégie ne peut faire mieux que  $B_n = A_{n-1} + 2 + A_{n-1} + 2 + B_{n-1}$  lorsque  $n > 1$ . Par conséquent,  $B_n = 2^{n+2} - 5$  pour tout  $n > 0$ . Curieusement, c'est exactement égal à  $2A_n - 1$  ; de plus,  $B_n = A_{n-1} + 1 + A_{n-1} + 1 + A_{n-1} + 1 + A_{n-1}$ .

**1.12** Si  $m_k > 0$  pour tout  $k$ , alors  $A(m_1, \dots, m_n) = 2A(m_1, \dots, m_{n-1}) + m_n$ . Cette équation est du type "Josèphe généralisé", et sa solution est  $(m_1 \dots m_n)_2 = 2^{n-1}m_1 + \dots + 2m_{n-1} + m_n$ . Remarquons en passant que la solution de la même généralisation pour l'exercice 11b satisfait la

récurrence

$$B(m_1, \dots, m_n) = \begin{cases} A(m_1, \dots, m_n), & \text{si } m_n = 1; \\ 2m_n - 1, & \text{si } n = 1; \\ 2A(m_1, \dots, m_{n-1}) + 2m_n \\ \quad + B(m_1, \dots, m_{n-1}), & \text{si } n > 1 \text{ et } m_n > 1. \end{cases}$$

**1.13** Considérons  $n$  droites qui définissent  $L_n$  régions. On peut remplacer chacune d'entre elles par une ligne brisée suffisamment "repliée" pour qu'il y ait neuf intersections entre chaque couple de lignes brisées distinctes. Cela entraîne que  $ZZ_n = ZZ_{n-1} + 9n - 8$  pour tout  $n > 0$ ; par conséquent,  $ZZ_n = 9S_n - 8n + 1 = \frac{9}{2}n^2 - \frac{7}{2}n + 1$ .

**1.14** Le nombre de régions tri-dimensionnelles définies par chaque nouvelle coupe est égal au nombre de régions bi-dimensionnelles délimitées, dans le nouveau plan, par ses intersections avec les plans précédents. Par conséquent  $P_n = P_{n-1} + L_{n-1}$ , et on trouve  $P_5 = 26$ . (En coupant six fois un fromage cubique, on peut obtenir 27 cubes plus petits, ou bien jusqu'à  $P_6 = 42$  parts de fromage de formes plus bizarres).

Incidemment, la solution de cette récurrence a une jolie expression en termes de coefficients binomiaux (voir le chapitre 5) :

$$\begin{aligned} X_n &= \binom{n}{0} + \binom{n}{1}; \\ L_n &= \binom{n}{0} + \binom{n}{1} + \binom{n}{2}; \\ P_n &= \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3}. \end{aligned}$$

Ici,  $X_n$  représente le nombre de régions uni-dimensionnelles définissables par  $n$  points sur une droite.

*Je parie que je peux dire ce qui se passe en dimension quatre !*

**1.15** La fonction  $I$  satisfait la même récurrence que  $J$  quand  $n > 1$ , mais  $I(1)$  n'est pas défini. Comme  $I(2) = 2$  et  $I(3) = 1$ , il n'existe pas de valeur de  $I(1) = \alpha$  qui nous permette d'utiliser notre méthode générale. Si on développe la récurrence, le résultat final dépend des deux premiers chiffres de la représentation binaire de  $n$ .

Si  $n = 2^m + 2^{m-1} + k$ , où  $0 \leq k < 2^{m+1} + 2^m - (2^m + 2^{m-1}) = 2^m + 2^{m-1}$ , la solution est  $I(n) = 2k+1$  pour tout  $n > 2$ . On peut l'exprimer différemment, en posant  $n = 2^m + l$ :

$$I(n) = \begin{cases} J(n) + 2^{m-1}, & \text{si } 0 \leq l < 2^{m-1}; \\ J(n) - 2^m, & \text{si } 2^{m-1} \leq l < 2^m. \end{cases}$$

**1.16** Soit  $g(n) = a(n)\alpha + b(n)\beta_0 + c(n)\beta_1 + d(n)\gamma$ . On sait, d'après (1.18), que  $a(n)\alpha + b(n)\beta_0 + c(n)\beta_1 = (\alpha \beta_{b_{m-1}} \beta_{b_{m-2}} \dots \beta_{b_1} \beta_{b_0})_3$  lorsque  $n =$

$(1 b_{m-1} \dots b_1 b_0)_2$  ; ceci définit  $a(n)$ ,  $b(n)$  et  $c(n)$ . En posant  $g(n) = n$  dans la récurrence, on en déduit que  $a(n) + c(n) - d(n) = n$  ; nous avons donc tout résolu. (En posant  $g(n) = 1$ , on obtient de plus l'égalité  $a(n) - 2b(n) - 2c(n) = 1$ , que l'on peut utiliser pour définir  $b(n)$  plus simplement en fonction de  $a(n)$  et  $a(n) + c(n)$ ).

**1.17** On a  $W_m \leq 2W_{m-k} + T_k$ , pour  $0 \leq k \leq m$  (en transférant les  $m-k$  disques du haut, puis en utilisant seulement trois axes pour transférer les  $k$  disques du bas, enfin en posant dessus les  $m-k$  du haut). En prenant pour  $k$  l'unique valeur qui minimise le membre droit de cette inégalité quand  $m = n(n+1)/2$ , on trouve la formule demandée (on ne peut cependant pas en conclure l'égalité des deux membres, car beaucoup d'autres stratégies sont concevables pour transférer la tour). En posant  $Y_n = (W_{n(n+1)/2} - 1)/2^n$ , on trouve que  $Y_n \leq Y_{n-1} + 1$  ; ainsi  $W_{n(n+1)/2} \leq 2^n(n-1) + 1$ .

**1.18** Il suffit de montrer que chacune des deux lignes partant de  $(n^{2j}, 0)$  coupe chacune des deux lignes partant de  $(n^{2k}, 0)$  et que tous les points d'intersection sont distincts.

Une ligne partant de  $(x_j, 0)$  et passant par  $(x_j - a_j, 1)$  coupe une ligne partant de  $(x_k, 0)$  et passant par  $(x_k - a_k, 1)$  au point de coordonnées  $(x_j - ta_j, t)$ , où  $t = (x_k - x_j)/(a_k - a_j)$ . Soit  $x_j = n^{2j}$  et  $a_j = n^j + (0 \text{ ou } n^{-n})$ . Alors le rapport  $t = (n^{2k} - n^{2j})/(n^k - n^j + (-n^{-n} \text{ ou } 0 \text{ ou } n^{-n}))$  est compris strictement entre  $n^j + n^k - 1$  et  $n^j + n^k + 1$  ; donc l'ordonnée du point d'intersection détermine  $j$  et  $k$  de façon unique, et les quatre points d'intersection qui ont le même  $j$  et le même  $k$  sont distincts.

**1.19** Non si  $n > 5$ . Soit une ligne brisée dont les deux demi-droites ont respectivement des pentes de  $\theta$  et  $\theta + 30^\circ$ , et une autre pour laquelle les pentes sont de  $\phi$  et  $\phi + 30^\circ$ . Ces deux lignes brisées ne peuvent se couper en quatre point que si  $30^\circ < |\theta - \phi| < 150^\circ$ . Il n'est pas possible de choisir plus de 5 angles respectant mutuellement cette propriété (c'est possible jusqu'à 5).

**1.20** Soit  $h(n) = a(n)\alpha + b(n)\beta_0 + c(n)\beta_1 + d(n)\gamma_0 + e(n)\gamma_1$ . Nous savons d'après (1.18) que  $a(n)\alpha + b(n)\beta_0 + c(n)\beta_1 = (\alpha \beta_{b_{m-1}} \beta_{b_{m-2}} \dots \beta_{b_1} \beta_{b_0})_4$  lorsque  $n = (1 b_{m-1} \dots b_1 b_0)_2$  ; ceci détermine  $a(n)$ ,  $b(n)$  et  $c(n)$ . En posant  $h(n) = n^2$ , on déduit que  $a(n) + c(n) + 4e(n) = n^2$ . Ainsi  $d(n) = (3a(n) + 3c(n) - n^2 - 2n)/4$  et  $e(n) = (n^2 - a(n) - c(n))/4$ .

**1.21** On peut prendre pour  $m$  le plus petit commun multiple de  $2n$ ,  $2n-1$ ,  $\dots$ ,  $n+1$ , ou n'importe lequel de leurs multiples communs. (Intuitivement, on peut penser qu'en choisissant  $m$  "au hasard", on réussira avec une probabilité de

$$\frac{n}{2n} \frac{n-1}{2n-1} \dots \frac{1}{n+1} = 1 / \binom{2n}{n} \sim \frac{\sqrt{\pi n}}{4^n} ;$$

on peut donc s'attendre à ce qu'il existe un  $m$  plus petit que  $4^n$ ).

**1.22** On prend un polygone à  $2^n$  côtés et on étiquette ses côtés par les éléments d'un cycle de de Bruijn de longueur  $2n$  (c'est une suite cyclique de 0 et de 1 dans laquelle toutes les suites de  $n$  éléments adjacents sont différentes ; voir [207, exercice 2.3.4.2–23] et [208, exercice 3.2.2–17]). Collez à chaque côté étiqueté 1 une petite extension convexe. Les  $n$  ensembles sont obtenus par copie du polygone obtenu, puis rotation d'un angle couvrant  $k$  côtés pour  $k = 0, 1, \dots, n - 1$ .

**1.23** Oui. Nous utilisons dans la preuve quelques principes de théorie des nombres élémentaire du chapitre 4. Soit  $L(n) = \text{ppcm}(1, 2, \dots, n)$ . On peut supposer que  $n > 2$  ; donc, d'après le postulat de Bertrand, il existe un nombre premier  $p$  entre  $n/2$  et  $n$ . On peut aussi supposer que  $j > n/2$ , car  $q' = L(n) + 1 - q$  épargne  $j' = n + 1 - j$  si et seulement si  $q$  épargne  $j$ . Il faut choisir  $q$  tel que  $q \equiv 1 \pmod{L(n)/p}$  et  $q \equiv j + 1 - n \pmod{p}$ . L'ordre d'élimination est alors  $1, 2, \dots, n - p, j + 1, j + 2, \dots, n, n - p + 1, \dots, j - 1$ .

**1.24** Les seuls exemples connus sont :  $X_n = 2i \sin \pi r + 1/X_{n-1}$ , où  $r$  est rationnel et  $0 \leq r < \frac{1}{2}$  (lorsque  $r$  varie, toutes les périodes de longueur  $\geq 2$  peuvent apparaître) ; la récurrence de Gauss, de période 5, dans l'exercice 8 ; la récurrence plus remarquable encore de H. Todd,  $X_n = (1 + X_{n-1} + X_{n-2})/X_{n-3}$ , de période 8 (voir [261]) ; et les récurrences obtenues en remplaçant un nombre constant de fois  $X_n$  par  $X_{m,n}$  dans les précédentes. On peut supposer que le premier coefficient non nul dans le dénominateur est 1, et que le premier coefficient non nul dans le numérateur (s'il y a lieu) a une partie réelle positive ou nulle. En utilisant un ordinateur, on peut montrer facilement qu'il n'existe pas d'autres solutions de période  $\leq 5$  lorsque  $k = 2$ . Lyness [261, 262], Kurshan et Gopinath [231] ont développé une théorie partielle de ce problème.

Il existe un intéressant exemple d'une récurrence d'un autre type, de période 9 lorsque les valeurs de départ sont réelles :  $X_n = |X_{n-1}| - X_{n-2}$  ; elle a été découverte par Morton Brown [43]. Les récurrences non linéaires de période  $\geq 5$  peuvent se baser sur les continuants [65].

**1.25** Si  $T^{(k)}(n)$  désigne le nombre minimum de mouvements nécessaires pour transférer  $n$  disques en utilisant  $k$  disques auxiliaires (donc  $T^{(1)}(n) = T_n$  et  $T^{(2)}(n) = W_n$ ), on a  $T^{(k)}(\binom{n+1}{k}) \leq 2T^{(k)}(\binom{n}{k}) + T^{(k-1)}(\binom{n}{k-1})$ . On ne connaît pas de couple  $(n, k)$  pour lequel cette inégalité n'est pas une égalité. Quand  $k$  est petit par rapport à  $n$ , la formule  $2^{n+1-k}\binom{n-1}{k-1}$  donne un bon majorant de  $T^{(k)}(\binom{n}{k})$ .

*Une fois, je suis monté sur un cycle de de Bruijn (c'était chez lui à Nuenen, en Hollande).*

**1.26** Pour tous  $m$  et  $n$ , la permutation correspondant à l'ordre d'exécution peut être calculée en  $O(n \log n)$  étapes [209, exercices 5.1.1–2 et 5.1.1–5]. Bjorn Poonen a démontré qu'il existe des sous-ensembles qui ne sont pas des sous-ensembles de Josèphe, contenant exactement quatre "méchants",

dès lors que  $n \equiv 0 \pmod{3}$  et  $n \geq 9$ ; en fait, il existe  $\epsilon > 0$  tel que le nombre de tels sous-ensembles soit supérieur ou égal à  $\epsilon \binom{n}{4}$ . Il a aussi montré, en calculant par ordinateur, que le seul autre  $n < 24$  pour lequel il y a des sous-ensembles qui ne sont pas de Josèphe est  $n = 20$ , pour lequel il existe 236 tels sous-ensembles avec  $k = 14$  et deux avec  $k = 13$  (l'un de ceux-ci est  $\{1, 2, 3, 4, 5, 6, 7, 8, 11, 14, 15, 16, 17\}$ ; l'autre est son image-miroir par rapport à 21). Il existe un seul sous-ensemble qui n'est pas de Josèphe pour  $n = 15$  et  $k = 9$ :  $\{3, 4, 5, 6, 8, 10, 11, 12, 13\}$ .

**2.1** Il n'y a pas une solution unique; trois bonnes réponses sont envisageables : (1) On peut dire que  $\sum_{k=m}^n q_k$  est toujours équivalent à  $\sum_{m \leq k \leq n} q_k$ ; donc la somme est nulle. (2) En sommant sur des valeurs décroissantes de  $k$ , on pourrait dire que la somme est égale à  $q_4 + q_3 + q_2 + q_1 + q_0$ . Toutefois, cela serait en conflit avec la convention généralement admise selon laquelle  $\sum_{k=1}^n q_k = 0$  lorsque  $n = 0$ . (3) On peut aussi dire que  $\sum_{k=m}^n q_k = \sum_{k \leq n} q_k - \sum_{k < m} q_k$ ; alors la somme est égale à  $-q_1 - q_2 - q_3$ . Cette convention peut paraître bizarre; elle obéit pourtant à une règle très utilisée selon laquelle  $\sum_{k=a}^b + \sum_{k=b+1}^c = \sum_{k=a}^c$  pour tous  $a, b, c$ .

Il vaut mieux n'utiliser la notation  $\sum_{k=m}^n$  que lorsque  $n - m \geq -1$ ; dans ce cas, les conventions (1) et (3) concordent.

**2.2** Cela fait  $|x|$ . Notons en passant que la quantité  $([x > 0] - [x < 0])$  est souvent appelée "signe( $x$ )" ou "signum"( $x$ ).

**2.3** La première somme donne évidemment  $a_0 + a_1 + a_2 + a_3 + a_4 + a_5$ ; la seconde est égale à  $a_4 + a_1 + a_0 + a_1 + a_4$ , car elle est indicée par  $k \in \{-2, -1, 0, +1, +2\}$ . On ne peut pas utiliser la règle de commutativité car la fonction  $p(k) = k^2$  n'est pas une permutation. Il existe des valeurs de  $n$  (par exemple  $n = 3$ ) pour lesquelles il n'y a pas de  $k$  tel que  $p(k) = n$ ; pour d'autres valeurs (par exemple  $n = 4$ ), on en trouve deux.

**2.4** (a)  $\sum_{i=1}^4 \sum_{j=i+1}^4 \sum_{k=j+1}^4 a_{ijk} = \sum_{i=1}^2 \sum_{j=i+1}^3 \sum_{k=j+1}^4 a_{ijk} = ((a_{123} + a_{124}) + a_{134}) + a_{234}$ .

(b)  $\sum_{k=1}^4 \sum_{j=1}^{k-1} \sum_{i=1}^{j-1} a_{ijk} = \sum_{k=3}^4 \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} a_{ijk} = a_{123} + (a_{124} + (a_{134} + a_{234}))$ .

**2.5** On utilise le même symbole " $k$ " pour deux indices différents, en dépit du fait que  $k$  est lié dans la somme interne. C'est une erreur bien connue en mathématiques (et en informatique). Le résultat peut quand même être correct si  $a_j = a_k$  pour tout  $j$  et tout  $k$  tels que  $1 \leq j, k \leq n$ .

**2.6** C'est  $[1 \leq j \leq n](n - j + 1)$ . Le premier facteur est nécessaire car on doit obtenir zéro lorsque  $j < 1$  ou  $j > n$ .

**2.7**  $m\overline{x^{m-1}}$ . Par conséquent, une version du calcul fini basée sur  $\nabla$  au lieu de  $\Delta$  accorderait une importance particulière à la puissance factorielle *montante*.

**2.8**  $0$  si  $m \geq 1$  ;  $1/|m|!$  si  $m \leq 0$ .

**2.9**  $x^{\overline{m+n}} = x^{\overline{m}}(x+m)^{\overline{n}}$ , pour  $m$  et  $n$  entiers. En posant  $m = -n$ , on trouve  $x^{\overline{-n}} = 1/(x-n)^{\overline{n}} = 1/(x-1)^{\underline{n}}$ .

**2.10** Voici un autre membre droit possible :  $\mathbb{E}u\Delta v + v\Delta u$ .

**2.11** Séparez le membre de gauche en deux sommes puis transformez  $k$  en  $k+1$  dans la seconde somme.

**2.12** Si  $p(k) = n$  alors  $n+c = k + ((-1)^k + 1)c$  et  $((-1)^k + 1)$  est pair ; donc  $(-1)^{n+c} = (-1)^k$  et  $k = n - (-1)^{n+c}c$ . Réciproquement, pour cette valeur de  $k$ ,  $p(k) = n$ .

**2.13** Soit  $R_0 = \alpha$  et  $R_n = R_{n-1} + (-1)^n(\beta + n\gamma + n^2\delta)$  pour  $n > 0$ . Alors  $R(n) = A(n)\alpha + B(n)\beta + C(n)\gamma + D(n)\delta$ . En posant  $R_n = 1$ , on obtient  $A(n) = 1$ . En posant  $R_n = (-1)^n$ , on obtient  $A(n) + 2B(n) = (-1)^n$ . En posant  $R_n = (-1)^n n$ , on obtient  $-B(n) + 2C(n) = (-1)^n n$ . En posant  $R_n = (-1)^n n^2$ , on obtient  $B(n) - 2C(n) + 2D(n) = (-1)^n n^2$ . Par conséquent,  $2D(n) = (-1)^n(n^2 + n)$  ; la somme recherchée est  $D(n)$ .

**2.14** La réécriture suggérée est légitime, du fait que  $k = \sum_{1 \leq j \leq k} 1$  lorsque  $1 \leq k \leq n$ . En sommant d'abord sur  $k$ , on réduit la somme à

$$\sum_{1 \leq j \leq n} (2^{n+1} - 2^j) = n2^{n+1} - (2^{n+1} - 2).$$

**2.15** La première étape a pour effet de remplacer  $k(k+1)$  par  $2 \sum_{1 \leq j \leq k} j$ . La seconde étape donne  $\square_n + \square_n = (\sum_{k=1}^n k)^2 + \square_n$ .

**2.16**  $x^{\overline{m}}(x-m)^{\underline{n}} = x^{\overline{m+n}} = x^{\underline{n}}(x-n)^{\overline{m}}$ , d'après (2.52).

**2.17** On utilise l'induction pour les deux premières égalités et (2.52) pour la troisième. La seconde ligne se déduit de la première.

**2.18** Utilisez les résultats suivants :  $(\Re z)^+ \leq |z|$ ,  $(\Re z)^- \leq |z|$ ,  $(\Im z)^+ \leq |z|$ ,  $(\Im z)^- \leq |z|$ , et  $|z| \leq (\Re z)^+ + (\Re z)^- + (\Im z)^+ + (\Im z)^-$ .

**2.19** Multipliez les deux membres par  $2^{n-1}/n!$ . Soit maintenant  $S_n = 2^n T_n/n! = S_{n-1} + 3 \cdot 2^{n-1} = 3(2^n - 1) + S_0$ . la solution est  $T_n = 3 \cdot n! + n!/2^{n-1}$ . (Nous verrons au chapitre 4 que  $T_n$  n'est entier que si  $n$  est nul ou est une puissance de 2).

**2.20** La méthode de perturbation donne

$$S_n + (n+1)H_{n+1} = S_n + \left( \sum_{0 \leq k \leq n} H_k \right) + n+1.$$

*"It is a profoundly erroneous truism, repeated by all copybooks and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle—they are strictly limited in number, they require fresh horses, and must only be made at decisive moments."*

—A. N. Whitehead [370]

**2.21** En extrayant le dernier terme de  $S_{n+1}$ , on obtient  $S_{n+1} = 1 - S_n$  ; en extrayant le premier terme, on obtient

$$\begin{aligned} S_{n+1} &= (-1)^{n+1} + \sum_{1 \leq k \leq n+1} (-1)^{n+1-k} \\ &= (-1)^{n+1} + \sum_{0 \leq k \leq n} (-1)^{n-k} \\ &= (-1)^{n+1} + S_n. \end{aligned}$$

Par conséquent,  $2S_n = 1 + (-1)^n$  et  $S_n = [n \text{ est pair}]$ . De manière similaire, on trouve

$$T_{n+1} = n + 1 - T_n = \sum_{k=0}^n (-1)^{n-k}(k+1) = T_n + S_n,$$

donc  $2T_n = n + 1 - S_n$  et  $T_n = \frac{1}{2}(n + [n \text{ est impair}])$ . Pour finir, la même approche permet de trouver que

$$\begin{aligned} U_{n+1} &= (n+1)^2 - U_n = U_n + 2T_n + S_n \\ &= U_n + n + [n \text{ est impair}] + [n \text{ est pair}] \\ &= U_n + n + 1. \end{aligned}$$

Donc  $U_n$  est égal au nombre triangulaire  $\frac{1}{2}(n+1)n$ .

**2.22** En doublant la somme générale, on obtient une somme pour  $1 \leq j, k \leq n$ , qui se partage pour donner deux fois  $(\sum_k a_k A_k)(\sum_k b_k B_k) - (\sum_k a_k B_k)(\sum_k b_k A_k)$ .

**2.23** (a) On obtient par cette méthode quatre sommes, dont l'évaluation donne  $2n + H_n - 2n + (H_n + \frac{1}{n+1} - 1)$ . (Il aurait été plus facile de remplacer le terme générique par  $1/k + 1/(k+1)$ ). (b) Soient  $u(x) = 2x + 1$  et  $\Delta v(x) = 1/x(x+1) = (x-1)^{-2}$  ; alors  $\Delta u(x) = 2$  et  $v(x) = -(x-1)^{-1} = -1/x$ . La réponse est  $2H_n - \frac{n}{n+1}$ .

**2.24** En sommant par parties, on trouve  $\sum x^m H_x \delta x = x^{m+1} H_x / (m+1) - x^{m+1} / (m+1)^2 + C$  ; donc  $\sum_{0 \leq k < n} k^m H_k = n^{m+1} (H_n - 1/(m+1)) / (m+1) + 0^{m+1} / (m+1)^2$ . En ce qui nous concerne,  $m = -2$  donc la somme donne  $1 - (H_n + 1)/(n+1)$ .

**2.25** Voici quelques analogies de base :

$$\begin{aligned} \sum_{k \in K} c a_k &= c \sum_{k \in K} a_k & \leftrightarrow \prod_{k \in K} a_k^c &= \left( \prod_{k \in K} a_k \right)^c \\ \sum_{k \in K} (a_k + b_k) &= \sum_{k \in K} a_k + \sum_{k \in K} b_k & \leftrightarrow \prod_{k \in K} a_k b_k &= \left( \prod_{k \in K} a_k \right) \left( \prod_{k \in K} b_k \right) \end{aligned}$$

$$\begin{array}{ll} \sum_{k \in K} a_k = \sum_{p(k) \in K} a_{p(k)} & \leftrightarrow \prod_{k \in K} a_k = \prod_{p(k) \in K} a_{p(k)} \\ \sum_{\substack{j \in J \\ k \in K}} a_{j,k} = \sum_{j \in J} \sum_{k \in K} a_{j,k} & \leftrightarrow \prod_{\substack{j \in J \\ k \in K}} a_{j,k} = \prod_{j \in J} \prod_{k \in K} a_{j,k} \\ \sum_{k \in K} a_k = \sum_k a_k [k \in K] & \leftrightarrow \prod_{k \in K} a_k = \prod_k a_k^{[k \in K]} \\ \sum_{k \in K} 1 = \#K & \leftrightarrow \prod_{k \in K} c = c^{\#K} \end{array}$$

**2.26**  $P^2 = (\prod_{1 \leq j, k \leq n} a_j a_k) (\prod_{1 \leq j=k \leq n} a_j a_k)$ . Le premier facteur est égal à  $(\prod_{k=1}^n a_k^n)^2$ ; le second facteur est  $\prod_{k=1}^n a_k^2$ . Par conséquent  $P = (\prod_{k=1}^n a_k)^{n+1}$ .

**2.27**  $\Delta(c^x) = c^x(c - x - 1) = c^{x+2}/(c - x)$ . En posant  $c = -2$  et en diminuant  $x$  de 2, on obtient  $\Delta(-(-2)^{x-2}) = (-2)^x/x$ ; la somme considérée est donc égale à  $(-2)^{-1} - (-2)^{n-1} = (-1)^n n! - 1$ .

**2.28** L'échange des sommations entre la deuxième et la troisième ligne n'est pas autorisé ; les termes de cette somme ne convergent pas absolument. Tout le reste est parfaitement correct, sauf qu'on aurait peut-être dû écrire le résultat de  $\sum_{k \geq 1} [k=j-1] k/j$  comme ceci :  $[j-1 \geq 1] (j-1)/j$ , et le simplifier explicitement.

*Par opposition à imparfaitement correct.*

**2.29** Développez en éléments simples pour obtenir

$$\frac{k}{4k^2 - 1} = \frac{1}{4} \left( \frac{1}{2k+1} + \frac{1}{2k-1} \right).$$

Maintenant, grâce au facteur  $(-1)^k$ , chacune des deux parties de chaque terme se neutralise avec un de ses voisins, et on trouve  $-1/4 + (-1)^n/(8n+4)$ .

**2.30**  $\sum_a^b x \delta x = \frac{1}{2}(b^2 - a^2) = \frac{1}{2}(b-a)(b+a-1)$ . Par conséquent,

$$(b-a)(b+a-1) = 2100 = 2^2 \cdot 3 \cdot 5^2 \cdot 7.$$

Il y a autant de solutions que de façons d'écrire  $2100 = x \cdot y$ , où  $x$  est pair et  $y$  impair, si on pose  $a = \frac{1}{2}|x-y| + \frac{1}{2}$  et  $b = \frac{1}{2}(x+y) + \frac{1}{2}$ . Ainsi, le nombre de solutions est égal au nombre de diviseurs de  $3 \cdot 5^2 \cdot 7$ , soit 12. De manière générale, il y a  $\prod_{p>2} (n_p + 1)$  façons de représenter  $\prod_p p^{n_p}$ , où les  $p$  sont des nombres premiers.

**2.31**  $\sum_{j,k \geq 2} j^{-k} = \sum_{j \geq 2} 1/j^2(1 - 1/j) = \sum_{j \geq 2} 1/j(j-1)$ . De manière similaire, on trouve que la seconde somme est égale à  $3/4$ .

**2.32** Si  $2n \leq x < 2n + 1$ , les sommes font  $0 + \dots + n + (x-n-1) + \dots + (x-2n) = n(x-n) = (x-1) + (x-3) + \dots + (x-2n+1)$ . Si  $2n - 1 \leq x < 2n$ , elles sont toutes deux égales à  $n(x-n)$ . (Avec les notations du chapitre 3, la formule  $\lfloor \frac{1}{2}(x+1) \rfloor (x - \lfloor \frac{1}{2}(x+1) \rfloor)$  est valable dans les deux cas).

**2.33** Si  $K$  est vide,  $\bigwedge_{k \in K} a_k = \infty$ . Voici les analogies de base :

$$\begin{aligned} \sum_{k \in K} c a_k &= c \sum_{k \in K} a_k &\longleftrightarrow \bigwedge_{k \in K} (c + a_k) &= c + \bigwedge_{k \in K} a_k \\ \sum_{k \in K} (a_k + b_k) &= \sum_{k \in K} a_k + \sum_{k \in K} b_k &\longleftrightarrow \bigwedge_{k \in K} \min(a_k, b_k) \\ &&&= \min\left(\bigwedge_{k \in K} a_k, \bigwedge_{k \in K} b_k\right) \\ \sum_{k \in K} a_k &= \sum_{p(k) \in K} a_{p(k)} &\longleftrightarrow \bigwedge_{k \in K} a_k &= \bigwedge_{p(k) \in K} a_{p(k)} \\ \sum_{\substack{j \in J \\ k \in K}} a_{j,k} &= \sum_{j \in J} \sum_{k \in K} a_{j,k} &\longleftrightarrow \bigwedge_{\substack{j \in J \\ k \in K}} a_{j,k} &= \bigwedge_{j \in J} \bigwedge_{k \in K} a_{j,k} \\ \sum_{k \in K} a_k &= \sum_k a_k [k \in K] &\longleftrightarrow \bigwedge_{k \in K} a_k &= \bigwedge_k a_k \cdot \infty^{[k \notin K]} \end{aligned}$$

Pour diriger  
la somme vers  
n'importe quelle  
valeur voulue, il  
suffit d'une permu-  
tation qui absorbe  
les termes d'un signe  
plus vite que ceux  
de l'autre.

**2.34** Soit  $K^+ = \{k \mid a_k \geq 0\}$  et  $K^- = \{k \mid a_k < 0\}$ . Si, par exemple,  $n$  est impair, prenons  $F_n = F_{n-1} \cup E_n$ , où  $E_n \subseteq K^-$  est choisi assez grand pour que

$$\sum_{k \in (F_{n-1} \cap K^+)} a_k - \sum_{k \in E_n} (-a_k) < A^-.$$

**2.35** On montre que la somme de Goldbach est égale à

$$\sum_{m,n \geq 2} m^{-n} = \sum_{m \geq 2} \frac{1}{m(m-1)} = 1$$

de la façon suivante. On remarque d'abord qu'elle est égale à une somme de séries géométriques,  $\sum_{k \in P, l \geq 1} k^{-l}$ . Par conséquent, on aura prouvé le résultat final si on sait trouver une bijection entre l'ensemble des couples  $(m, n)$  tels que  $m, n \geq 2$  et l'ensemble des couples  $(k, l)$  tels que  $k \in P$  et  $l \geq 1$ , de sorte que  $m^n = k^l$  pour toute paire de couples en correspondance. Si  $m \notin P$ , posons  $(m, n) \longleftrightarrow (m^n, 1)$ ; et si  $m = a^b \in P$ , posons  $(m, n) \longleftrightarrow (a^n, b)$ .

**2.36** (a) Par définition,  $g(n) - g(n-1) = f(n)$ . (b) D'après (a),  $g(g(n)) - g(g(n-1)) = \sum_k f(k)[g(n-1) < k \leq g(n)] = n(g(n) - g(n-1)) = nf(n)$ .

(c) D'après (a) de nouveau,  $g(g(g(n))) - g(g(g(n-1)))$  est égal à

$$\begin{aligned} \sum_k f(k) [g(g(n-1)) < k \leq g(g(n))] \\ = \sum_{j,k} j [j = f(k)] [g(g(n-1)) < k \leq g(g(n))] \\ = \sum_{j,k} j [j = f(k)] [g(n-1) < j \leq g(n)] \\ = \sum_j j (g(j) - g(j-1)) [g(n-1) < j \leq g(n)] \\ = \sum_j j f(j) [g(n-1) < j \leq g(n)] = n \sum_j j [g(n-1) < j \leq g(n)]. \end{aligned}$$

Colin Mallows fait remarquer que cette suite peut être aussi définie par la récurrence

$$f(1) = 1; \quad f(n+1) = 1 + f(n + 1 - f(f(n))), \quad \text{pour } n \geq 0.$$

**2.37** (RLG pense que ça ne marchera probablement pas ; DEK pense que ça marchera probablement ; OP ne se mouille pas).

**3.1**  $m = \lfloor \lg n \rfloor$ ;  $l = n - 2^m = n - 2^{\lfloor \lg n \rfloor}$ .

**3.2** (a)  $[x + 0,5]$ . (b)  $\lceil x - 0,5 \rceil$ .

**3.3** C'est égal à  $\lfloor mn - \{m\alpha\}n/\alpha \rfloor = mn - 1$ , car  $0 < \{m\alpha\} < 1$ .

**3.4** Quelque chose qui ne nécessite pas de preuve, et dont la réponse peut être trouvée avec de l'intuition et un peu de chance.

**3.5** On a  $\lfloor nx \rfloor = \lfloor n\lfloor x \rfloor + n\{x\} \rfloor = n\lfloor x \rfloor + \lfloor n\{x\} \rfloor$  d'après (3.8) et (3.6). Par conséquent,  $\lfloor nx \rfloor = n\lfloor x \rfloor \iff \lfloor n\{x\} \rfloor = 0 \iff 0 \leq n\{x\} < 1 \iff \{x\} < 1/n$ , car  $n$  est un entier positif. (Remarquez que, dans ce cas,  $n\lfloor x \rfloor \leq \lfloor nx \rfloor$  pour tout  $x$ ).

**3.6**  $\lfloor f(x) \rfloor = \lfloor f(\lceil x \rceil) \rfloor$ .

**3.7**  $\lfloor n/m \rfloor + n \bmod m$ .

**3.8** Si toutes les boîtes contiennent  $< \lceil n/m \rceil$  objets, alors  $n \leq (\lceil n/m \rceil - 1)m$ , donc  $n/m + 1 \leq \lceil n/m \rceil$ , ce qui contredit (3.5). L'autre preuve est similaire.

**3.9** On a  $m/n - 1/q = (n \text{ marmot } m)/qn$ . Le processus se termine nécessairement, car  $0 \leq (n \text{ marmot } m) < m$ . Les dénominateurs de la représentation croissent strictement, car  $qn/(n \text{ marmot } m) > q$ ; donc ils sont distincts.

**3.10**  $\lceil x + \frac{1}{2} \rceil - \lceil (2x + 1)/4 \rceil$  n'est pas un entier] est l'entier le plus proche de  $x$  si  $\{x\} \neq \frac{1}{2}$ ; sinon, c'est le plus proche entier pair (voir l'exercice 2). Ainsi, la formule donne une façon "objective" d'arrondir un nombre.

**3.11** Si  $n$  est un entier,  $\alpha < n < \beta \iff \lfloor \alpha \rfloor < n < \lceil \beta \rceil$ . Le nombre d'entiers satisfaisant  $a < n < b$  lorsque  $a$  et  $b$  sont des entiers est égal à  $(b - a - 1)$  ( $b > a$ ). On obtiendrait donc une réponse fausse dans le cas où  $\alpha = \beta$  avec  $\alpha, \beta$  entiers.

**3.12** En soustrayant  $\lfloor n/m \rfloor$  des deux côtés, d'après (3.6), on obtient  $\lceil (n \bmod m)/m \rceil = \lfloor ((n \bmod m) + m - 1)/m \rfloor$ . Les deux côtés sont maintenant égaux à  $\lceil n \bmod m \rceil$ , car  $0 \leq n \bmod m < m$ .

On peut aussi donner une preuve plus courte, mais moins directe, en observant que le premier terme de (3.24) doit être égal au dernier terme de (3.25).

**3.13** S'ils forment une partition, on doit avoir  $N(\alpha, n) + N(\beta, n) = n$  pour tout entier strictement positif  $n$ . En utilisant la formule donnée dans le chapitre pour  $N(\alpha, n)$ , on en déduit que  $1/\alpha + 1/\beta = 1$ . Par conséquent,  $\alpha$  et  $\beta$  sont soit tous deux rationnels, soit tous deux irrationnels. S'ils sont irrationnels, on obtient bien une partition, comme cela est démontré dans le chapitre. Par contre, s'ils sont tous deux irrationnels, on peut les écrire avec un même numérateur  $m$ , et la valeur  $m - 1$  n'appartient à aucun des deux spectres, tandis que  $m$  apparaît dans les deux. (En revanche, Golomb [151] a montré que les ensembles  $\{\lfloor n\alpha \rfloor \mid n \geq 1\}$  et  $\{\lceil n\beta \rceil - 1 \mid n \geq 1\}$  forment toujours une partition lorsque  $1/\alpha + 1/\beta = 1$ ).

**3.14** D'après (3.22), c'est évident si  $ny = 0$ ; sinon, c'est vrai aussi, d'après (3.21) et (3.6).

**3.15** Remplacez  $n$  par  $\lceil mx \rceil$  dans (3.24):  $\lceil mx \rceil = \lceil x \rceil + \lceil x - \frac{1}{m} \rceil + \cdots + \lceil x - \frac{m-1}{m} \rceil$ .

**3.16** Pour prouver la formule  $n \bmod 3 = 1 + \frac{1}{3}((\omega - 1)\omega^n - (\omega + 2)\omega^{2n})$ , il suffit de la vérifier pour  $0 \leq n < 3$ .

Nous verrons dans l'exercice 7.25 une formule générale pour  $n \bmod m$ , pour tout entier strictement positif  $m$ .

**3.17**  $\sum_{j,k} [0 \leq k < m][1 \leq j \leq x + k/m] = \sum_{j,k} [0 \leq k < m][1 \leq j \leq \lceil x \rceil] \times [k \geq m(j - x)] = \sum_{1 \leq j \leq \lceil x \rceil} \sum_k [0 \leq k < m] - \sum_{j=\lceil x \rceil} \sum_k [0 \leq k < m(j - x)] = m\lceil x \rceil - \lceil m(\lceil x \rceil - x) \rceil = -\lceil -mx \rceil = \lfloor mx \rfloor$ .

**3.18** On a

$$S = \sum_{0 \leq j < \lceil n\alpha \rceil} \sum_{k \geq n} [j\alpha^{-1} \leq k < (j + v)\alpha^{-1}] .$$

Si  $j \leq n\alpha - 1 \leq n\alpha - v$ , il n'y a pas de contribution car  $(j + v)\alpha^{-1} \leq n$ . Par conséquent, la seule valeur de  $j$  qui peut compter est  $j = \lfloor n\alpha \rfloor$ , et dans ce cas on trouve  $\lceil (\lfloor n\alpha \rfloor + v)\alpha^{-1} \rceil - n \leq \lceil v\alpha^{-1} \rceil$ .

**3.19** Cela arrive si et seulement si  $b$  est un nombre entier. (Si  $b$  est un entier,  $\log_b x$  est une fonction continue et croissante qui ne prend des valeurs entières que pour des arguments entiers. Si  $b$  n'est pas un entier, la formule est fausse lorsque  $x = b$ ).

**3.20** On a  $\sum_k kx[\alpha \leq kx \leq \beta] = x \sum_k k[\lceil \alpha/x \rceil \leq k \leq \lfloor \beta/x \rfloor]$ , ce qui donne  $\frac{1}{2}x(\lfloor \beta/x \rfloor \lfloor \beta/x + 1 \rfloor - \lceil \alpha/x \rceil \lceil \alpha/x - 1 \rceil)$ .

**3.21** Si  $10^n \leq 2^M < 10^{n+1}$ , il y a exactement  $n + 1$  puissances de 2 satisfaisant la propriété, car, pour tout entier  $k > 0$ , il existe exactement une puissance de deux à  $k$  chiffres qui la respecte. La réponse est donc  $1 + \lfloor M \log 2 \rfloor$ .

Note : il est plus difficile de trouver le nombre de puissances de 2 commençant par le chiffre  $l$ , pour  $l > 1$ ; c'est  $\sum_{0 \leq n \leq M} (\lfloor n \log 2 - \log l \rfloor - \lfloor n \log 2 - \log(l+1) \rfloor)$ .

**3.22** En posant  $n = 2^{j-1}q$  où  $q$  est le plus petit nombre impair possible, on trouve que tous les termes de  $S_n$  et  $S_{n-1}$  (resp.  $T_n$  et  $T_{n-1}$ ) sont identiques, sauf le  $j$ ième. On obtient  $S_n = S_{n-1} + 1$  et  $T_n = T_{n-1} + 2^k q$ , donc  $S_n = n$  et  $T_n = n(n+1)$ .

**3.23**  $X_n = m \iff \frac{1}{2}m(m-1) < n \leq \frac{1}{2}m(m+1) \iff m^2 - m + \frac{1}{4} < 2n < m^2 + m + \frac{1}{4} \iff m - \frac{1}{2} < \sqrt{2n} < m + \frac{1}{2}$ .

**3.24** Soit  $\beta = \alpha/(\alpha + 1)$ . Il y a exactement une occurrence de chaque entier positif ou nul de plus dans  $\text{Spec}(\beta)$  que dans  $\text{Spec}(\alpha)$ . Pourquoi ? Parce que  $N(\beta, n) = N(\alpha, n) + n + 1$ .

**3.25** Continuons la preuve commencée dans le chapitre. Si nous trouvions une valeur de  $m$  telle que  $K_m \leq m$ , nous pourrions réfuter l'inégalité en  $n + 1$ , pour  $n = 2m + 1$  (ainsi que pour  $n = 3m + 1$  et  $n = 3m + 2$ ). Cependant, pour qu'un tel  $m = n' + 1$  existe, il faudrait que  $2K_{\lfloor n'/2 \rfloor} \leq n'$  ou  $3K_{\lfloor n'/3 \rfloor} \leq n'$ , c'est-à-dire que

$$K_{\lfloor n'/2 \rfloor} \leq \lfloor n'/2 \rfloor \quad \text{ou} \quad K_{\lfloor n'/3 \rfloor} \leq \lfloor n'/3 \rfloor.$$

On peut continuer ainsi, pour trouver finalement  $K_0 \leq 0$ ; mais c'est absurde car  $K_0 = 1$ .

Ce qu'il nous faut prouver en réalité, c'est que  $K_n$  est strictement supérieur à  $n$ , pour tout  $n > 0$ . C'est en fait facile à prouver par induction, bien que ce résultat soit plus fort que celui que nous n'avons pas pu démontrer !

(Cet exercice nous donne une bonne leçon. Il porte plus sur la nature de l'induction que sur les propriétés de la partie entière par défaut).

*"In trying to devise a proof by mathematical induction, you may fail for two opposite reasons. You may fail because you try to prove too much: Your P(n) is too heavy a burden. Yet you may also fail because you try to prove too little: Your P(n) is too weak a support. In general, you have to balance the statement of your theorem so that the support is just enough for the burden."*

— G. Pólya [297]

**3.26** Par induction, en prenant une hypothèse plus forte :

$$D_n^{(q)} \leq (q-1) \left( \left( \frac{q}{q-1} \right)^{n+1} - 1 \right), \quad \text{pour } n \geq 0.$$

**3.27** Si  $D_n^{(3)} = 2^m b - a$ , où  $a$  est égal à 0 ou à 1, alors  $D_{n+m}^{(3)} = 3^m b - a$ .

**3.28** Il faut remarquer que  $a_n = m^2$  implique  $a_{n+2k+1} = (m+k)^2 + m - k$  et  $a_{n+2k+2} = (m+k)^2 + 2m$ , pour  $0 \leq k \leq m$ ; donc  $a_{n+2m+1} = (2m)^2$ . La solution peut s'écrire sous une jolie forme, découverte par Carl Witty :

$$a_{n-l} = 2^l + \left\lfloor \left( \frac{n-l}{2} \right)^2 \right\rfloor, \quad \text{lorsque } 2^l + l \leq n < 2^{l+1} + l + 1.$$

**3.29**  $D(\alpha', \lfloor \alpha n \rfloor)$  est au plus égal au maximum de la valeur absolue de

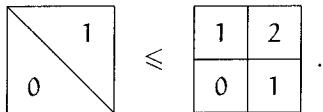
$$s(\alpha', \lfloor n\alpha \rfloor, v') = -s(\alpha, n, v) - S + \epsilon + \{0 \text{ ou } 1\} + v' - \{0 \text{ ou } 1\}.$$

**3.30**  $X_n = \alpha^{2^n} + \alpha^{-2^n}$ , par induction; et  $X_n$  est un entier.

**3.31** Voici une preuve "élégante", "impressionnante", écrite de telle façon que rien ne permet de savoir comment on l'a trouvée.

$$\begin{aligned} \lfloor x \rfloor + \lfloor y \rfloor + \lfloor x+y \rfloor &= \lfloor x + \lfloor y \rfloor \rfloor + \lfloor x+y \rfloor \\ &\leq \lfloor x + \frac{1}{2} \lfloor 2y \rfloor \rfloor + \lfloor x + \frac{1}{2} \lfloor 2y \rfloor + \frac{1}{2} \rfloor \\ &= \lfloor 2x + \lfloor 2y \rfloor \rfloor = \lfloor 2x \rfloor + \lfloor 2y \rfloor. \end{aligned}$$

Toutefois, il existe aussi une preuve simple et visuelle, basée sur le fait qu'on n'a besoin de considérer que le cas  $0 \leq x, y < 1$ . Les fonctions ressemblent alors à ceci :



Il est possible de démontrer un résultat un peu plus fort :

$$\lfloor x \rfloor + \lfloor y \rfloor + \lfloor x+y \rfloor \leq \lfloor 2x \rfloor + \lfloor 2y \rfloor;$$

mais en fait il n'est plus fort que lorsque  $\{x\} = \frac{1}{2}$ . En remplaçant  $(x, y)$  par  $(-x, x+y)$  dans cette identité, et en appliquant la règle de réflexivité (3.4), on obtient

$$\lfloor y \rfloor + \lfloor x+y \rfloor + \lfloor 2x \rfloor \leq \lfloor x \rfloor + \lfloor 2x+2y \rfloor.$$

**3.32** Soit  $f(x)$  la somme en question. Comme  $f(x) = f(-x)$ , nous pouvons supposer que  $x \geq 0$ . La somme est définie pour tout réel  $x$ , car ses termes sont bornés par  $2^k$  lorsque  $k \rightarrow -\infty$  et par  $x^2/2^k$  lorsque  $k \rightarrow +\infty$ .

On a  $f(2x) = 2 \sum_k 2^{k-1} \|x/2^{k-1}\|^2 = 2f(x)$ . Soit  $f(x) = l(x) + r(x)$ , où  $l(x)$  désigne la partie où  $k \leq 0$  et  $r(x)$  celle où  $k > 0$ . Alors  $l(x+1) = l(x)$  et  $l(x) \leq 1/2$  pour tout  $x$ . Si  $0 \leq x < 1$ , on a  $r(x) = x^2/2 + x^2/4 + \dots = x^2$  et  $r(x+1) = (x-1)^2/2 + (x+1)^2/4 + (x+1)^2/8 + \dots = x^2 + 1$ . Ainsi,  $f(x+1) = f(x) + 1$  pour  $0 \leq x < 1$ .

On peut maintenant prouver par induction que si  $0 \leq x < 1$ , alors  $f(x+n) = f(x) + n$  pour tout entier  $n \geq 0$ . En particulier,  $f(n) = n$ . Donc, plus généralement,  $f(x) = 2^{-m}f(2^m x) = 2^{-m}[2^m x] + 2^{-m}f(\{2^m x\})$ . Or,  $f(\{2^m x\}) = l(\{2^m x\}) + r(\{2^m x\}) \leq \frac{1}{2} + 1$ ; donc  $|f(x) - x| \leq |2^{-m}[2^m x] - x| + 2^{-m} \cdot \frac{3}{2} \leq 2^{-m} \cdot \frac{5}{2}$  pour tout entier  $m$ .

On arrive donc inéluctablement à la conclusion suivante :  $f(x) = |x|$  pour tout réel  $x$ .

**3.33** Soit  $r = n - \frac{1}{2}$  le rayon du cercle. (a) L'échiquier est traversé par  $2n - 1$  segments horizontaux et  $2n - 1$  segments verticaux, allant de bord en bord. Le cercle croise deux fois chacun de ces segments. Comme  $r^2$  n'est pas entier, le théorème de Pythagore nous permet d'affirmer que le cercle ne peut pas passer par le coin d'une case. Par conséquent, le nombre de cellules traversées par le cercle est égal au nombre de fois qu'il croise les segments, soit  $8n - 4 = 8r$  (c'est aussi le nombre de cases du bord de l'échiquier). (b)  $f(n, k) = 4 \lfloor \sqrt{r^2 - k^2} \rfloor$ .

On déduit de (a) et (b) que

$$\frac{1}{4}\pi r^2 - 2r \leq \sum_{0 < k < r} \lfloor \sqrt{r^2 - k^2} \rfloor \leq \frac{1}{4}\pi r^2, \quad r = n - \frac{1}{2}.$$

La recherche de bornes plus précises pour cette somme est un problème fameux en théorie des nombres. Il a été étudié par Gauss et par bien d'autres (voir Dickson [78, volume 2, chapitre 6]).

**3.34** (a) Soit  $m = \lceil \lg n \rceil$ . Pour simplifier les calculs aux bornes, ajoutons  $2^m - n$  termes à  $f(n)$  :

$$\begin{aligned} f(n) + (2^m - n)m &= \sum_{k=1}^{2^m} \lceil \lg k \rceil = \sum_{j,k} j[j = \lceil \lg k \rceil][1 \leq k \leq 2^m] \\ &= \sum_{j,k} j[2^{j-1} < k \leq 2^j][1 \leq j \leq m] \\ &= \sum_{j=1}^m j 2^{j-1} = 2^m(m-1) + 1. \end{aligned}$$

Par conséquent,  $f(n) = nm - 2^m + 1$ .

(b) On sait que  $\lceil n/2 \rceil = \lfloor (n+1)/2 \rfloor$ , et il s'ensuit que la solution de la récurrence  $g(n) = a(n) + g(\lceil n/2 \rceil) + g(\lfloor n/2 \rfloor)$  doit satisfaire  $\Delta g(n) = \Delta a(n) + \Delta g(\lfloor n/2 \rfloor)$ . En particulier, lorsque  $a(n) = n-1$ , l'égalité  $\Delta f(n) =$

$1 + \Delta f(\lfloor n/2 \rfloor)$  est satisfaite par le nombre de chiffres de la représentation binaire de  $n$ , soit  $\lceil \lg(n+1) \rceil$ . Il suffit maintenant de passer de  $\Delta$  à  $\Sigma$ .

On peut aussi trouver une solution plus directe à partir des identités suivantes :  $\lceil \lg 2j \rceil = \lceil \lg j \rceil + 1$  et  $\lceil \lg(2j-1) \rceil = \lceil \lg j \rceil + \lfloor j > 1 \rfloor$ , pour  $j \geq 1$ .

**3.35**  $(n+1)^2 n! e = A_n + (n+1)^2 + (n+1) + B_n$ , où

$$A_n = \frac{(n+1)^2 n!}{0!} + \frac{(n+1)^2 n!}{1!} + \cdots + \frac{(n+1)^2 n!}{(n-1)!}$$

est un multiple de  $n$  et

$$\begin{aligned} B_n &= \frac{(n+1)^2 n!}{(n+2)!} + \frac{(n+1)^2 n!}{(n+3)!} + \cdots \\ &= \frac{n+1}{n+2} \left( 1 + \frac{1}{n+3} + \frac{1}{(n+3)(n+4)} + \cdots \right) \\ &< \frac{n+1}{n+2} \left( 1 + \frac{1}{n+3} + \frac{1}{(n+3)(n+3)} + \cdots \right) \\ &= \frac{(n+1)(n+3)}{(n+2)^2} \end{aligned}$$

est plus petit que 1. La réponse est donc  $2 \bmod n$ .

**3.36** La somme est égale à

$$\begin{aligned} \sum_{k,l,m} 2^{-l} 4^{-m} [m = \lceil \lg l \rceil] [l = \lceil \lg k \rceil] [1 < k < 2^{2^n}] \\ &= \sum_{k,l,m} 2^{-l} 4^{-m} [2^m \leq l < 2^{m+1}] [2^l \leq k < 2^{l+1}] [0 \leq m < n] \\ &= \sum_{l,m} 4^{-m} [2^m \leq l < 2^{m+1}] [0 \leq m < n] \\ &= \sum_m 2^{-m} [0 \leq m < n] = 2(1 - 2^{-n}). \end{aligned}$$

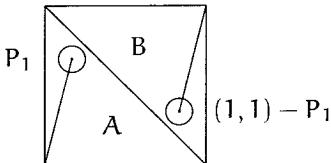
**3.37** Considérez d'abord le cas  $m < n$ , qui se décompose en deux sous-cas, selon que  $m < \frac{1}{2}n$  ou non ; puis montrez que les deux membres de l'égalité sont modifiés de la même manière lorsqu'on augmente  $m$  de  $n$  unités.

**3.38** Il existe au plus un  $x_k$  qui n'est pas entier. Supprimons tous les  $x_k$  entiers de l'expression et supposons qu'il nous en reste  $n$ . Lorsque  $\{x\} \neq 0$ , la valeur moyenne de  $\{mx\}$  quand  $m \rightarrow \infty$  est comprise entre  $\frac{1}{4}$  et  $\frac{1}{2}$  ; donc la valeur moyenne de  $\{mx_1\} + \cdots + \{mx_n\} - \{mx_1 + \cdots + mx_n\}$  ne peut pas être nulle si  $n > 1$ .

Le problème, c'est que l'argument que nous venons de donner repose sur un théorème difficile qui concerne la distribution uniforme. On peut

En fait, ce n'est qu'un problème de niveau 4, malgré la façon dont il est présenté.

aussi trouver une preuve élémentaire. La voici pour  $n = 2$  : soit  $P_m$  le point  $(\{mx\}, \{my\})$ . Partageons le carré unitaire  $0 \leq x, y < 1$  en régions triangulaires  $A$  et  $B$ , selon que  $x + y < 1$  ou  $x + y \geq 1$ . Nous voulons montrer que si  $\{x\}$  et  $\{y\}$  sont non nuls, il existe un  $m$  tel que  $P_m \in B$ . Si  $P_1 \in B$ , c'est gagné. Sinon, il existe un disque  $D$  de rayon  $\epsilon > 0$  et de centre  $P_1$  tel que  $D \subseteq A$ . D'après le principe des boîtes de Dirichlet, pour  $N$  suffisamment grand, la suite  $P_1, \dots, P_N$  contient deux points tels que  $|P_k - P_j| < \epsilon$  et  $k > j$ .



Il s'ensuit que  $P_{k-j-1}$  est à distance plus petite que  $\epsilon$  de  $(1,1) - P_1$  ; donc  $P_{k-j-1} \in B$ .

**3.39** Remplacez  $j$  par  $b - j$  et ajoutez le terme  $j = 0$  à la somme, de façon à pouvoir utiliser l'exercice 15 pour calculer la somme sur  $j$ . Le résultat,

$$\lceil x/b^k \rceil - \lceil x/b^{k+1} \rceil + b - 1,$$

se simplifie “télescopiquement” lorsqu'on le somme sur  $k$ .

**3.40** Soit  $\lfloor 2\sqrt{n} \rfloor = 4k + r$ , avec  $-2 \leq r < 2$ , et soit  $m = \lfloor \sqrt{n} \rfloor$ . Ce qui suit peut être prouvé par induction :

segment	$r$	$m$	$x$	$y$	si et seulement si
$W_k$	$-2$	$2k-1$	$m(m+1)-n-k$	$k$	$(2k-1)(2k-1) \leq n \leq (2k-1)(2k)$
$S_k$	$-1$	$2k-1$	$-k$	$m(m+1)-n+k$	$(2k-1)(2k) < n < (2k)(2k)$
$E_k$	$0$	$2k$	$n-m(m+1)+k$	$-k$	$(2k)(2k) \leq n \leq (2k)(2k+1)$
$N_k$	$1$	$2k$	$k$	$n-m(m+1)-k$	$(2k)(2k+1) < n < (2k+1)(2k+1)$

Ainsi, lorsque  $k \geq 1$ ,  $W_k$  est un segment de longueur  $2k$ , qui est emprunté de droite à gauche par le chemin, et  $y(n) = k$  ;  $S_k$  est un segment de longueur  $2k-2$ , emprunté de haut en bas, et  $x(n) = -k$  ; etc. (a) Voici donc la formule recherchée :

$$y(n) = (-1)^m \left( (n - m(m+1)) \cdot [\lfloor 2\sqrt{n} \rfloor \text{ est impair}] - \lceil \frac{1}{2}m \rceil \right).$$

(b) Pour tout segment,  $k = \max(|x(n)|, |y(n)|)$ . Pour les segments  $W_k$  et  $S_k$ , on a  $x < y$  et  $n+x+y = m(m+1) = (2k)^2 - 2k$  ; pour les segments  $E_k$  et  $N_k$ , on a  $x \geq y$  et  $n-x-y = m(m+1) = (2k)^2 + 2k$ . Par conséquent, le signe est  $(-1)^{|x(n)|+|y(n)|}$ .

**3.41** Comme  $1/\phi + 1/\phi^2 = 1$ , les suites correspondantes partitionnent bien l'ensemble des entiers strictement positifs. Puisque  $f$  et  $g$  sont déterminées de façon unique par la condition  $g(n) = f(f(n)) + 1$ , il nous suffit de prouver que  $\lfloor [n\phi]\phi \rfloor + 1 = \lfloor n\phi^2 \rfloor$  pour tout  $n > 0$ . Cela se déduit de l'exercice 3, en prenant  $\alpha = \phi$  et  $n = 1$ .

**3.42** Non. On peut prouver, par un argument similaire à celui donné dans le chapitre pour le problème du partitionnement par  $\text{Spec}(\sqrt{2})$  et  $\text{Spec}(2+\sqrt{2})$ , ou pour l'exercice 13, qu'il y a une tripartition si et seulement si  $1/\alpha + 1/\beta + 1/\gamma = 1$  et

$$\left\{ \frac{n+1}{\alpha} \right\} + \left\{ \frac{n+1}{\beta} \right\} + \left\{ \frac{n+1}{\gamma} \right\} = 1$$

pour tout  $n > 0$ . Cependant, d'après le théorème sur la distribution uniforme, la valeur moyenne de  $\{(n+1)/\alpha\}$  est égale à  $1/2$  si  $\alpha$  est irrationnel. Les paramètres ne peuvent pas être tous rationnels, et si  $\gamma = m/n$  la moyenne est égale à  $3/2 - 1/(2n)$ . Donc  $\gamma$  doit être un entier, mais cela ne marche pas non plus. (Il existe aussi une preuve qui ne fait pas appel au théorème sur la distribution uniforme ; voir [155]).

**3.43** La première étape du dépliage de la récurrence de  $K_n$  nous donne le minimum des quatre nombres  $1 + a + a \cdot b \cdot K_{\lfloor (n-1-a)/(a \cdot b) \rfloor}$ , où chacun des nombres  $a$  et  $b$  vaut soit 2, soit 3. (On obtient cette formule simple en appliquant (3.11) pour se débarrasser des parties entières internes et en utilisant l'identité  $x + \min(y, z) = \min(x+y, x+z)$  ; on ne doit pas prendre en compte les termes d'indice négatif, c'est-à-dire ceux pour lesquels  $n-1-a < 0$ ).

En continuant ainsi, on arrive à la constatation suivante :  $K_n$  est le plus petit nombre  $> n$  dans le multi-ensemble  $S$  de tous les nombres de la forme

$$1 + a_1 + a_1 a_2 + a_1 a_2 a_3 + \cdots + a_1 a_2 a_3 \dots a_m,$$

où  $m \geq 0$  et chaque  $a_k$  est égal soit à 2, soit à 3. Ainsi,

$$S = \{1, 3, 4, 7, 9, 10, 13, 15, 19, 21, 22, 27, 28, 31, 31, \dots\};$$

le nombre 31 est “deux fois” dans  $S$  car il a deux représentations possibles,  $1+2+4+8+16 = 1+3+9+18$ . (Notons en passant que Michael Fredman [134] a démontré que  $\lim_{n \rightarrow \infty} K_n/n = 1$  ; en d'autres termes, il n'y a pas d'énorme trou dans  $S$ ).

**3.44** Soit  $d_n^{(q)} = D_{n-1}^{(q)}$  marmot  $(q-1)$ , tel que  $D_n^{(q)} = (qD_{n-1}^{(q)} + d_n^{(q)})/(q-1)$  et  $a_n^{(q)} = \lceil D_{n-1}^{(q)} / (q-1) \rceil$ . Alors  $D_{k-1}^{(q)} \leq (q-1)n \iff a_k^{(q)} \leq n$  et le résultat s'ensuit. (Cette solution a été trouvée par Euler [116] ; il

a calculé les  $a_k$ , puis les les  $d_k$ , sans s'apercevoir qu'une seule suite  $D_n^{(q)}$  aurait suffi).

**3.45** Soit  $\alpha > 1$  tel que  $\alpha + 1/\alpha = 2m$ . Alors on trouve  $2Y_n = \alpha^{2^n} + \alpha^{-2^n}$ , *Trop facile.* et il s'ensuit que  $Y_n = \lceil \alpha^{2^n}/2 \rceil$ .

**3.46** La formule de la suggestion de déduit de (3.9), car  $2n(n+1) = \lfloor 2(n+\frac{1}{2})^2 \rfloor$ . Soit  $n+\theta = (\sqrt{2}^l + \sqrt{2}^{l-1})m$  et  $n'+\theta' = (\sqrt{2}^{l+1} + \sqrt{2}^l)m$ , où  $0 \leq \theta, \theta' < 1$ . Alors  $\theta' = 2\theta \bmod 1 = 2\theta - d$ , où  $d$  est égal à 0 ou 1. Nous voulons prouver que  $n' = \lfloor \sqrt{2}(n + \frac{1}{2}) \rfloor$ ; cette égalité est vraie si et seulement si

$$0 \leq \theta'(2 - \sqrt{2}) + \sqrt{2}(1 - d) < 2.$$

Pour résoudre la récurrence, remarquons que  $\text{Spec}(1 + 1/\sqrt{2})$  et  $\text{Spec}(1 + \sqrt{2})$  forment une partition de l'ensemble des entiers strictement positifs ; par conséquent, tout entier strictement positif  $a$  peut s'écrire de façon unique sous la forme  $a = \lfloor (\sqrt{2}^l + \sqrt{2}^{l-1})m \rfloor$ , où  $l$  est un entier  $\geq 0$  et  $m$  est un entier impair. Il s'ensuit que  $L_n = \lfloor (\sqrt{2}^{l+n} + \sqrt{2}^{l+n-1})m \rfloor$ .

**3.47** (a)  $c = -\frac{1}{2}$ . (b)  $c$  est un entier. (c)  $c = 0$ . (d)  $c$  est quelconque. Voir la réponse à l'exercice 1.2.4–40 dans [207] pour des résultats plus généraux.

**3.48** Soient  $x^{(0)} = 1$  et  $x^{(k+1)} = x[x^{(k)}]$ ; soient encore  $a_k = \{x^{(k)}\}$  et  $b_k = \lfloor x^{(k)} \rfloor$ . L'identité en question s'écrit alors  $x^3 = 3x^{(3)} + 3a_1a_2 + a_1^3 - 3b_1b_2 + b_1^3$ . Comme  $a_k + b_k = x^{(k)} = xb_{k-1}$  pour  $k \geq 0$ , on a  $(1 - xz)(1 + b_1z + b_2z^2 + \dots) = 1 - a_1z - a_2z^2 - \dots$ ; donc

$$\frac{1}{1 - xz} = \frac{1 + b_1z + b_2z^2 + \dots}{1 - a_1z - a_2z^2 - \dots}.$$

Prenons le logarithme de chacun des deux membres afin de séparer les  $a$  des  $b$ . Puis dérivons par rapport à  $z$ , pour obtenir

$$\frac{x}{1 - xz} = \frac{a_1 + 2a_2z + 3a_3z^2 + \dots}{1 - a_1z - a_2z^2 - \dots} + \frac{b_1 + 2b_2z + 3b_3z^2 + \dots}{1 + b_1z + b_2z^2 + \dots}.$$

Dans le membre gauche, le coefficient de  $z^{n-1}$  est  $x^n$ ; dans le membre droit, c'est une formule qui, pour  $n = 3$ , correspond à l'identité cherchée.

On peut aussi trouver des identités similaires pour le produit plus général  $x_0x_1\dots x_{n-1}$  [170].

**3.49** (Solution due à Heinrich Rolletschek). On peut remplacer  $(\alpha, \beta)$  par  $(\{\beta\}, \alpha + \lfloor \beta \rfloor)$  sans que  $\lfloor n\alpha \rfloor + \lfloor n\beta \rfloor$  ne soit modifié. Donc la condition  $\alpha = \{\beta\}$  est nécessaire. Elle est aussi suffisante : soit  $m = \lfloor \beta \rfloor$  le plus petit élément du multi-ensemble donné, et  $S$  le multi-ensemble obtenu en soustrayant, pour tout  $n$ , la valeur  $mn$  au  $n$ ième plus petit élément

*Voici un problème plus intéressant (non résolu à ce jour) : en supposant que  $\alpha$  et  $\beta$  sont  $< 1$ , sous quelles conditions le multi-ensemble donné permet de déterminer la paire (non ordonnée)  $\{\alpha, \beta\}$  ?*

du multi-ensemble donné. Si  $\alpha = \{\beta\}$ , alors deux éléments consécutifs quelconques de  $S$  diffèrent soit de 0 soit de 2, et par conséquent le multi-ensemble  $\frac{1}{2}S = \text{Spec}(\alpha)$  détermine  $\alpha$ .

**3.50** D'après des notes non publiées de William A. Veech, il suffit que  $\alpha\beta$ ,  $\beta$ , et 1 soient linéairement indépendants sur l'ensemble des rationnels.

**3.51** H. S. Wilf fait observer que, si nous connaissions  $f(x)$  sur tout intervalle  $(\phi \dots \phi + \epsilon)$ , l'équation fonctionnelle  $f(x^2 - 1) = f(x)^2$  permettrait de déterminer  $f(x)$  pour tout  $x \geq \phi$ .

**3.52** Il y a une infinité de façons de partitionner l'ensemble des entiers strictement positifs en deux ou plusieurs spectres généralisés avec  $\alpha_k$  irrationnel. En voici un exemple :

$$\text{Spec}(2\alpha; 0) \cup \text{Spec}(4\alpha; -\alpha) \cup \text{Spec}(4\alpha; -3\alpha) \cup \text{Spec}(\beta; 0) ;$$

mais en fait, toutes ces partitions dérivent d'une certaine manière de celle de base,  $\text{Spec}(\alpha) \cup \text{Spec}(\beta)$  (voir [158]). Les seuls exemples rationnels connus, comme

$$\text{Spec}(7; -3) \cup \text{Spec}(\frac{7}{2}; -1) \cup \text{Spec}(\frac{7}{4}; 0),$$

sont basés sur des paramètres respectant les conditions de la conjecture donnée, due à A. S. Fraenkel [128].

**3.53** On peut trouver des résultats partiels dans [95, pages 30–31]. La réponse à la question est probablement négative.

**4.1** 1, 2, 4, 6, 16, 12.

**4.2** Remarquez que  $m_p + n_p = \min(m_p, n_p) + \max(m_p, n_p)$ . La récurrence  $\text{ppcm}(m, n) = (n/(n \bmod m)) \text{ppcm}(n \bmod m, m)$  est valable, mais inutilisable en pratique pour calculer les ppcm. La meilleure méthode connue pour trouver  $\text{ppcm}(m, n)$  consiste à calculer  $\text{pgcd}(m, n)$  puis diviser  $mn$  par ce pgcd.

**4.3** C'est vrai si  $x$  est un nombre entier, mais  $\pi(x)$  est défini pour tout nombre réel. La formule correcte,

$$\pi(x) - \pi(x - 1) = [\lfloor x \rfloor \text{ est premier}] ,$$

se vérifie facilement.

**4.4** Entre  $\frac{1}{0}$  et  $\frac{0}{-1}$ , on aurait un arbre de Stern–Brocot à l'envers avec des dénominateurs négatifs, etc. Par conséquent, on obtiendrait toutes les fractions  $m/n$  telles que  $m \perp n$ . La condition  $m'n - mn' = 1$  est toujours vraie dans cette nouvelle construction. (On l'appelle la *couronne de Stern–Brocot*, à cause de la figure qu'on obtient si on forme un cycle en confondant

le  $\frac{0}{1}$  initial et le  $\frac{0}{1}$  final. On trouve d'intéressantes applications de la couronne de Stern-Brocot dans le domaine de l'informatique graphique, du fait qu'elle représente toutes les directions rationnelles du plan).

**4.5**  $L^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$  et  $R^k = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$ . Ceci est vrai même si  $k < 0$ . Nous verrons au chapitre 6 une formule pour n'importe quel produit des matrices  $L$  et  $R$ .

**4.6**  $a = b$ . (Nous avons vu au chapitre 3 que  $x \bmod 0 = x$  par définition).

**4.7** Il faudrait que  $m \bmod 10 = 0$ ,  $m \bmod 9 = k$  et  $m \bmod 8 = 1$ ; or  $m$  ne peut pas être pair et impair à la fois.

**4.8** Nous voulons que  $10x+6y \equiv 10x+y \pmod{15}$ ; donc  $5y \equiv 0 \pmod{15}$ ; donc  $y \equiv 0 \pmod{3}$ . Il faut que  $y = 0$  ou  $3$ , et  $x = 0$  ou  $1$ .

**4.9**  $3^{2k+1} \bmod 4 = 3$ , donc  $(3^{2k+1}-1)/2$  est impair. Le nombre qui nous intéresse est divisible par  $(3^7-1)/2$  et  $(3^{11}-1)/2$  (ainsi que par d'autres nombres).

$$\mathbf{4.10} \quad 999\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{37}\right) = 648.$$

**4.11**  $\sigma(0) = 1$ ;  $\sigma(1) = -1$ ;  $\sigma(n) = 0$  pour  $n > 1$ . (Les fonctions de Möbius généralisées, définies sur des structures partiellement ordonnées quelconques, présentent des propriétés intéressantes et importantes. Elles ont été découvertes par Weisner [366] puis développées par bien d'autres, notamment par Gian-Carlo Rota [313]).

**4.12** D'après (4.7) et (4.9), nous trouvons que  $\sum_{d|m} \sum_{k|d} \mu(d/k) g(k) = \sum_{k|m} \sum_{d|(m/k)} \mu(d) g(k) = \sum_{k|m} g(k)[m/k=1] = g(m)$ .

**4.13** (a)  $n_p \leq 1$  pour tout  $p$ ; (b)  $\mu(n) \neq 0$ .

**4.14** C'est vrai si  $k > 0$ . Utilisez (4.12), (4.14) et (4.15).

**4.15** Non. Par exemple,  $e_n \bmod 5 = [2 \text{ ou } 3]$ ;  $e_n \bmod 11 = [2, 3, 7, \text{ ou } 10]$ .

**4.16**  $1/e_1 + 1/e_2 + \dots + 1/e_n = 1 - 1/(e_n(e_n-1)) = 1 - 1/(e_{n+1}-1)$ .

**4.17** On a  $f_n \bmod f_m = 2$ ; donc  $\text{pgcd}(f_n, f_m) = \text{pgcd}(2, f_m) = 1$ . (Notons en passant que la relation  $f_n = f_0 f_1 \dots f_{n-1} + 2$  est similaire à la récurrence qui définit les nombres d'Euclide  $e_n$ ).

**4.18** Si  $n = qm$  et  $q$  est impair,  $2^n + 1 = (2^m + 1)(2^{n-m} - 2^{n-2m} + \dots - 2^m + 1)$ .

*En fait, quand on écrit "mod y", on "fait comme si" y était nul" en quelque sorte. S'il l'est effectivement, on n'a pas besoin de faire semblant.*

**4.19** La première somme est égale à  $\pi(n)$ , car le terme général est  $[k+1 \text{ est premier}]$ . La somme interne de la seconde somme est égale à  $\sum_{1 \leq k < m} [k \nmid m]$ , donc elle est supérieure à 1 si et seulement si  $m$  est composite; on obtient de nouveau  $\pi(n)$ . Pour finir,  $[\{m/n\}] = [n \nmid m]$ , donc la troisième somme est une application du théorème de Wilson. Bien entendu, il faudrait être fou pour calculer  $\pi(n)$  avec une de ces formules.

**4.20** Soient  $p_1 = 2$  et  $p_n$  le plus petit nombre premier supérieur à  $2^{p_{n-1}}$ . Alors  $2^{p_{n-1}} < p_n < 2^{p_{n-1}+1}$ . On peut donc prendre  $b = \lim_{n \rightarrow \infty} \lg^{(n)} p_n$ , où  $\lg^{(n)}$  est la fonction  $\lg$  itérée  $n$  fois. La valeur numérique donnée provient du fait que  $p_2 = 5$  et  $p_3 = 37$ . En calculant  $p_4 = 2^{37} + 9$ , on trouve une valeur plus précise

$$b \approx 1,2516475977905$$

(mais pas d'indice pour trouver  $p_5$ ).

**4.21** D'après le postulat de Bertrand,  $P_n < 10^n$ . Soit

$$K = \sum_{k \geq 1} 10^{-k^2} P_k = 0,200300005\dots.$$

Alors  $10^{n^2} K \equiv P_n + \text{une fraction } (\bmod 10^{2n-1})$ .

**4.22**  $(b^{mn} - 1)/(b - 1) = ((b^m - 1)/(b - 1))(b^{m(n-m)} + \dots + 1)$ . (Les seuls nombres premiers de la forme  $(10^p - 1)/9$  avec  $p < 20000$  sont ceux pour lesquels  $p = 2, 19, 23, 317$ , ou  $1031$ ). Les nombres de cette forme sont appelés “repunits” en anglais.

**4.23**  $\rho(2k+1) = 0$  ;  $\rho(2k) = \rho(k) + 1$ , pour  $k \geq 1$ . On peut montrer par induction que  $\rho(n) = \rho(n - 2^m)$ , si  $n > 2^m$  et  $m > \rho(n)$ . Si on numérote les disques de la tour de  $0$  à  $n-1$ , le  $k$ ième mouvement concerne le disque numéro  $\rho(k)$ . C'est facile à voir si  $k$  est une puissance de  $2$ . Si  $2^m < k < 2^{m+1}$ , on a  $\rho(k) < m$  ; les mouvements  $k$  et  $k - 2^m$  sont en correspondance dans la suite de mouvements qui déplace  $m+1$  disques en  $T_m + 1 + T_m$  étapes.

**4.24** Si un chiffre  $d$  compte pour  $dp^m$  dans  $n$ , alors il compte pour  $dp^{m-1} + \dots + d = d(p^m - 1)/(p - 1)$  dans  $\epsilon_p(n!)$ . Donc  $\epsilon_p(n!) = (n - v_p(n))/(p - 1)$ .

**4.25**  $m \setminus\!/ n \iff m_p = 0$  ou  $m_p = n_p$ , pour tout  $p$ . Par conséquent, (a) est vrai. Par contre, (b) est faux pour  $m = 12$  et  $n = 18$ , notre exemple favori (c'est une erreur assez courante que de croire le contraire).

**4.26** Oui, car  $\mathcal{G}_N$  définit un sous-arbre de l'arbre de Stern-Brocot.

**4.27** Prolonger la plus courte des deux suites avec des  $M$  (parce que  $M$  est entre  $L$  et  $R$  dans l'alphabet) jusqu'à ce qu'elle atteigne la longueur de l'autre suite ; puis les comparer selon l'ordre alphabétique. Ainsi, pour les trois premiers niveaux de l'arbre, on a  $LL < LM < LR < MM < RL < RM < RR$ . Une autre solution possible consiste à ajouter la suite infinie  $RL^\infty$  aux deux suites, puis les comparer lettre à lettre jusqu'à ce qu'on trouve  $L < R$ .

**4.28** Nous n'avons besoin que de la première partie de la représentation :

$$\begin{array}{ccccccccccccccccc} R & R & R & L & L & L & L & L & L & R & R & R & R & R & R \\ \frac{1}{1}, \frac{2}{1}, \frac{3}{1}, \frac{4}{1}, \frac{7}{2}, \frac{10}{3}, \frac{13}{4}, \frac{16}{5}, \frac{19}{6}, \frac{22}{7}, \frac{25}{8}, \frac{47}{15}, \frac{69}{22}, \frac{91}{29}, \frac{113}{36}, \frac{135}{43}, \dots \end{array}$$

La fraction  $\frac{4}{1}$  apparaît, non parce qu'elle est plus proche de  $\pi$  que  $\frac{3}{1}$ , mais parce que c'est un meilleur majorant que  $\frac{1}{0}$ . Similairement,  $\frac{25}{8}$  est un meilleur minorant que  $\frac{3}{1}$ . Tous les majorants et les minorants les plus simples apparaissent ; remarquons cependant que la fraction qui améliore vraiment une approximation donnée n'apparaît que juste avant le prochain L.

**4.29**  $1/\alpha$ . Pour obtenir  $1-x$  à partir de la représentation binaire de  $x$ , on échange les 0 et les 1 ; pour obtenir  $1/\alpha$  à partir de la représentation de Stern-Brocot de  $\alpha$ , on échange les L et les R. (Pour être complet, il faut aussi considérer les cas finis ; cela marche aussi, car la correspondance préserve l'ordre).

**4.30** Les  $m$  entiers  $x \in [A \dots A+m)$  sont tous distincts mod  $m$ , donc leurs résidus  $(x \bmod m_1, \dots, x \bmod m_r)$  prennent toutes les valeurs  $m_1 \dots m_r = m$  possibles. D'après le principe des boîtes, l'un de ces  $r$ -uplets est égal à  $(a_1 \bmod m_1, \dots, a_r \bmod m_r)$ .

**4.31** Si  $b \equiv 1 \pmod d$ , un nombre en base  $b$  est divisible par  $d$  si et seulement si la somme de ses chiffres est divisible par  $d$ . Ceci est dû au fait que  $(a_m \dots a_0)_b = a_m b^m + \dots + a_0 b^0 \equiv a_m + \dots + a_0$ .

**4.32** L'ensemble des  $\varphi(m)$  nombres  $\{kn \bmod m \mid k \perp m \text{ et } 0 \leq k < m\}$  est égal à l'ensemble des nombres  $\{k \mid k \perp m \text{ et } 0 \leq k < m\}$ . Faites leur produit puis divisez le tout par  $\prod_{0 \leq k < m, k \perp m} k$ .

**4.33** Bien sûr,  $h(1) = 1$ . Si  $m \perp n$ ,  $h(mn) = \sum_{d \mid mn} f(d) g(mn/d) = \sum_{c \mid m, d \mid n} f(cd) g((m/c)(n/d)) = \sum_{c \mid m} \sum_{d \mid n} f(c) g(m/c) f(d) g(n/d)$  ; c'est égal à  $h(m) h(n)$ , car  $c \perp d$  pour chacun des termes de la somme.

**4.34**  $g(m) = \sum_{d \mid m} f(d) = \sum_{d \mid m} f(m/d) = \sum_{d \geq 1} f(m/d)$  si  $f(x)$  est nul lorsque  $x$  n'est pas un entier.

**4.35** Les bases sont

$$I(0, n) = 0; \quad I(m, 0) = 1.$$

Lorsque  $m, n > 0$ , il y a deux règles. La première est triviale si  $m > n$ , la seconde l'est si  $m < n$  :

$$\begin{aligned} I(m, n) &= I(m, n \bmod m) - \lfloor n/m \rfloor I(n \bmod m, m); \\ I(m, n) &= I(m \bmod n, n). \end{aligned}$$

**4.36** Toute factorisation de chacun des nombres donnés en nombres qui ne sont pas des unités doit satisfaire  $m^2 - 10n^2 = \pm 2$  ou  $\pm 3$ . C'est impossible mod 10.

**4.37** Soient  $a_n = 2^{-n} \ln(e_n - \frac{1}{2})$  et  $b_n = 2^{-n} \ln(e_n + \frac{1}{2})$ . Alors

$$e_n = \lfloor E^{2^n} + \frac{1}{2} \rfloor \iff a_n \leq \ln E < b_n.$$

Comme  $a_{n-1} < a_n < b_n < b_{n-1}$ , on peut prendre  $E = \lim_{n \rightarrow \infty} e^{a_n}$ . Il s'avère que

$$E^2 = \frac{3}{2} \prod_{n \geq 1} \left( 1 + \frac{1}{(2e_n - 1)^2} \right)^{1/2^n},$$

ce qui converge rapidement vers  $(1,26408473530530111\dots)^2$ . Toutefois, ces résultats ne nous apprennent rien sur  $e_n$ . Il faudrait pour cela trouver une expression de  $E$  ne dépendant pas des nombres d'Euclide.

**4.38** Si  $r = n \bmod m$ , alors  $a^n - b^n = (a^m - b^m)(a^{n-m}b^0 + a^{n-2m}b^m + \dots + a^r b^{n-m-r}) + b^{m\lfloor n/m \rfloor}(a^r - b^r)$ .

**4.39** Si  $a_1 \dots a_t$  et  $b_1 \dots b_u$  sont des carrés parfaits, le nombre

$$a_1 \dots a_t b_1 \dots b_u / c_1^2 \dots c_v^2,$$

où  $\{a_1, \dots, a_t\} \cap \{b_1, \dots, b_u\} = \{c_1, \dots, c_v\}$ , en est un aussi. (On peut montrer que la suite  $\langle S(1), S(2), S(3), \dots \rangle$  contient chaque entier strictement positif non premier exactement une fois).

**4.40** Soient

$$f(n) = \prod_{1 \leq k \leq n, p \nmid k} k = n! / p^{\lfloor n/p \rfloor} \lfloor n/p \rfloor!$$

et

$$g(n) = n! / p^{\epsilon_p(n!)},$$

Alors

$$g(n) = f(n)f(\lfloor n/p \rfloor)f(\lfloor n/p^2 \rfloor) \dots = f(n)g(\lfloor n/p \rfloor).$$

On a aussi  $f(n) \equiv a_0!(p-1)!^{\lfloor n/p \rfloor} \equiv a_0!(-1)^{\lfloor n/p \rfloor} \pmod{p}$ , et  $\epsilon_p(n!) = \lfloor n/p \rfloor + \epsilon_p(\lfloor n/p \rfloor!)$ . Avec ces récurrences, on arrive facilement au résultat par induction. (Il y a beaucoup d'autres solutions possibles).

**4.41** (a) Si  $n^2 \equiv -1 \pmod{p}$ , alors  $(n^2)^{(p-1)/2} \equiv -1$ ; or, d'après Fermat, ce devrait être  $+1$ . (b) Soit  $n = ((p-1)/2)!$ ; nous avons  $n \equiv (-1)^{(p-1)/2} \prod_{1 \leq k < p/2} (p-k) = (p-1)!/n$ , donc  $n^2 \equiv (p-1)!$ .

**4.42** Remarquons tout d'abord que  $k \perp l \iff k \perp l + ak$  pour tout entier  $a$ , car, de par l'algorithme d'Euclide,  $\text{pgcd}(k, l) = \text{pgcd}(k, l + ak)$ . Alors

$$\begin{aligned} m \perp n \text{ et } n' \perp n &\iff mn' \perp n \\ &\iff mn' + nm' \perp n. \end{aligned}$$

De façon similaire, on a

$$m' \perp n' \text{ et } n \perp n' \iff mn' + nm' \perp n'.$$

Par conséquent,

$$m \perp n \text{ et } m' \perp n' \text{ et } n \perp n' \iff mn' + nm' \perp nn'.$$

**4.43** On veut multiplier par  $L^{-1}R$ , puis par  $R^{-1}L^{-1}RL$ , puis  $L^{-1}R$ , puis  $R^{-2}L^{-1}RL^2$  etc ; le  $n$ ième facteur est  $R^{-\rho(n)}L^{-1}RL^{\rho(n)}$ , car il doit permettre de supprimer  $\rho(n)$  lettres R. Or,  $R^{-m}L^{-1}RL^m = \begin{pmatrix} 0 & -1 \\ 1 & 2m+1 \end{pmatrix}$ .

**4.44** On peut trouver le nombre rationnel le plus simple de l'intervalle

$$[0,3155\dots 0,3165) = [\frac{631}{2000} \dots \frac{633}{2000})$$

en écrivant les représentations de Stern–Brocot de  $\frac{631}{2000}$  et  $\frac{633}{2000}$  et en stoppant à l'étape précédent celle où on écrirait L pour la première et R pour la seconde :

```
(m1, n1, m2, n2) := (631, 2000, 633, 2000);
tant que m1 > n1 ou m2 < n2 faire
  si m2 < n2 alors (écrire(L); (n1, n2) := (n1, n2) - (m1, m2))
  sinon (écrire(R); (m1, m2) := (m1, m2) - (n1, n2))
```

Cet algorithme écrit LLLRRRRR =  $\frac{6}{19} \approx 0,3158$ . On voit aussi qu'il faut frapper la balle au moins 287 fois pour obtenir une moyenne de 0,334 .

**4.45**  $x^2 \equiv x \pmod{10^n} \iff x(x-1) \equiv 0 \pmod{2^n}$  et  $x(x-1) \equiv 0 \pmod{5^n} \iff x \bmod{2^n} = [0 \text{ ou } 1]$  et  $x \bmod{5^n} = [0 \text{ ou } 1]$ . (La dernière étape provient du fait que  $x(x-1) \bmod{5} = 0$  implique que soit  $x$ , soit  $x-1$  est un multiple de 5, auquel cas l'autre facteur est premier par rapport à  $5^n$ ).

Il existe donc au plus quatre solutions. Deux d'entre elles ( $x = 0$  et  $x = 1$ ) ne peuvent prétendre au titre de "nombre à  $n$  chiffres", sauf si  $n = 1$ . Les deux autres sont respectivement de la forme  $x$  et  $10^n + 1 - x$ , et au moins un de ces nombres est  $\geq 10^{n-1}$ . Lorsque  $n = 4$ , l'autre solution,  $10001 - 9376 = 625$ , n'a pas quatre chiffres. On conjecture qu'il existe deux solutions pour environ 90% des cas, mais cela n'a pas été prouvé.

(De tels nombres, qui s'auto-reproduisent, sont dits automorphes).

*John 0,316*

— banderole vue pendant les championnats du monde 1993, au moment où John Kruk s'apprétait à frapper la balle.

**4.46** (a) Si  $j'j - k'k = \text{pgcd}(j, k)$ , on a  $n^{k'k}n^{\text{pgcd}(j, k)} = n^{j'j} \equiv 1$  et  $n^{k'k} \equiv 1$ . (b) Soit  $n = pq$ , où  $p$  est le plus petit diviseur premier de  $n$ . Si  $2^n \equiv 1 \pmod{n}$ , alors  $2^n \equiv 1 \pmod{p}$ . De même,  $2^{p-1} \equiv 1 \pmod{p}$ ; donc  $2^{\text{pgcd}(p-1, n)} \equiv 1 \pmod{p}$ . Or,  $\text{pgcd}(p-1, n) = 1$  d'après la définition de  $p$ .

**4.47** Si  $n^{m-1} \equiv 1 \pmod{m}$ , alors  $n \perp m$ . S'il existe  $1 \leq j < k < m$  tels que  $n^k \equiv n^j$ , alors on obtient  $n^{k-j} \equiv 1$  en divisant par  $n^j$ . Ainsi, si les nombres  $n^1 \pmod{m}, \dots, n^{m-1} \pmod{m}$  ne sont pas distincts, il existe un  $k < m-1$  tel que  $n^k \equiv 1$ . D'après l'exercice 46(a), le plus petit de ces  $k$  doit diviser  $m-1$ ; il existe alors un nombre premier  $p$  et un entier positif  $q$  tels que  $kq = (m-1)/p$ . Or c'est impossible car  $n^{kq} \not\equiv 1$ . Par conséquent, les nombres  $n^1 \pmod{m}, \dots, n^{m-1} \pmod{m}$  sont distincts et premiers avec  $m$ . Donc les nombres  $1, \dots, m-1$  sont premiers avec  $m$ , ce qui entraîne que  $m$  est premier.

**4.48** En appariant chaque nombre avec son inverse, réécrivons le produit  $(\pmod{m})$  en  $\prod_{1 \leq n < m, n^2 \pmod{m}=1} n$ . Nous pouvons alors faire appel à notre savoir sur les solutions de  $n^2 \pmod{m} = 1$ . En utilisant les résidus, on trouve que le résultat est  $m-1$  si  $m = 4, p^k$  ou  $2p^k$  ( $p > 2$ ); sinon, le résultat est  $+1$ .

**4.49** (a) Il y a trois cas possibles :  $m < n$  ( $\Phi(N)-1$  occurrences),  $m = n$  (une occurrence) ou  $m > n$  (encore  $\Phi(N)-1$  occurrences). Donc  $R(N) = 2\Phi(N)-1$ . (b) D'après (4.62), on obtient

$$2\Phi(N)-1 = -1 + \sum_{d \geq 1} \mu(d)[N/d][1+N/d];$$

donc le résultat attendu est vrai si et seulement si

$$\sum_{d \geq 1} \mu(d)[N/d] = 1, \quad \text{pour } N \geq 1.$$

En posant  $f(x) = [x \geq 1]$ , on se ramène à un cas particulier de (4.61).

**4.50** (a) Pour toute fonction  $f$ , on a

$$\begin{aligned} \sum_{0 \leq k < m} f(k) &= \sum_{d \mid m} \sum_{0 \leq k < m} f(k)[d = \text{pgcd}(k, m)], \\ &= \sum_{d \mid m} \sum_{0 \leq k < m} f(k)[k/d \perp m/d] \\ &= \sum_{d \mid m} \sum_{0 \leq k < m/d} f(kd)[k \perp m/d] \\ &= \sum_{d \mid m} \sum_{0 \leq k < d} f(km/d)[k \perp d]; \end{aligned}$$

nous avons déjà vu un cas particulier de ceci dans le calcul de (4.63). Il existe aussi un calcul analogue si on remplace  $\sum$  par  $\prod$ . On a donc

$$z^m - 1 = \prod_{0 \leq k < m} (z - \omega^k) = \prod_{d \mid m} \prod_{\substack{0 \leq k < d \\ k \perp d}} (z - \omega^{km/d}) = \prod_{d \mid m} \Psi_d(z)$$

du fait que  $\omega^{m/d} = e^{2\pi i/d}$ .

Le point (b) découle du point (a) en prenant l'analogue de (4.56) pour des produits au lieu de sommes. Notons en passant que cette formule indique que  $\Psi_m(z)$  a des coefficients entiers, puisqu'on obtient  $\Psi_m(z)$  en multipliant et divisant des polynômes dont le coefficient de plus haut degré est 1.

**4.51**  $(x_1 + \dots + x_n)^p = \sum_{k_1 + \dots + k_n = p} p! / (k_1! \dots k_n!) x_1^{k_1} \dots x_n^{k_n}$ , et le coefficient est divisible par  $p$ , sauf s'il existe un  $k_j = p$ . Donc  $(x_1 + \dots + x_n)^p \equiv x_1^p + \dots + x_n^p \pmod{p}$ . En prenant tous les  $x$  égaux à 1, on trouve  $n^p \equiv n$ .

**4.52** Si  $p > n$ , c'est trivial. Dans le cas contraire, on a  $x \perp p$ , donc  $x^{k(p-1)} \equiv 1 \pmod{p}$ . Cela signifie qu'il y a au moins  $\lfloor (n-1)/(p-1) \rfloor$  multiples de  $p$  parmi les nombres donnés. Comme  $n \geq p$ , on a  $(n-1)/(p-1) \geq n/p$ .

**4.53** Montrez d'abord que si  $m \geq 6$  et  $m$  n'est pas premier, alors  $(m-2)! \equiv 0 \pmod{m}$ . (Si  $m = p^2$ , le produit  $(m-2)!$  contient les facteurs  $p$  et  $2p$ ; sinon, il contient  $d$  et  $m/d$  tels que  $d < m/d$ ). Puis considérez les cas suivants :

Cas 0,  $n < 5$ . La condition n'est vraie que pour  $n = 1$ .

Cas 1,  $n \geq 5$  et  $n$  est premier. Alors  $(n-1)!/(n+1)$  est un entier et ne peut pas être un multiple de  $n$ .

Cas 2,  $n \geq 5$ ,  $n$  est composite ainsi que  $n+1$ . Alors  $n$  et  $n+1$  divisent  $(n-1)!$  et  $n \perp n+1$ ; donc  $n(n+1) \mid (n-1)!$ .

Cas 3,  $n \geq 5$ ,  $n$  est composite et  $n+1$  est premier. Alors  $(n-1)! \equiv 1 \pmod{n+1}$  d'après le théorème de Wilson, et

$$\lceil (n-1)!/(n+1) \rceil = ((n-1)! + n)/(n+1);$$

c'est divisible par  $n$ .

La réponse est donc : soit  $n = 1$ , soit  $n \neq 4$  et  $n$  est composite.

**4.54** Comme  $\epsilon_2(1000!) > 500$  et  $\epsilon_5(1000!) = 249$ , il existe un entier pair  $a$  tel que  $1000! = a \cdot 10^{249}$ ; et comme  $1000 = (1300)_5$ , l'exercice 40 nous dit que  $a \cdot 2^{249} = 1000!/5^{249} \equiv -1 \pmod{5}$ . On a aussi  $2^{249} \equiv 2$ , donc  $a \equiv 2$ , donc  $a \bmod 10 = 2$  ou 7. Par conséquent, la réponse est  $2 \cdot 10^{249}$ .

**4.55** On peut démontrer par induction que  $P_{2n}/P_n^4(n+1)$  est un entier. Ce résultat plus fort permet à l'induction d'aboutir. On peut aussi opérer

*Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.*

—L. Kronecker [365]

différemment, en montrant que tout nombre premier  $p$  divise le numérateur au moins autant de fois qu'il divise le dénominateur. Cela revient à prouver l'inégalité

$$\sum_{k=1}^{2n} \lfloor k/m \rfloor \geq 4 \sum_{k=1}^n \lfloor k/m \rfloor$$

qui se déduit de

$$\lfloor (2n-1)/m \rfloor + \lfloor 2n/m \rfloor \geq \lfloor n/m \rfloor.$$

Cette dernière est vraie lorsque  $0 \leq n < m$ , et chacun des deux membres augmente de 4 si  $n$  augmente de  $m$ .

**4.56** Soient les fonctions  $f(m) = \sum_{k=1}^{2n-1} \min(k, 2n-k)[m \setminus k]$  et  $g(m) = \sum_{k=1}^{n-1} (2n-2k-1)[m \setminus (2k+1)]$ . Le nombre de fois que  $p$  divise le numérateur du produit donné est égal à  $f(p) + f(p^2) + f(p^3) + \dots$ , et le nombre de fois que  $p$  divise le dénominateur est égal à  $g(p) + g(p^2) + g(p^3) + \dots$ . Or, d'après l'exercice 2.32,  $f(m) = g(m)$  pour tout  $m$  impair. Par conséquent, d'après l'exercice 3.22, le produit considéré est égal à  $2^{n(n-1)}$ .

**4.57** D'après la suggestion, on peut effectuer un changement de sommation standard, car

$$\sum_{1 \leq m \leq n} [d \setminus m] = \sum_{0 < k \leq n/d} [m = dk] = \lfloor n/d \rfloor.$$

Si on note  $\Sigma(n)$  la somme de la suggestion, on a

$$\Sigma(m+n) - \Sigma(m) - \Sigma(n) = \sum_{d \in S(m,n)} \varphi(d).$$

D'autre part, on sait que, d'après (4.54),  $\Sigma(n) = \frac{1}{2}n(n+1)$ . Donc  $\Sigma(m+n) - \Sigma(m) - \Sigma(n) = mn$ .

**4.58** La fonction  $f(m)$  est multiplicative et vaut  $1+p+\dots+p^k$  si  $m=p^k$ . C'est une puissance de 2 si et seulement si  $p$  est un nombre de Mersenne premier et  $k=1$ . Donc  $k$  doit être impair, et dans ce cas la somme vaut

$$(1+p)(1+p^2+p^4+\dots+p^{k-1}),$$

alors  $(k-1)/2$  doit être impair etc. Voici donc la condition nécessaire et suffisante :  $m$  est un produit de nombres de Mersenne premiers distincts.

**4.59** Preuve de la suggestion : si  $n=1$ , on a  $x_1=\alpha=2$ , donc pas de problème. Si  $n>1$ , on peut supposer que  $x_1 \leq \dots \leq x_n$ . Cas 1 :  $x_1^{-1}+\dots+x_{n-1}^{-1}+(x_n-1)^{-1} \geq 1$  et  $x_n > x_{n-1}$ . Alors on peut trouver

$\beta \geq x_n - 1 \geq x_{n-1}$  tel que  $x_1^{-1} + \cdots + x_{n-1}^{-1} + \beta^{-1} = 1$ . Donc  $x_n \leq \beta + 1 \leq e_n$  et  $x_1 \dots x_n \leq x_1 \dots x_{n-1}(\beta + 1) \leq e_1 \dots e_n$ , par induction. Il existe un entier positif  $m$  tel que  $\alpha = x_1 \dots x_n/m$ ; donc  $\alpha \leq e_1 \dots e_n = e_{n+1} - 1$ , et on a  $x_1 \dots x_n(\alpha + 1) \leq e_1 \dots e_n e_{n+1}$ . Cas 2 :  $x_1^{-1} + \cdots + x_{n-1}^{-1} + (x_n - 1)^{-1} \geq 1$  et  $x_n = x_{n-1}$ . Soient  $a = x_n$  et  $a^{-1} + (a-1)^{-1} = (a-2)^{-1} + \zeta^{-1}$ . On peut alors montrer que  $a \geq 4$  et  $(a-2)(\zeta + 1) \geq a^2$ . Il existe donc un  $\beta \geq \zeta$  tel que  $x_1^{-1} + \cdots + x_{n-2}^{-1} + (a-2)^{-1} + \beta^{-1} = 1$ . Il s'ensuit, par induction, que  $x_1 \dots x_n \leq x_1 \dots x_{n-2}(a-2)(\zeta + 1) \leq x_1 \dots x_{n-2}(a-2)(\beta + 1) \leq e_1 \dots e_n$ , et on peut terminer comme dans le cas précédent. Cas 3 :  $x_1^{-1} + \cdots + x_{n-1}^{-1} + (x_n - 1)^{-1} < 1$ . Soient  $a = x_n$  et  $a^{-1} + \alpha^{-1} = (a-1)^{-1} + \beta^{-1}$ . On peut montrer que  $(a-1)(\beta + 1) > a(\alpha + 1)$ , car cette identité est équivalente à  $a\alpha^2 - a^2\alpha + a\alpha - a^2 + \alpha + a > 0$ , qui est une conséquence du fait que  $a\alpha(a-a) + (1+a)\alpha \geq (1+a)\alpha > a^2 - a$ . On peut donc remplacer  $x_n$  et  $\alpha$  par  $a-1$  et  $\beta$  et répéter cette transformation jusqu'à se retrouver dans le cas 1 ou le cas 2.

Voici une autre conséquence de la suggestion : si  $1/x_1 + \cdots + 1/x_n < 1$ , alors  $1/x_1 + \cdots + 1/x_n \leq 1/e_1 + \cdots + 1/e_n$  (voir l'exercice 16).

**4.60** Le point important est que  $\theta < \frac{2}{3}$ . Alors on peut prendre  $p_1$  assez grand (pour que les conditions du fait donné soient respectées) et  $p_n$  comme étant le plus petit nombre premier supérieur à  $p_{n-1}^3$ . Soient alors  $a_n = 3^{-n} \ln p_n$  et  $b_n = 3^{-n} \ln(p_n + 1)$ . Si on sait prouver que  $a_{n-1} \leq a_n < b_n \leq b_{n-1}$ , on peut prendre  $P = \lim_{n \rightarrow \infty} e^{a_n}$ , tout comme dans l'exercice 37. Or, cette hypothèse est équivalente à  $p_{n-1}^3 \leq p_n < (p_{n-1} + 1)^3$ . S'il n'existe pas de nombre premier  $p_n$  dans cet intervalle, il existe nécessairement un nombre premier  $p < p_{n-1}^3$  tel que  $p + cp^\theta > (p_{n-1} + 1)^3$ . Cela implique que  $cp^\theta > 3p^{2/3}$ , ce qui est impossible si  $p$  est suffisamment grand.

On peut presque à coup sûr prendre  $p_1 = 2$ , car tout indique que les bornes que l'on connaît sur les tailles des intervalles entre les nombres premiers sont bien plus faibles que la réalité (voir l'exercice 69). Alors  $p_2 = 11$ ,  $p_3 = 1361$ ,  $p_4 = 2521008887$  et  $1,306377883863 < P < 1,306377883869$ .

*"Man made  
the integers:  
All else is  
Dieudonné."*

— R. K. Guy

**4.61** Soient  $\hat{m}$  et  $\hat{n}$  les membres droits. Remarquez que  $\hat{m}\hat{n}' - m'\hat{n} = 1$ , donc que  $\hat{m} \perp \hat{n}$ . On a aussi  $\hat{m}/\hat{n} > m'/n'$  et  $N = ((n+N)/n')n' - n \geq \hat{n} > ((n+N)/n' - 1)n' - n = N - n' \geq 0$ . Par conséquent,  $\hat{m}/\hat{n} \geq m''/n''$ . Si on suppose que l'égalité est fausse, on a  $n'' = (\hat{m}\hat{n}' - m'\hat{n})n'' = \hat{n}'(\hat{m}\hat{n}'' - m''\hat{n}) + \hat{n}(m''n' - m'n'') \geq n' + \hat{n} > N$ , ce qui est absurde.

Notons en passant que cet exercice implique que  $(m+m'')/(n+n'') = m'/n'$ , bien que la première fraction ne soit pas toujours réduite.

**4.62**  $2^{-1} + 2^{-2} + 2^{-3} - 2^{-6} - 2^{-7} + 2^{-12} + 2^{-13} - 2^{-20} - 2^{-21} + 2^{-30} + 2^{-31} - 2^{-42} - 2^{-43} + \cdots$  peut s'écrire

$$\frac{1}{2} + 3 \sum_{k \geq 0} (2^{-4k^2-6k-3} - 2^{-4k^2-10k-7}).$$

En utilisant la “fonction theta”,  $\theta(z, \lambda) = \sum_k e^{-\pi \lambda k^2 + 2izk}$ , on peut en donner une formule close :

$$e \leftrightarrow \frac{1}{2} + \frac{3}{8}\theta\left(\frac{4}{\pi} \ln 2, 3i \ln 2\right) - \frac{3}{128}\theta\left(\frac{4}{\pi} \ln 2, 5i \ln 2\right).$$

**4.63** Tout entier  $n > 2$  admet un diviseur premier  $d$ , ou bien est divisible par  $d = 4$ . Dans les deux cas, s'il existe une solution avec l'exposant  $n$ , alors il existe une solution  $(a^{n/d})^d + (b^{n/d})^d = (c^{n/d})^d$  avec l'exposant  $d$ . Comme il n'y a pas de solution avec  $d = 4$ ,  $d$  est forcément premier.

La suggestion découle de la formule du binôme, car  $(a^p + (x-a)^p)/x \equiv p a^{p-1} \pmod{x}$  lorsque  $p$  est impair. Le plus petit contre-exemple, si (4.46) est faux, est tel que  $a \perp x$ . Si  $x$  n'est pas divisible par  $p$ , alors  $x$  est premier avec  $c^p/x$ . Cela signifie que, pour tout  $q$  premier tel que  $q^e \nmid x$  et  $q^f \mid c$ , on a  $e = fp$ . Par conséquent, il existe  $m$  tel que  $x = m^p$ . D'autre part, si  $x$  est divisible par  $p$ , alors  $c^p/x$  est divisible par  $p$  mais pas par  $p^2$ , et  $c^p$  n'a pas d'autre facteur commun avec  $x$ .

**4.64** Les fractions égales de  $\mathcal{P}_N$  sont en “ordre de tuyaux d'orgue” :

$$\frac{2m}{2n}, \frac{4m}{4n}, \dots, \frac{rm}{rn}, \dots, \frac{3m}{3n}, \frac{m}{n}.$$

Supposons que  $\mathcal{P}_N$  est correct et montrons que  $\mathcal{P}_{N+1}$  l'est aussi. Pour cela, nous devons prouver que, si  $kN$  est impair, alors

$$\frac{k-1}{N+1} = \mathcal{P}_{N,kN}$$

et si  $kN$  est pair, alors

$$\mathcal{P}_{N,kN-1} \mathcal{P}_{N,kN} \frac{k-1}{N+1} \mathcal{P}_{N,kN} \mathcal{P}_{N,kN+1}.$$

Dans les deux cas, il nous sera utile de connaître le nombre de fractions qui sont strictement plus petites que  $(k-1)/(N+1)$  dans  $\mathcal{P}_N$  :

$$\begin{aligned} \sum_{n=1}^N \sum_m \left[ 0 \leq \frac{m}{n} < \frac{k-1}{N+1} \right] &= \sum_{n=1}^N \left\lceil \frac{(k-1)n}{N+1} \right\rceil = \sum_{n'=0}^N \left\lfloor \frac{(k-1)n+N}{N+1} \right\rfloor \\ &= \frac{(k-2)N}{2} + \frac{d-1}{2} + d \left\lfloor \frac{N}{d} \right\rfloor \end{aligned}$$

d'après (3.32), avec  $d = \text{pgcd}(k-1, N+1)$ . Comme  $N \bmod d = d-1$ , on trouve finalement  $\frac{1}{2}(kN - d + 1)$ .

De plus, en raison de la structure de l'ordre de tuyaux d'orgue, le nombre de fractions égales à  $(k-1)/(N+1)$  dans  $\mathcal{P}_N$  qui devraient la précéder dans  $\mathcal{P}_{N+1}$  est  $\frac{1}{2}(d-1 - [d \text{ pair}])$ .

*J'ai découvert une merveilleuse preuve du Dernier Théorème de Fermat, mais je n'ai pas la place de l'écrire ici.*

Si  $kN$  est impair, alors  $d$  est pair et  $(k-1)/(N+1)$  est précédé de  $\frac{1}{2}(kN-1)$  éléments de  $\mathcal{P}_N$ . C'est exactement le nombre qu'il faut pour que cela marche. Si  $kN$  est pair, alors  $d$  est impair et  $(k-1)/(N+1)$  est précédé de  $\frac{1}{2}(kN)$  éléments de  $\mathcal{P}_N$ . Si  $d=1$ , aucun d'eux n'est égal à  $(k-1)/(N+1)$  et  $\mathcal{P}_{N,kN}$  désigne le symbole " $<$ " ; sinon,  $(k-1)/(N+1)$  s'insère entre deux éléments égaux et  $\mathcal{P}_{N,kN}$  désigne " $=$ ". (C. S. Peirce [288] a découvert indépendamment l'arbre de Stern-Brocot, à peu près en même temps que  $\mathcal{P}_N$ ).

**4.65** La question analogue pour les nombres (analogues) de Fermat  $f_n$  est un problème fameux non résolu. Celui-ci peut être plus facile ou plus difficile.

*"No square less than  
25 × 10<sup>14</sup> divides a  
Euclid number."*

— Ilan Vard

**4.66** On sait qu'aucun carré inférieur à  $36 \times 10^{18}$  ne divise un nombre de Mersenne ou un nombre de Fermat. Jusqu'à présent cependant, il n'existe aucune preuve de la conjecture de Schinzel selon laquelle il existe une infinité de nombres de Mersenne sans carré. On ne sait même pas s'il existe une infinité de nombres  $p$  tels que  $p \nmid (a \pm b)$ , où tous les facteurs premiers de  $a$  et  $b$  sont  $\leq 31$ .

**4.67** M. Szegedy a prouvé cette conjecture pour tout  $n$  assez grand. Voir [348], [95, pp. 78–79] et [55].

**4.68** C'est une conjecture bien plus faible que celle de l'exercice suivant.

**4.69** Cramér [66] a montré que cette conjecture est plausible en termes probabilistes, ce qui est confirmé par le calcul expérimental : Brent [37] a montré que  $P_{n+1} - P_n \leq 602$  pour  $P_{n+1} < 2,686 \times 10^{12}$ . Cependant, le résultat bien plus faible de l'exercice 60 est le meilleur connu en 1994 [255]. La réponse à l'exercice 68 est "oui" si  $P_{n+1} - P_n < 2P_n^{1/2}$  pour tout entier  $n$  suffisamment grand. Selon Guy [169, problème A8], Paul Erdős offrait 10000 dollars à qui trouverait la preuve qu'il existe une infinité de nombres  $n$  tels que

*Paul Erdős est  
décédé en 1996  
(N.d.T.).*

$$P_{n+1} - P_n > \frac{c \ln n \ln \ln n \ln \ln \ln \ln n}{(\ln \ln \ln n)^2}$$

pour tout  $c > 0$ .

**4.70** Selon l'exercice 24, cela est vrai si et seulement si  $\nu_2(n) = \nu_3(n)$ . Les méthodes de [96] pourraient aider à faire tomber cette conjecture.

**4.71** Si  $k=3$ , la plus petite solution est  $n = 4700063497 = 19 \cdot 47 \cdot 5263229$  ; on n'en connaît pas d'autre.

**4.72** On sait que c'est vrai pour une infinité d'entiers  $a$ , dont  $-1$  (évidemment) et  $0$  (bien moins évidemment). Lehmer [244] a formulé la fameuse conjecture suivante :  $\varphi(n) \mid (n-1)$  si et seulement si  $n$  est premier.

**4.73** On sait que cette question est équivalente à l'hypothèse de Riemann (selon laquelle la fonction complexe zêta  $\zeta(z)$  est non nulle lorsque la partie réelle de  $z$  est supérieure à  $1/2$ ).

**4.74** D'après les résultats expérimentaux, il semble qu'il y a à peu près  $p(1 - 1/e)$  éléments distincts, exactement comme si les factorielles étaient distribuées au hasard modulo  $p$ .

Combien vaut  $11^4$  en base 11 ?

**5.1**  $(11)_r^4 = (14641)_r$ , dans tout système de numération à base  $r \geq 7$ , en vertu de la formule du binôme.

**5.2** Le rapport  $\binom{n}{k+1}/\binom{n}{k} = (n-k)/(k+1)$  est  $\leq 1$  lorsque  $k \geq \lfloor n/2 \rfloor$  et  $\geq 1$  lorsque  $k < \lceil n/2 \rceil$ , donc le maximum est atteint lorsque  $k = \lfloor n/2 \rfloor$  et  $k = \lceil n/2 \rceil$ .

**5.3** Développez en factorielles. Les deux produits valent  $f(n)/f(n-k)f(k)$ , où  $f(n) = (n+1)!n!(n-1)!$ .

**5.4**  $\binom{-1}{k} = (-1)^k \binom{k+1-1}{k} = (-1)^k \binom{k}{k} = (-1)^k [k \geq 0]$ .

**5.5** Si  $0 < k < p$ , il existe un  $p$  dans le numérateur de  $\binom{p}{k}$  qui ne peut pas se simplifier avec un facteur du dénominateur. Comme  $\binom{p}{k} = \binom{p-1}{k} + \binom{p-1}{k-1}$ , on a forcément  $\binom{p-1}{k} \equiv (-1)^k \pmod{p}$  pour tout  $0 \leq k < p$ .

**5.6** Voici ce qu'il aurait fallu faire après avoir éliminé les deux premiers  $k$ :

$$\begin{aligned} & \frac{1}{n+1} \sum_k \binom{n+k}{k} \binom{n+1}{k+1} (-1)^k \\ &= \frac{1}{n+1} \sum_{k \geq 0} \binom{n+k}{n} \binom{n+1}{k+1} (-1)^k \\ &= \frac{1}{n+1} \sum_k \binom{n+k}{n} \binom{n+1}{k+1} (-1)^k \\ &\quad - \frac{1}{n+1} \binom{n-1}{n} \binom{n+1}{0} (-1)^{-1}. \end{aligned}$$

Dans le calcul d'origine, on a oublié d'ajouter ce terme, qui vaut  $[n=0]$ .

**5.7** Oui, car  $r^{-k} = (-1)^k / (-r-1)^k$ . On a aussi  $r^{\bar{k}}(r+\frac{1}{2})^{\bar{k}} = (2r)^{\bar{2k}}/2^{2k}$ .

**5.8**  $f(k) = (k/n-1)^n$  est un polynôme de degré  $n$  dont le coefficient directeur est  $n^{-n}$ . D'après (5.40), la somme fait  $n!/n^n$ . Lorsque  $n$  est grand, on sait, d'après la formule d'approximation de Stirling, que cela donne approximativement  $\sqrt{2\pi n}/e^n$ . Remarquez que c'est tout à fait différent de  $(1-1/e)$ , qui est ce qu'on obtiendrait en prenant l'approximation  $(1-k/n)^n \sim e^{-k}$ , valable pour  $k$  fixé lorsque  $n \rightarrow \infty$ .

**5.9**  $\mathcal{E}_t(z)^t = \sum_{k \geq 0} t(tk+t)^{k-1} z^k/k! = \sum_{k \geq 0} (k+1)^{k-1} (tz)^k/k! = \mathcal{E}_1(tz)$ , selon (5.60).

**5.10**  $\sum_{k \geq 0} 2z^k/(k+2) = F(2, 1; 3; z)$ , car  $t_{k+1}/t_k = (k+2)z/(k+3)$ .

**5.11** La première est une fonction de Bessel, tandis que la seconde est une *Mais pas une fonction d'imbécile.* gaussienne.

$$z^{-1} \sin z = \sum_{k \geq 0} (-1)^k z^{2k}/(2k+1)! = F(1; 1, \frac{3}{2}; -z^2/4);$$

$$z^{-1} \arcsin z = \sum_{k \geq 0} z^{2k} (\frac{1}{2})^{\overline{k}}/(2k+1)k! = F(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}; z^2).$$

**5.12** (a) Oui si  $n \neq 0$ , car le rapport des termes vaut  $n$ . (b) Oui si  $n$  est un entier ; le rapport des termes est égal à  $(k+1)^n/k^n$ . Remarquez que ce terme provient de (5.115), en posant  $m = n+1$ ,  $a_1 = \dots = a_m = 1$ ,  $b_1 = \dots = b_n = 0$ ,  $z = 1$ , et en multipliant par  $0^n$ . (c) Oui, car le rapport des termes est  $(k+1)(k+3)/(k+2)$ . (d) Non. Le rapport des termes est égal à  $1 + 1/(k+1)H_k$ , et  $H_k \sim \ln k$  n'est pas une fonction rationnelle. (e) Oui. L'inverse d'un terme hypergéométrique est toujours un terme hypergéométrique. Le fait que  $t(k) = \infty$  lorsque  $k < 0$  ou  $k > n$  n'empêche pas  $t(k)$  d'être un terme hypergéométrique. (f) Bien sûr. (g) Non, par exemple si  $t(k) = 2^k$  et  $T(k) = 1$ . (h) Oui. Pour tout  $n$ , le rapport des termes  $t(n-1-k)/t(n-1-(k+1))$  est une fonction rationnelle (c'est l'inverse du rapport des termes de  $t$ , où  $k$  est remplacé par  $n-1-k$ ). (i) Oui. Le rapport des termes peut s'écrire

$$\frac{a t(k+1)/t(k) + b t(k+2)/t(k) + c t(k+3)/t(k)}{a + b t(k+1)/t(k) + c t(k+2)/t(k)},$$

et  $t(k+m)/t(k) = (t(k+m)/t(k+m-1)) \dots (t(k+1)/t(k))$  est une fonction rationnelle de  $k$ . (j) Non. Si deux fonctions rationnelles  $p_1(k)/q_1(k)$  et  $p_2(k)/q_2(k)$  sont égales pour une infinité de  $k$ , alors elles sont égales pour tout  $k$ , car  $p_1(k)q_2(k) = q_1(k)p_2(k)$  est une identité polynomiale. Par conséquent, le rapport des termes  $[(k+1)/2]/[k/2]$  vaudrait 1 si c'était une fonction rationnelle. (k) Non. Le rapport des termes devrait être égal à  $(k+1)/k$ , puisqu'il vaut  $(k+1)/k$  pour tout  $k > 0$ . Dans ce cas,  $t(-1)$  ne pourrait être nul que si  $t(0)$  était un multiple de  $0^2$ , tandis que  $t(1)$  ne pourrait être égal à 1 que si  $t(0) = 0^1$ .

**5.13**  $R_n = n!^{n+1}/P_n^2 = Q_n/P_n = Q_n^2/n!^{n+1}$ .

**5.14** Le premier facteur de (5.25) est égal à  $\binom{l-k}{l-k-m}$  lorsque  $k \leq l$ , ce qui fait  $(-1)^{l-k-m} \binom{-m-1}{l-k-m}$ . Comme  $m \geq 0$ , la somme pour  $k \leq l$  est en fait la somme pour tout  $k$  (la condition  $n \geq 0$  n'est pas vraiment nécessaire, bien que  $k$  doive prendre des valeurs négatives si  $n < 0$ ). Pour passer de (5.25) à (5.26), commencez par remplacer  $s$  par  $-1-n-q$ .

**5.15** Si  $n$  est impair, la somme est nulle car on peut remplacer  $k$  par  $n-k$ . Si  $n = 2m$ , la somme est égale à  $(-1)^m (3m)!/m!^3$  d'après (5.29), en prenant  $a = b = c = m$ .

Chaque valeur d'un terme hypergéométrique  $t(k)$  peut s'écrire  $0^{e(k)} v(k)$ , où  $e(k)$  est un entier et  $v(k) \neq 0$ . Supposons que le rapport des termes  $t(k+1)/t(k)$  est égal à  $p(k)/q(k)$ , et que  $p$  et  $q$  ont été complètement factorisés sur le corps des nombres complexes. Alors, pour tout  $k$ ,  $e(k+1)$  est égal à  $e(k)$  plus le nombre de facteurs nuls de  $p(k)$  moins le nombre de facteurs nuls de  $q(k)$ , et  $v(k+1)$  est égal à  $v(k)$  fois le produit des facteurs non nuls de  $p(k)$ , divisé par le produit des facteurs non nuls de  $q(k)$ .

**5.16** Cela fait  $(2a)!(2b)!(2c)!/(a+b)!(b+c)!(c+a)!$  fois (5.29), si on développe les binomiaux en factorielles.

**5.17** On déduit de  $\binom{2n-1/2}{n} = \binom{4n}{2n}/2^{2n}$  et de  $\binom{2n-1/2}{2n} = \binom{4n}{2n}/2^{4n}$  que  $\binom{2n-1/2}{n} = 2^{2n} \binom{2n-1/2}{2n}$ .

**5.18**  $\binom{3r}{3k} \binom{3k}{k,k,k} / 3^{3k}$ .

**5.19**  $B_{1-t}(-z)^{-1} = \sum_{k \geq 0} \binom{k-tk-1}{k} (-1/(k-tk-1)) (-z)^k$  selon (5.60), et c'est égal à  $\sum_{k \geq 0} \binom{tk}{k} (1/(tk-k+1)) z^k = B_t(z)$ .

**5.20**  $F(-a_1, \dots, -a_m; -b_1, \dots, -b_n; (-1)^{m+n} z)$ ; voir l'exercice 2.17.

**5.21**  $\lim_{n \rightarrow \infty} (n+m)^{\underline{m}}/n^m = 1$ .

**5.22** En multipliant et divisant des instances de (5.83), on trouve

$$\begin{aligned} \frac{(-1/2)!}{x! (x-1/2)!} &= \lim_{n \rightarrow \infty} \binom{n+x}{n} \binom{n+x-1/2}{n} n^{-2x} / \binom{n-1/2}{n} \\ &= \lim_{n \rightarrow \infty} \binom{2n+2x}{2n} n^{-2x}, \end{aligned}$$

d'après (5.34) et (5.36). On a aussi

$$1/(2x)! = \lim_{n \rightarrow \infty} \binom{2n+2x}{2n} (2n)^{-2x}.$$

Voici l'équivalent en termes de fonction Gamma :

$$\Gamma(x) \Gamma(x + \frac{1}{2}) = \Gamma(2x) \Gamma(\frac{1}{2}) / 2^{2x-1}.$$

**5.23**  $(-1)^n n_i$ , voir (5.50).

**5.24** Cette somme est égale à  $\binom{n}{m} F\left(\begin{smallmatrix} m-n, -m \\ 1/2 \end{smallmatrix} \middle| 1\right) = \binom{2n}{2m}$ , d'après (5.35) et (5.93).

**5.25** C'est équivalent à l'identité suivante, que l'on prouve facilement :

$$(a-b) \frac{a^{\bar{k}}}{(b+1)^{\bar{k}}} = a \frac{(a+1)^{\bar{k}}}{(b+1)^{\bar{k}}} - b \frac{a^{\bar{k}}}{b^{\bar{k}}}.$$

C'est aussi équivalent à la relation suivante, qui fait appel à des opérateurs :  $a - b = (\vartheta + a) - (\vartheta + b)$ .

De même, on a

$$\begin{aligned} (a_1 - a_2) F\left(\begin{smallmatrix} a_1, a_2, a_3, \dots, a_m \\ b_1, \dots, b_n \end{smallmatrix} \middle| z\right) \\ = a_1 F\left(\begin{smallmatrix} a_1+1, a_2, a_3, \dots, a_m \\ b_1, \dots, b_n \end{smallmatrix} \middle| z\right) - a_2 F\left(\begin{smallmatrix} a_1, a_2+1, a_3, \dots, a_m \\ b_1, \dots, b_n \end{smallmatrix} \middle| z\right), \end{aligned}$$

car  $a_1 - a_2 = (a_1 + k) - (a_2 + k)$ . Si  $a_1 - b_1$  est un entier positif ou nul d, cette deuxième identité nous permet d'exprimer  $F(a_1, \dots, a_m; b_1, \dots, b_n; z)$  comme une combinaison linéaire de  $F(a_2 + j, a_3, \dots, a_m; b_2, \dots, b_n; z)$  pour  $0 \leq j \leq d$ , éliminant de la sorte un paramètre du haut et un paramètre du bas. On obtient ainsi, par exemple, des formules closes pour  $F(a, b; a-1; z)$ ,  $F(a, b; a-2; z)$ , etc.

Gauss [143, §7] trouva des relations analogues entre  $F(a, b; c; z)$  et toute fonction hypergéométrique “contiguë”, dans laquelle un paramètre est modifié de  $\pm 1$ . Rainville [301] a généralisé aux cas où les paramètres sont plus nombreux.

**5.26** Si le rapport des termes de la série hypergéométrique d'origine est  $t_{k+1}/t_k = r(k)$ , celui de la nouvelle est  $t_{k+2}/t_{k+1} = r(k+1)$ . Par conséquent,

$$F\left(\begin{matrix} a_1, \dots, a_m \\ b_1, \dots, b_n \end{matrix} \middle| z\right) = 1 + \frac{a_1 \dots a_m z}{b_1 \dots b_n} F\left(\begin{matrix} a_1+1, \dots, a_m+1, 1 \\ b_1+1, \dots, b_n+1, 2 \end{matrix} \middle| z\right).$$

**5.27** C'est la somme des termes pairs de  $F(2a_1, \dots, 2a_m; 2b_1, \dots, 2b_m; z)$ . On a  $(2a)^{\overline{2k+2}}/(2a)^{\overline{2k}} = 4(k+a)(k+a+\frac{1}{2})$ , etc.

**5.28**  $F\left(\begin{matrix} a, b \\ c \end{matrix} \middle| z\right) = (1-z)^{-a} F\left(\begin{matrix} a, c-b \\ c \end{matrix} \middle| \frac{-z}{1-z}\right) = (1-z)^{-a} F\left(\begin{matrix} c-b, a \\ c \end{matrix} \middle| \frac{-z}{1-z}\right) = (1-z)^{c-a-b} F\left(\begin{matrix} c-a, c-b \\ c \end{matrix} \middle| z\right)$ . Pour sa part, Euler prouva cette identité en montrant que les deux membres satisfont la même équation différentielle. La règle de réflexion est souvent attribuée à Euler, mais il semble qu'elle n'apparaît pas dans ses publications.

En mettant en équations les coefficients de  $z^n$ , on obtient la formule de Pfaff-Saalschütz (5.97).

**5.29** D'après la convolution de Vandermonde, les coefficients de  $z^n$  sont égaux. (La preuve originale de Kummer était différente : il considérait  $\lim_{m \rightarrow \infty} F(m, b-a; b; z/m)$  dans la règle de réflexion (5.101)).

**5.30** Dérivez encore une fois pour obtenir  $z(1-z)F''(z) + (2-3z)F'(z) - F(z) = 0$ . Alors  $F(z) = F(1, 1; 2; z)$  d'après (5.108).

**5.31** La condition  $f(k) = T(k+1) - T(k)$  entraîne que  $f(k+1)/f(k) = (T(k+2)/T(k+1) - 1)/(1 - T(k)/T(k+1))$  est une fonction rationnelle de  $k$ .

**5.32** Quand il s'agit de sommer un polynôme en  $k$ , la méthode de Gosper se ramène à la “méthode des coefficients indéterminés”. On sait que  $q(k) = r(k) = 1$ , et on essaie de résoudre  $p(k) = s(k+1) - s(k)$ . La méthode suggère de prendre  $d = \deg(p) + 1$  pour le degré de  $s(k)$ .

**5.33** La solution de  $k = (k-1)s(k+1) - (k+1)s(k)$  est  $s(k) = -k + \frac{1}{2}$  ; donc la réponse est  $(1-2k)/2k(k-1) + C$ .

**5.34** La limite donne le bon résultat car, d'une part tous les termes tels que  $k > c$  sont nuls, et d'autre part le  $\epsilon - c$  et le  $-c$  se neutralisent mutuellement

lors du passage à la limite. La seconde somme partielle est donc égale à  $\lim_{\epsilon \rightarrow 0} F(-m, -n; \epsilon - m; 1) = \lim_{\epsilon \rightarrow 0} (\epsilon + n - m)^{\overline{m}} / (\epsilon - m)^{\overline{m}} = (-1)^m \binom{n-1}{m}$ .

**5.35** (a)  $2^{-n} 3^n [n \geq 0]$ . (b)  $(1 - \frac{1}{2})^{-k-1} [k \geq 0] = 2^{k+1} [k \geq 0]$ .

**5.36** La somme des chiffres de  $m + n$  est égale à la somme des chiffres de  $m$  plus la somme des chiffres de  $n$ , moins  $p - 1$  fois le nombre de retenues, car chaque retenue diminue la somme des chiffres de  $p - 1$ . On trouvera dans [226] des extensions de ce résultat aux coefficients binomiaux généralisés.

**5.37** En divisant la première identité par  $n!$ , on trouve la convolution de Vandermonde  $\binom{x+y}{n} = \sum_k \binom{x}{k} \binom{y}{n-k}$ . La seconde identité peut se déduire de la formule  $x^{\overline{k}} = (-1)^k (-x)^{\underline{k}}$ , par exemple, en changeant le signe de  $x$  et  $y$ .

**5.38** Prenez le plus grand  $c$  possible tel que  $\binom{c}{3} \leq n$ . Alors  $0 \leq n - \binom{c}{3} < \binom{c+1}{3} - \binom{c}{3} = \binom{c}{2}$ . Remplacez  $n$  par  $n - \binom{c}{3}$  et continuez de la même façon. Notez qu'on peut généraliser cette décomposition en écrivant de façon unique

$$n = \binom{a_1}{1} + \binom{a_2}{2} + \cdots + \binom{a_m}{m}, \quad 0 \leq a_1 < a_2 < \cdots < a_m$$

pour tout  $m$  fixé.

**5.39** Par induction sur  $m + n$ , on trouve que, pour tous  $m > 0$  et  $n > 0$ ,  $x^m y^n = \sum_{k=1}^m \binom{m+n-1-k}{n-1} a^n b^{m-k} x^k + \sum_{k=1}^n \binom{m+n-1-k}{m-1} a^{n-k} b^m y^k$ .

**5.40**  $(-1)^{m+1} \sum_{k=1}^n \sum_{j=1}^m \binom{r}{j} \binom{m-rk-s-1}{m-j} = (-1)^m \sum_{k=1}^n \left( \binom{m-rk-s-1}{m} - \binom{m-r(k-1)-s-1}{m} \right) = (-1)^m \left( \binom{m-rn-s-1}{m} - \binom{m-s-1}{m} \right) = \binom{rn+s}{m} - \binom{s}{m}$ .

**5.41**  $\sum_{k \geq 0} n!/(n-k)! (n+k+1)! = (n!/(2n+1)!) \sum_{k>n} \binom{2n+1}{k}$ , ce qui donne  $2^{2n} n!/(2n+1)!$ .

**5.42** Considérons  $n$  comme une variable réelle indéterminée. En utilisant la méthode de Gosper avec  $q(k) = k+1$  et  $r(k) = k-1-n$ , on trouve la solution  $s(k) = 1/(n+2)$ . La somme indéfinie vaut donc  $(-1)^{x-1} \frac{n+1}{n+2} / \binom{n+1}{x}$ . Alors

$$\sum_{k=0}^n (-1)^k / \binom{n}{k} = (-1)^{x-1} \frac{n+1}{n+2} / \left( \binom{n+1}{x} \right) \Big|_0^{n+1} = 2 \frac{n+1}{n+2} [\text{n pair}].$$

Notons en passant qu'on peut déduire de cet exercice l'expression

$$\frac{1}{n \binom{n-1}{k}} = \frac{1}{(n+1) \binom{n}{k+1}} + \frac{1}{(n+1) \binom{n}{k}},$$

formule “duale” de la récurrence de base (5.8).

La phrase encadrée de l'autre côté de cette page est vraie.

**5.43** Après avoir fait le remplacement conseillé, on peut appliquer (5.21) et sommer sur  $k$ . On peut alors appliquer encore une fois (5.21) puis finir le travail avec la convolution de Vandermonde. (Andrews [10] en a donné une preuve combinatoire. D'autre part, cette identité fournit une preuve rapide de (5.29), comme expliqué dans [207, exercice 1.2.6-62]).

**5.44** En développant en factorielles, on montre que

$$\binom{m}{j} \binom{n}{k} \binom{m+n}{m} = \binom{m+n-j-k}{m-j} \binom{j+k}{j} \binom{m+n}{j+k}.$$

Ainsi, la seconde somme est égale à  $1/\binom{m+n}{m}$  fois la première. Or, la première est juste un cas particulier de (5.32), avec  $l=0$ ,  $n=b$ ,  $r=a$  et  $s=m+n-b$ . Elle vaut donc  $\binom{a+b}{a} \binom{m+n-a-b}{n-a}$ .

**5.45** Selon (5.9),  $\sum_{k \leq n} \binom{k-1/2}{k} = \binom{n+1/2}{n}$ . Si cette réponse n'est pas encore assez "close", on peut encore appliquer (5.35) pour trouver  $(2n+1)\binom{2n}{n}4^{-n}$ .

**5.46** D'après (5.69), cette convolution est l'opposé du coefficient de  $z^{2n}$  dans  $\mathcal{B}_{-1}(z)\mathcal{B}_{-1}(-z)$ . Comme  $(2\mathcal{B}_{-1}(z)-1)(2\mathcal{B}_{-1}(-z)-1) = \sqrt{1-16z^2}$ , on a  $\mathcal{B}_{-1}(z)\mathcal{B}_{-1}(-z) = \frac{1}{4}\sqrt{1-16z^2} + \frac{1}{2}\mathcal{B}_{-1}(z) + \frac{1}{2}\mathcal{B}_{-1}(-z) - \frac{1}{4}$ . Selon la formule du binôme,

$$(1-16z^2)^{1/2} = \sum_n \binom{1/2}{n} (-16)^n z^{2n} = -\sum_n \binom{2n}{n} \frac{4^n z^{2n}}{2n-1},$$

La phrase encadrée de l'autre côté de cette page est fausse.

donc la réponse est  $\binom{2n}{n}4^{n-1}/(2n-1) + \binom{4^{n-1}}{2n}/(4n-1)$ .

**5.47** C'est le coefficient de  $z^n$  dans  $(\mathcal{B}_r(z)^s/Q_r(z))(\mathcal{B}_r(z)^{-s}/Q_r(z)) = Q_r(z)^{-2}$ , avec  $Q_r(z) = 1-r+r\mathcal{B}_r(z)^{-1}$ , d'après (5.61).

**5.48**  $F(2n+2, 1; n+2; \frac{1}{2}) = 2^{2n+1}/\binom{2n+1}{n+1}$ , un cas particulier de (5.111).

**5.49** L'identité de Saalschütz (5.97) entraîne que

$$\binom{x+n}{n} \frac{y}{y+n} F\left(\begin{matrix} -x, -n, -n-y \\ -x-n, 1-n-y \end{matrix} \middle| 1\right) = \frac{(y-x)^{\overline{n}}}{(y+1)^{\overline{n}}}.$$

**5.50** Le membre gauche est

$$\begin{aligned} & \sum_{k \geq 0} \frac{a^{\bar{k}} b^{\bar{k}}}{c^{\bar{k}}} \frac{(-z)^k}{k!} \sum_{m \geq 0} \binom{k+a+m-1}{m} z^m \\ &= \sum_{n \geq 0} z^n \sum_{k \geq 0} \frac{a^{\bar{k}} b^{\bar{k}}}{c^{\bar{k}} k!} (-1)^k \binom{n+a-1}{n-k} \end{aligned}$$

et le coefficient de  $z^n$  est égal à

$$\binom{n+a-1}{n} F\left(\begin{matrix} a, b, -n \\ c, a \end{matrix} \middle| 1\right) = \frac{a^{\bar{n}}}{n!} \frac{(c-b)^{\bar{n}}}{c^{\bar{n}}}$$

d'après la convolution de Vandermonde (5.92).

**5.51** (a) La réflexion donne  $F(a, -n; 2a; 2) = (-1)^n F(a, -n; 2a; 2)$ . (Notons que cette formule implique la remarquable identité  $\Delta^{2m+1} f(0) = 0$ , lorsque  $f(n) = 2^n x^n / (2x)^n$ ).

(b) La limite terme à terme est égale à  $\sum_{0 \leq k \leq m} \binom{m}{k} \frac{2m+1}{2m+1-k} (-2)^k$  plus un terme supplémentaire pour  $k = 2m - 1$ . Ce terme supplémentaire est égal à

$$\begin{aligned} & \frac{(-m) \dots (-1)(1) \dots (m)(-2m+1) \dots (-1) 2^{2m+1}}{(-2m) \dots (-1)(2m-1)!} \\ &= (-1)^{m+1} \frac{m! m! 2^{2m+1}}{(2m)!} = \frac{-2}{\binom{-1/2}{m}}; \end{aligned}$$

dons, d'après (5.104), la limite vaut  $-1/\binom{-1/2}{m}$ , soit exactement l'opposé de ce que nous avons calculé précédemment.

**5.52** Les termes des deux séries sont nuls pour  $k > N$ . Cette identité revient à remplacer  $k$  par  $N - k$ . Notez que

$$\begin{aligned} a^{\bar{N}} &= a^{\bar{N-k}} (a + N - k)^{\bar{k}} \\ &= a^{\bar{N-k}} (a + N - 1)^{\bar{k}} = a^{\bar{N-k}} (1 - a - N)^{\bar{k}} (-1)^k. \end{aligned}$$

**5.53** Lorsque  $b = -\frac{1}{2}$ , le membre gauche de (5.110) vaut  $1 - 2z$  et le membre droit est égal à  $(1 - 4z + 4z^2)^{1/2}$ , indépendamment de  $a$ . Ce membre droit est la série formelle

$$1 + \binom{1/2}{1} 4z(z-1) + \binom{1/2}{2} 16z^2(z-1)^2 + \dots,$$

que l'on peut développer et réarranger pour obtenir  $1 - 2z + 0z^2 + 0z^3 + \dots$ . Cependant, ces manipulations font apparaître des séries divergentes pour  $z = 1$  lors des étapes intermédiaires. C'est pour cela qu'elles ne sont pas valables.

**5.54** Si  $m + n$  est impair, disons égal à  $2N - 1$ , il nous faut montrer que

$$\lim_{\epsilon \rightarrow 0} F\left(\begin{matrix} N-m-\frac{1}{2}, -N+\epsilon \\ -m+\epsilon \end{matrix} \middle| 1\right) = 0.$$

On peut appliquer l'équation (5.92) car  $-m + \epsilon > -m - \frac{1}{2} + \epsilon$ . Comme  $N \leq m$ , le facteur  $\Gamma(c - b) = \Gamma(N - m)$  du dénominateur est infini ; tous les

autres facteurs sont finis. Si  $m + n$  est pair, posons  $n = m - 2N$ . D'après (5.93), on a

$$\lim_{\epsilon \rightarrow 0} F\left(\begin{array}{c} -N, N-m-\frac{1}{2}+\epsilon \\ -m+\epsilon \end{array} \middle| 1\right) = \frac{(N-1/2)^N}{m^N}.$$

Il reste à montrer que

$$\binom{m}{m-2N} \frac{(N-1/2)!}{(-1/2)!} \frac{(m-N)!}{m!} = \binom{m-N}{m-2N} 2^{-2N}.$$

C'est fait dans l'exercice 22, avec  $x = N$ .

**5.55** Soit  $Q(k) = (k + A_1) \dots (k + A_M)Z$  et  $R(k) = (k + B_1) \dots (k + B_N)$ . Alors  $t(k+1)/t(k) = P(k)Q(k-1)/P(k-1)R(k)$ , où  $P(k) = Q(k) - R(k)$  est un polynôme non nul.

**5.56** La solution de  $-(k+1)(k+2) = s(k+1) + s(k)$  est  $s(k) = -\frac{1}{2}k^2 - k - \frac{1}{4}$ ; donc  $\sum \binom{-3}{k} \delta k = \frac{1}{8}(-1)^{k-1}(2k^2 + 4k + 1) + C$ . D'autre part,

$$\begin{aligned} & (-1)^{k-1} \left[ \frac{k+1}{2} \right] \left[ \frac{k+2}{2} \right] \\ &= \frac{(-1)^{k-1}}{4} \left( k+1 - \frac{1+(-1)^k}{2} \right) \left( k+2 - \frac{1-(-1)^k}{2} \right) \\ &= \frac{(-1)^{k-1}}{8} (2k^2 + 4k + 1) + \frac{1}{8}. \end{aligned}$$

**5.57** On a  $t(k+1)/t(k) = (k-n)(k+1+\theta)(-z)/(k+1)(k+\theta)$ . Posons alors  $p(k) = k+\theta$ ,  $q(k) = (k-n)(-z)$  et  $r(k) = k$ . La fonction  $s(k)$  est forcément une constante  $\alpha_0$ , et on a

$$k + \theta = (-z(k-n) - k) \alpha_0;$$

donc  $\alpha_0 = -1/(1+z)$  et  $\theta = -nz/(1+z)$ . La somme vaut

$$\sum \binom{n}{k} z^k \left( k - \frac{nz}{1+z} \right) \delta k = -\frac{n}{1+z} \binom{n-1}{k-1} z^k + C.$$

(Le cas particulier  $z = 1$  est considéré en (5.18)).

**5.58** Si  $m > 0$ , on peut remplacer  $\binom{k}{m}$  par  $\frac{k}{m} \binom{k-1}{m-1}$  et en déduire la formule  $T_{m,n} = \frac{n}{m} T_{m-1,n-1} - \frac{1}{m} \binom{n-1}{m}$ . Par conséquent, le facteur de sommation  $\binom{n}{m}^{-1}$  est celui qu'il nous faut :

$$\frac{T_{m,n}}{\binom{n}{m}} = \frac{T_{m-1,n-1}}{\binom{n-1}{m-1}} - \frac{1}{m} + \frac{1}{n}.$$

Développons cela pour obtenir

$$\frac{T_{m,n}}{\binom{n}{m}} = T_{0,n-m} - H_m + H_n - H_{n-m}.$$

Pour finir,  $T_{0,n-m} = H_{n-m}$ , donc  $T_{m,n} = \binom{n}{m}(H_n - H_m)$ . (On peut aussi obtenir ce résultat avec des fonctions génératrices ; voir l'exemple 2 de la section 7.5).

$$5.59 \quad \sum_{j \geq 0, k \geq 1} \binom{n}{j} [j = \lfloor \log_m k \rfloor] = \sum_{j \geq 0, k \geq 1} \binom{n}{j} [m^j \leq k < m^{j+1}] = \sum_{j \geq 0} \binom{n}{j} (m^{j+1} - m^j) = (m-1) \sum_{j \geq 0} \binom{n}{j} m^j = (m-1)(m+1)^n.$$

5.60  $\binom{2n}{n} \approx 4^n / \sqrt{\pi n}$  constitue le cas  $m = n$  de

$$\binom{m+n}{n} \approx \sqrt{\frac{1}{2\pi} \left( \frac{1}{m} + \frac{1}{n} \right)} \left( 1 + \frac{m}{n} \right)^n \left( 1 + \frac{n}{m} \right)^m.$$

5.61 Soient  $\lfloor n/p \rfloor = q$  et  $n \bmod p = r$ . L'identité  $(x+1)^p \equiv x^p + 1 \pmod{p}$  entraîne que

$$(x+1)^{pq+r} \equiv (x+1)^r(x^p+1)^q \pmod{p}.$$

*La phrase encadrée de l'autre côté de cette page n'est pas une phrase.*

Le coefficient de  $x^m$  dans le membre gauche est  $\binom{n}{m}$ . Du côté droit, c'est  $\sum_k \binom{r}{m-pk} \binom{q}{k}$ , qui est égal à  $\binom{r}{m \bmod p} \binom{q}{\lfloor m/p \rfloor}$  car  $0 \leq r < p$ .

5.62  $\binom{np}{mp} = \sum_{k_1+\dots+k_n=mp} \binom{p}{k_1} \dots \binom{p}{k_n} \equiv \binom{n}{m} \pmod{p^2}$ , car tous les termes de la somme sont des multiples de  $p^2$ , sauf les  $\binom{n}{m}$  termes dans lesquels exactement  $m$  variables  $k$  sont égales à  $p$ . (Stanley [335, exercice 1.6(d)] montre que la congruence est valide modulo  $p^3$  lorsque  $p > 3$ ).

5.63  $S_n = \sum_{k=0}^n (-4)^k \binom{n+k}{n-k} = \sum_{k=0}^n (-4)^{n-k} \binom{2n-k}{k}$ . Comme le dénominateur de (5.74) est nul pour  $z = -1/4$ , on ne peut pas utiliser cette formule. La récurrence  $S_n = -2S_{n-1} - S_{n-2}$  mène à la solution  $S_n = (-1)^n(2n+1)$ .

5.64  $\sum_{k \geq 0} ((\binom{n}{2k} + \binom{n}{2k+1}) / (k+1)) = \sum_{k \geq 0} \binom{n+1}{2k+1} / (k+1)$ , ce qui donne

$$\frac{2}{n+2} \sum_{k \geq 0} \binom{n+2}{2k+2} = \frac{2^{n+2}-2}{n+2}.$$

5.65 Multipliez les deux membres par  $n^{n-1}$  et remplacez  $k$  par  $n-1-k$  pour obtenir

$$\begin{aligned} \sum_k \binom{n-1}{k} n^k (n-k)! &= (n-1)! \sum_{k=0}^{n-1} (n^{k+1}/k! - n^k/(k-1)!) \\ &= (n-1)! n^n / (n-1)!. \end{aligned}$$

(Les sommes partielles peuvent être calculées par l'algorithme de Gosper). Voici une preuve combinatoire de la même formule :  $\binom{n}{k}kn^{n-1-k}k!$  est le nombre d'applications de  $\{1, \dots, n\}$  dans lui-même telles que  $f(1), \dots, f(k)$  sont distincts mais  $f(k+1) \in \{f(1), \dots, f(k)\}$ . En sommant sur  $k$ , on doit trouver  $n^n$ .

**5.66** C'est un problème dans lequel il y a "à l'évidence", à chaque étape, une seule chose à faire. Remplacez d'abord  $k-j$  par  $l$ , puis  $\lfloor \sqrt{l} \rfloor$  par  $k$  pour obtenir

$$\sum_{j,k \geq 0} \binom{-1}{j-k} \binom{j}{m} \frac{2k+1}{2^j}.$$

Cette série infinie converge car, pour  $j$  fixé, les termes sont donnés par un polynôme en  $j$  divisé par  $2^j$ . Sommez maintenant sur  $k$ , ce qui donne

$$\sum_{j \geq 0} \binom{j}{m} \frac{j+1}{2^j}.$$

Faites absorber le  $j+1$  et appliquez (5.57) pour aboutir à la réponse 4( $m+1$ ).

**5.67**  $3\binom{2n+2}{n+5}$  d'après (5.26), car

$$\binom{\binom{k}{2}}{2} = 3\binom{k+1}{4}.$$

**5.68** En utilisant le fait que

$$\sum_{k \leq n/2} \binom{n}{k} = 2^{n-1} + \frac{1}{2}\binom{n}{n/2} [\text{n est pair}],$$

on obtient  $n(2^{n-1} - \binom{n-1}{\lfloor n/2 \rfloor})$ .

**5.69** Comme  $\binom{k+1}{2} + \binom{l-1}{2} \leq \binom{k}{2} + \binom{l}{2} \iff k < l$ , le minimum est atteint lorsque les  $k$  sont aussi égaux que possible. Donc, d'après la formule d'équipartition du chapitre 3, le minimum est

$$\begin{aligned} (n \bmod m) \binom{\lceil n/m \rceil}{2} + (n - (n \bmod m)) \binom{\lfloor n/m \rfloor}{2} \\ = n \binom{\lfloor n/m \rfloor}{2} + (n \bmod m) \left\lfloor \frac{n}{m} \right\rfloor. \end{aligned}$$

On obtient un résultat similaire pour tout autre index du bas à la place du 2.

*La phrase encadrée de l'autre côté de cette page n'est pas encadrée.*

**5.70** C'est  $F(-n, \frac{1}{2}; 1; 2)$  ; mais c'est aussi  $(-2)^{-n} \binom{2n}{n} F(-n, -n; \frac{1}{2} - n; \frac{1}{2})$  si on remplace  $k$  par  $n - k$ . Or,  $F(-n, -n; \frac{1}{2} - n; \frac{1}{2}) = F(-\frac{n}{2}, -\frac{n}{2}; \frac{1}{2} - n; 1)$  d'après l'identité de Gauss (5.111). (Autre méthode :  $F(-n, -n; \frac{1}{2} - n; \frac{1}{2}) = 2^{-n} F(-n, \frac{1}{2}; \frac{1}{2} - n; -1)$  d'après la règle de réflexion (5.101), et la formule de Kummer (5.94) lie ceci à (5.55)). Finalement, la réponse est 0 si  $n$  est impair et  $2^{-n} \binom{n}{n/2}$  s'il est pair. (On trouvera dans [164, §1.2] une preuve différente. Cette somme apparaît dans l'étude d'un algorithme de recherche [195]).

**5.71** (a) Remarquez que

$$S(z) = \sum_{k \geq 0} a_k \frac{z^{m+k}}{(1-z)^{m+2k+1}} = \frac{z^m}{(1-z)^{m+1}} A(z/(1-z)^2).$$

(b) Ici,  $A(z) = \sum_{k \geq 0} \binom{2k}{k} (-z)^k / (k+1) = (\sqrt{1+4z} - 1) / 2z$ , donc on a  $A(z/(1-z)^2) = 1 - z$ . Ainsi,  $S_n = [z^n] (z/(1-z))^m = \binom{n-1}{n-m}$ .

**5.72** L'expression donnée vaut  $m(m-n) \dots (m-(k-1)n) n^{k-v(k)} / k!$ . Tout diviseur premier  $p$  de  $n$  divise le numérateur au moins  $k-v(k)$  fois et divise le dénominateur au plus  $k-v(k)$  fois, puisque c'est le nombre de fois que 2 divise  $k!$ . Un nombre premier  $p$  qui ne divise pas  $n$  doit diviser le produit  $m(m-n) \dots (m-(k-1)n)$  au moins autant de fois qu'il divise  $k!$ , car  $m(m-n) \dots (m-(p^r-1)n)$  est un multiple de  $p^r$  pour tout  $r \geq 1$  et pour tout  $m$ .

**5.73** Si on prend  $X_n = n!$ , on trouve  $\alpha = \beta = 1$  ; si on prend  $X_n = n_1$ , on trouve  $\alpha = 1$  et  $\beta = 0$ . La solution générale est donc  $X_n = \alpha n_1 + \beta(n! - n_1)$ .

**5.74**  $\binom{n+1}{k} - \binom{n-1}{k-1}$ , pour  $1 \leq k \leq n$ .

**5.75** La récurrence  $S_k(n+1) = S_k(n) + S_{(k-1) \bmod 3}(n)$  permet de vérifier par induction que, pour  $n$  donné, deux des fonctions donnent la même valeur, et que  $S_{(-n) \bmod 3}(n)$  diffère de ces deux-là de  $(-1)^n$ . Comme ces trois valeurs partagent leur somme  $S_0(n) + S_1(n) + S_2(n) = 2^n$  aussi également que possible, il y a  $2^n \bmod 3$  occurrences de  $\lceil 2^n/3 \rceil$  et  $3 - (2^n \bmod 3)$  occurrences de  $\lfloor 2^n/3 \rfloor$ .

**5.76**  $Q_{n,k} = (n+1) \binom{n}{k} - \binom{n}{k+1}$ .

**5.77** Les termes sont nuls, sauf lorsque  $k_1 \leq \dots \leq k_m$ , c'est-à-dire lorsque le produit est le coefficient multinomial

$$\binom{k_m}{k_1, k_2 - k_1, \dots, k_m - k_{m-1}}.$$

Par conséquent, la somme sur  $k_1, \dots, k_{m-1}$  est égale à  $m^{k_m}$ , et la somme finale sur  $k_m$  vaut  $(m^{n+1} - 1)/(m - 1)$ .

**5.78** Etendez la somme jusqu'à  $k = 2m^2 + m - 1$ ; les termes nouveaux sont  $\binom{1}{4} + \binom{2}{6} + \cdots + \binom{m-1}{2m} = 0$ . Comme  $m \perp (2m+1)$ , les couples  $(k \bmod m, k \bmod (2m+1))$  sont distincts. De plus, les nombres  $(2j+1) \bmod (2m+1)$ , pour  $j$  variant de 0 à  $2m$ , sont les nombres  $0, 1, \dots, 2m$ , éventuellement dans le désordre. Par conséquent, la somme fait

$$\sum_{\substack{0 \leq k < m \\ 0 \leq j < 2m+1}} \binom{k}{j} = \sum_{0 \leq k < m} 2^k = 2^m - 1.$$

**5.79** (a) Comme la somme est égale à  $2^{2n-1}$ , le pgcd doit être une puissance de 2. Si  $n = 2^k q$  avec  $q$  impair,  $\binom{2n}{1}$  est divisible par  $2^{k+1}$  et pas par  $2^{k+2}$ . Chacun des  $\binom{2n}{2j+1}$  est divisible par  $2^{k+1}$  (voir l'exercice 36), donc c'est bien le pgcd. (b) Si  $p^r \leq n+1 < p^{r+1}$ , c'est en additionnant  $k$  et  $n-k$ , avec  $k = p^r - 1$ , qu'on effectue le plus de retenues en base  $p$ . Dans ce cas, le nombre de retenues est égal à  $r - \epsilon_p(n+1)$ , et  $r = \epsilon_p(L(n+1))$ .

**5.80** Montrez d'abord par induction que  $k! \geq (k/e)^k$ .

**5.81** Soit  $f_{l,m,n}(x)$  le membre gauche. Il suffit de montrer que  $f_{l,m,n}(1) > 0$  et que  $f'_{l,m,n}(x) < 0$  pour  $0 \leq x \leq 1$ . D'après (5.23), la valeur de  $f_{l,m,n}(1)$  est  $(-1)^{n-m-1} \binom{l+m+\theta}{l+n}$ , et elle est strictement positive car le coefficient binomial a exactement  $n-m-1$  facteurs strictement négatifs. L'inégalité est vraie lorsque  $l=0$ , pour la même raison. Si  $l>0$ , on a  $f'_{l,m,n}(x) = -l f_{l-1,m,n+1}(x)$ , ce qui est négatif par induction.

**5.82** Soit  $\epsilon_p(a)$  l'exposant de  $p$  dans la décomposition en facteurs premiers de  $a$ , et soit  $m = n - k$ . L'identité à prouver se réduit à

$$\begin{aligned} & \min(\epsilon_p(m) - \epsilon_p(m+k), \epsilon_p(m+k+1) - \epsilon_p(k+1), \epsilon_p(k) - \epsilon_p(m+1)) \\ &= \min(\epsilon_p(k) - \epsilon_p(m+k), \epsilon_p(m) - \epsilon_p(k+1), \epsilon_p(m+k+1) - \epsilon_p(m+1)). \end{aligned}$$

Pour simplifier, convenons d'écrire cela  $\min(x_1, y_1, z_1) = \min(x_2, y_2, z_2)$ . Remarquez que  $x_1 + y_1 + z_1 = x_2 + y_2 + z_2$ . La relation générale  $\epsilon_p(a) < \epsilon_p(b) \implies \epsilon_p(a) = \epsilon_p(|a \pm b|)$  nous permet de conclure que  $x_1 \neq x_2 \implies \min(x_1, x_2) = 0$ ; c'est vrai aussi pour  $(y_1, y_2)$  et  $(z_1, z_2)$ . La fin de la preuve ne pose pas de difficulté.

**5.83** (Solution de P. Paule). Soit  $r$  un entier positif ou nul. La somme donnée représente le coefficient de  $x^l y^m$  dans

$$\begin{aligned} & \sum_{j,k} (-1)^{j+k} \frac{(1+x)^{j+k}}{x^k} \binom{r}{j} \binom{n}{k} (1+y)^{s+n-j-k} y^j \\ &= \left(1 - \frac{(1+x)y}{1+y}\right)^r \left(1 - \frac{1+x}{(1+y)x}\right)^n (1+y)^{s+n} \\ &= (-1)^n (1-xy)^{n+r} (1+y)^{s-r}/x^n, \end{aligned}$$

par conséquent il est clair qu'elle vaut  $(-1)^l \binom{n+r}{n+l} \binom{s-r}{m-n-l}$ . (Voir aussi l'exercice 106).

**5.84** En suivant le conseil donné, on obtient

$$z\mathcal{B}_t(z)^{r-1}\mathcal{B}'_t(z) = \sum_{k \geq 0} \binom{tk+r}{k} \frac{kz^k}{tk+r},$$

ainsi qu'une formule similaire pour  $\mathcal{E}_t(z)$ . Par conséquent, les formules  $(zt\mathcal{B}_t^{-1}(z)\mathcal{B}'_t(z)+1)\mathcal{B}_t(z)^r$  et  $(zt\mathcal{E}_t^{-1}(z)\mathcal{E}'_t(z)+1)\mathcal{E}_t(z)^r$  donnent respectivement les membres droits de (5.61). Il nous faut encore prouver que

$$(zt\mathcal{B}_t^{-1}(z)\mathcal{B}'_t(z)+1)\mathcal{B}_t(z)^r = \frac{1}{1-t+t\mathcal{B}_t(z)^{-1}},$$

$$(zt\mathcal{E}_t^{-1}(z)\mathcal{E}'_t(z)+1)\mathcal{E}_t(z)^r = \frac{1}{1-zt\mathcal{E}(z)^t},$$

ce qui découle de (5.59).

**5.85** Pour tout polynôme  $f(x) = a_n x^n + \dots + a_1 x + a_0$  de degré  $\leq n$ , on peut prouver par induction que

$$\sum_{0 \leq \epsilon_1, \dots, \epsilon_n \leq 1} (-1)^{\epsilon_1 + \dots + \epsilon_n} f(\epsilon_1 x_1 + \dots + \epsilon_n x_n) = (-1)^n n! a_n x_1 \dots x_n.$$

L'identité à démontrer constitue le cas particulier où  $a_n = 1/n!$  et  $x_k = k^3$ .

La phrase encadrée de l'autre côté de cette page est auto-référente.

**5.86** (a) Commencez par développer en prenant  $n(n-1)$  variables d'indice  $l_{ij}$  pour tous  $i \neq j$ . En posant  $k_{ij} = l_{ij} - l_{ji}$  pour  $1 \leq i < j < n$  et en utilisant la contrainte  $\sum_{i \neq j} (l_{ij} - l_{ji}) = 0$  pour tout  $i < n$ , on peut effectuer les sommes sur  $l_{jn}$  pour  $1 \leq j < n$ , puis sur  $l_{ji}$  pour  $1 \leq i < j < n$ , avec la convolution de Vandermonde. (b)  $f(z) - 1$  est un polynôme de degré  $< n$  qui a  $n$  racines ; il est donc identiquement nul. (c) Considérez les termes constants de

$$\prod_{\substack{1 \leq i, j \leq n \\ i \neq j}} \left(1 - \frac{z_i}{z_j}\right)^{a_i} = \sum_{k=1}^n \prod_{\substack{1 \leq i, j \leq n \\ i \neq j}} \left(1 - \frac{z_i}{z_j}\right)^{a_i - [i=k]},$$

**5.87** Le premier terme vaut  $\sum_k \binom{n-k}{k} z^{mk}$ , d'après (5.61), et voici les termes généraux du second terme :

$$\begin{aligned} & \frac{1}{m} \sum_{k \geq 0} \binom{(n+1)/m + (1+1/m)k}{k} (\zeta z)^{k+n+1} \\ &= \frac{1}{m} \sum_{k > n} \binom{(1+1/m)k - n - 1}{k-n-1} (\zeta z)^k. \end{aligned}$$

Comme  $\sum_{0 \leq j < m} (\zeta^{2j+1})^k = m(-1)^l [k=m]$ , ces termes se somment en

$$\begin{aligned} & \sum_{k>n/m} \binom{(1+1/m)mk - n - 1}{mk - n - 1} (-z^m)^k \\ &= \sum_{k>n/m} \binom{(m+1)k - n - 1}{k} (-z^m)^k = \sum_{k>n/m} \binom{n - mk}{k} z^{mk}. \end{aligned}$$

Notons en passant que les fonctions  $B_m(z^m)$  et  $\zeta^{2j+1} z B_{1+1/m}(\zeta^{2j+1} z)^{1/m}$  sont les  $m+1$  racines complexes de l'équation  $w^{m+1} - w^m = z^m$ .

**5.88** Utilisez les faits que  $\int_0^\infty (e^{-t} - e^{-nt}) dt/t = \ln n$  et que  $(1 - e^{-t})/t \leq 1$ . (On a  $\binom{x}{k} = O(k^{-x-1})$  lorsque  $k \rightarrow \infty$ , selon (5.83) ; cela implique que la série de Stirling  $\sum_k s_k \binom{x}{k}$  converge lorsque  $x > -1$ . Hermite [186] a démontré que la somme vaut  $\ln \Gamma(1+x)$ ).

**5.89** En additionnant l'identité donnée à (5.19), on obtient  $y^{-r}(x+y)^{m+r}$  dans les deux membres en appliquant la formule du binôme. La dérivation donne

$$\begin{aligned} & \sum_{k>m} \binom{m+r}{k} \binom{m-k}{n} x^k y^{m-k-n} \\ &= \sum_{k>m} \binom{-r}{k} \binom{m-k}{n} (-x)^k (x+y)^{m-k-n}, \end{aligned}$$

La phrase encadrée de l'autre côté de cette page n'est pas auto-référente.

et on peut remplacer  $k$  par  $k+m+1$  puis appliquer (5.15) pour obtenir

$$\begin{aligned} & \sum_{k \geq 0} \binom{m+r}{m+1+k} \binom{-n-1}{k} (-x)^{m+1+k} y^{-1-k-n} \\ &= \sum_{k \geq 0} \binom{-r}{m+1+k} \binom{-n-1}{k} x^{m+1+k} (x+y)^{-1-k-n}. \end{aligned}$$

Sous forme hypergéométrique, cela donne

$$F\left(\frac{1-r, n+1}{m+2} \mid \frac{-x}{y}\right) = \left(1 + \frac{x}{y}\right)^{-n-1} F\left(\frac{m+1+r, n+1}{m+2} \mid \frac{x}{x+y}\right),$$

ce qui correspond au cas particulier  $(a, b, c, z) = (n+1, m+1+r, m+2, -x/y)$  de la règle de réflexion (5.101). (Ainsi, (5.105) est liée à la fois à la réflexion et à la formule de l'exercice 52).

**5.90** Si  $r$  est un entier positif ou nul, la somme est finie et le calcul effectué dans le problème 1 est valide tant qu'aucun des termes de la somme pour  $0 \leq k \leq r$  n'a de zéro au dénominateur. Sinon, la somme est infinie, et le  $k$ ième terme  $\binom{k-r-1}{k}/\binom{k-s-1}{k}$  vaut approximativement  $k^{s-r}(-s-1)!/(-r-1)!$  d'après (5.83). Il est donc nécessaire que  $r > s+1$  pour que la série converge

(si  $r$  et  $s$  sont complexes, la condition devient  $\Re r > \Re s + 1$ , car  $|k^z| = k^{\Re z}$ ). La somme vaut

$$F\left(\begin{matrix} -r, 1 \\ -s \end{matrix} \middle| 1\right) = \frac{\Gamma(r-s-1)\Gamma(-s)}{\Gamma(r-s)\Gamma(-s-1)} = \frac{s+1}{s+1-r}$$

d'après (5.92). C'est exactement la même formule que lorsque  $r$  et  $s$  sont entiers.

**5.91** (Il vaut mieux disposer d'un ordinateur pour faire cette preuve). Notons en passant que si  $c = (a+1)/2$ , on se ramène à une identité qui se trouve équivalente à l'identité de Gauss (5.110) en appliquant la règle de réflexion de Pfaff. En effet, si  $w = -z/(1-z)$ , on a  $4w(1-w) = -4z/(1-z)^2$ , et

$$\begin{aligned} F\left(\begin{matrix} \frac{1}{2}a, \frac{1}{2}a+\frac{1}{2}-b \\ 1+a-b \end{matrix} \middle| 4w(1-w)\right) &= F\left(\begin{matrix} a, a+1-2b \\ 1+a-b \end{matrix} \middle| \frac{-z}{1-z}\right) \\ &= (1-z)^a F\left(\begin{matrix} a, b \\ 1+a-b \end{matrix} \middle| z\right). \end{aligned}$$

**5.92** On peut prouver ces identités, comme Clausen l'a fait il y a plus de 150 ans, en montrant que les deux membres satisfont la même équation différentielle. Les équations qui lient les coefficients de  $z^n$  peuvent s'écrire avec des coefficients binomiaux :

$$\begin{aligned} \sum_k \frac{\binom{r}{k} \binom{s}{k} \binom{r}{n-k} \binom{s}{n-k}}{\binom{r+s-1/2}{k} \binom{r+s-1/2}{n-k}} &= \frac{\binom{2r}{n} \binom{r+s}{n} \binom{2s}{n}}{\binom{2r+2s}{n} \binom{r+s-1/2}{n}}; \\ \sum_k \frac{\binom{-1/4+r}{k} \binom{-1/4+s}{k} \binom{-1/4-r}{n-k} \binom{-1/4-s}{n-k}}{\binom{-1+r+s}{k} \binom{-1-r-s}{n-k}} \\ &= \frac{\binom{-1/2}{n} \binom{-1/2+r-s}{n} \binom{-1/2-r+s}{n}}{\binom{-1+r+s}{n} \binom{-1-r-s}{n}}. \end{aligned}$$

On peut aussi les exprimer avec des séries hypergéométriques :

$$F\left(\begin{matrix} a, b, \frac{1}{2}-a-b-n, -n \\ \frac{1}{2}+a+b, 1-a-n, 1-b-n \end{matrix} \middle| 1\right) = \frac{(2a)^\overline{n} (a+b)^\overline{n} (2b)^\overline{n}}{(2a+2b)^\overline{n} a^\overline{n} b^\overline{n}};$$

$$\begin{aligned} F\left(\begin{array}{l} \frac{1}{4}+a, \frac{1}{4}+b, a+b-n, -n \\ 1+a+b, \frac{3}{4}+a-n, \frac{3}{4}+b-n \end{array} \middle| 1\right) \\ = \frac{(1/2)^{\bar{n}} (1/2+a-b)^{\bar{n}} (1/2-a+b)^{\bar{n}}}{(1+a+b)^{\bar{n}} (1/4-a)^{\bar{n}} (1/4-b)^{\bar{n}}}. \end{aligned}$$

**5.93**  $\alpha^{-1} \prod_{j=1}^k (f(j) + \alpha) / f(j).$

**5.94** L'algorithme de Gosper trouve la réponse :  $-(\frac{a-1}{k-1})(\frac{-a-1}{n-k})a/n + C.$  On peut en déduire que, si  $m \geq 0$  est un entier inférieur à  $n$ , alors

$$\sum \binom{a}{k} \binom{m-a}{n-k} \delta k = \sum_j \binom{m}{j} \frac{-a}{n-j} \binom{a-1}{k-1} \binom{-a-1}{n-j-k} + C.$$

**5.95** Les coefficients directeurs de  $p$  et de  $r$  doivent être égaux à 1, et  $p$  ne doit avoir de facteur commun ni avec  $q$  ni avec  $r$ .

Supposons que  $p(k+1)q(k)/p(k)r(k+1) = P(k+1)Q(k)/P(k)R(k+1)$ , où les deux triplets de polynômes  $(p, q, r)$  et  $(P, Q, R)$  satisfont le nouveau critère. Soient  $p_0(k) = p(k)/g(k)$  et  $P_0(k) = P(k)/g(k)$ , où  $g(k) = \text{pgcd}(p(k), P(k))$  est le produit de tous les facteurs communs de  $p$  et  $P$ . Alors

$$p_0(k+1)q(k)P_0(k)R(k+1) = p_0(k)r(k+1)P_0(k+1)Q(k).$$

Supposons que  $p_0(k) \neq 1$ . Alors il existe un nombre complexe  $\alpha$  tel que  $p_0(\alpha) = 0$ ; ce qui implique que  $q(\alpha) \neq 0$ ,  $r(\alpha) \neq 0$  et  $P_0(\alpha) \neq 0$ . On a donc forcément  $p_0(\alpha+1)R(\alpha+1) = 0$  et  $p_0(\alpha-1)Q(\alpha-1) = 0$ . Soit maintenant  $N$  un entier strictement positif tel que  $p_0(\alpha+N) \neq 0$  et  $p_0(\alpha-N) \neq 0$ . En répétant l'argument  $N$  fois, on trouve  $R(\alpha+1) \dots R(\alpha+N) = 0 = Q(\alpha-1) \dots Q(\alpha-N)$ , ce qui contredit (5.118). Par conséquent  $p_0(k) = 1$ . De même,  $P_0(k) = 1$ , donc  $p(k) = P(k)$ . Selon (5.118),  $q(\alpha) = 0$  implique  $r(\alpha+1) \neq 0$ ; donc  $q(k) \mid Q(k)$ . De même,  $Q(k) \mid q(k)$ , donc  $q(k) = Q(k)$  puisqu'ils ont le même coefficient directeur. Donc  $r(k) = R(k)$ .

**5.96** Si  $r(k)$  est une fonction rationnelle non nulle et  $T(k)$  un terme hypergéométrique, alors  $r(k)T(k)$  est aussi un terme hypergéométrique, que l'on dit *similaire* à  $T(k)$ . (Il est possible que  $r(k)$  vaille  $\infty$  et que  $T(k)$  vaille 0, ou vice versa, pour un nombre fini de valeurs de  $k$ ). En particulier,  $T(k+1)$  est toujours similaire à  $T(k)$ . Si  $T_1(k)$  et  $T_2(k)$  sont des termes hypergéométriques similaires, alors  $T_1(k) + T_2(k)$  est un terme hypergéométrique. Si  $T_1(k), \dots, T_m(k)$  sont tous deux à deux non similaires, avec  $m > 1$ , alors  $T_1(k) + \dots + T_m(k)$  ne peut pas être nul pour presque tout  $k$ , c'est-à-dire pour tous les  $k$  sauf un nombre fini d'entre eux. Voici pourquoi : supposons que c'est possible et considérons un contre-exemple pour lequel  $m$  est minimum. Soit  $r_j(k) = T_j(k+1)/T_j(k)$ . Comme

$T_1(k) + \cdots + T_m(k) = 0$ , on a  $r_m(k)T_1(k) + \cdots + r_m(k)T_m(k) = 0$  et  $r_1(k)T_1(k) + \cdots + r_m(k)T_m(k) = T_1(k+1) + \cdots + T_m(k+1) = 0$ ; donc  $(r_m(k) - r_1(k))T_1(k) + \cdots + (r_m(k) - r_{m-1}(k))T_{m-1}(k) = 0$ . On ne peut pas avoir  $r_m(k) - r_j(k) = 0$  pour tout  $j < m$ , car  $T_j$  et  $T_m$  ne sont pas similaires. Comme  $m$  est minimum, on en déduit que  $m = 2$ . Dans ce cas,  $T_1(k)$  et  $T_2(k)$  sont forcément similaires car ils sont tous les deux nuls pour presque tous les  $k$ .

Soit  $t(k)$  un terme hypergéométrique tel que  $t(k+1)/t(k) = r(k)$ , et supposons que  $t(k) = (T_1(k+1) + \cdots + T_m(k+1)) - (T_1(k) + \cdots + T_m(k))$ , où  $m$  est minimal. Alors  $T_1, \dots, T_m$  sont deux à deux non similaires. Soit  $r_j(k)$  la fonction rationnelle telle que

$$r(k)(T_j(k+1) - T_j(k)) - (T_j(k+2) - T_j(k+1)) = r_j(k)T_j(k).$$

Supposons que  $m > 1$ . Comme  $0 = r(k)t(k) - t(k+1) = r_1(k)T_1(k) + \cdots + r_m(k)T_m(k)$ , on a  $r_j(k) = 0$  pour toutes les valeurs de  $j$  sauf au plus une. Si  $r_j(k) = 0$ , la fonction  $\bar{t}(k) = T_j(k+1) - T_j(k)$  satisfait  $\bar{t}(k+1)/\bar{t}(k) = t(k+1)/t(k)$ . Donc l'algorithme de Gosper trouvera une solution.

**5.97** Supposons que  $z$  est différent de  $-d - 1/d$ , pour tout entier  $d > 0$ . Alors, dans le déroulement de l'algorithme de Gosper, on a  $p(k) = 1$ ,  $q(k) = (k+1)^2$  et  $r(k) = k^2 + kz + 1$ . Comme  $\deg(Q) < \deg(R)$  et  $\deg(p) - \deg(R) + 1 = -1$ , il y a une seule possibilité :  $z = d + 2$  où  $d$  est un entier positif ou nul. Si on essaie de prendre  $s(k) = \alpha_d k^d + \cdots + \alpha_0$ , cela ne va pas pour  $d = 0$ , mais cela convient pour tout  $d > 0$ . (Les équations linéaires obtenues en mettant en équations les coefficients de  $k^d, k^{d-1}, \dots, k^1$  de (5.122) expriment  $\alpha_{d-1}, \dots, \alpha_0$  comme des multiples positifs de  $\alpha_d$ , et l'équation  $1 = \alpha_d + \cdots + \alpha_1$  détermine  $\alpha_d$ ). Par exemple, si  $z = 3$ , la somme indéfinie vaut  $(k+2)k!^2 / \prod_{j=1}^{k-1} (j^2 + 3j + 1) + C$ .

Si  $z = -d - 1/d$ , les termes  $t(k)$  sont infinis pour tout  $k \geq d$ . Il y a deux façons raisonnables de procéder. On peut supprimer le zéro du dénominateur en redéfinissant

$$t(k) = \frac{k!^2}{\prod_{j=d+1}^k (j^2 - j(d+1/d) + 1)} = \frac{(d-1/d)! k!^2}{(k-1/d)! (k-d)!},$$

en forçant de la sorte  $t(k)$  à être nul pour  $0 \leq k < d$  et strictement positif pour  $k \geq d$ . Dans ce cas, l'algorithme de Gosper donne  $p(k) = k^d$ ,  $q(k) = k+1$ ,  $r(k) = k - 1/d$ , et on peut résoudre (5.122) pour trouver  $s(k)$ , car le coefficient de  $k^j$  dans le membre droit est égal à  $(j+1+1/d)\alpha^j$  plus des multiples de  $\{\alpha_{j+1}, \dots, \alpha_d\}$ . Par exemple, si  $d = 2$ , la somme indéfinie vaut  $(3/2)! k! (\frac{2}{7}k^2 - \frac{26}{35}k + \frac{32}{105}) / (k-3/2)! + C$ .

Une autre méthode consiste à essayer de sommer les termes d'origine, mais seulement dans l'intervalle  $0 \leq k < d$ . On a alors le droit de remplacer

$p(k) = k^d$  par

$$p'(k) = \sum_{j=1}^d (-1)^{d-j} j \binom{d}{j} k^{j-1}.$$

Toute suite finie est trivialement sommable, car on peut trouver un polynôme égal à  $t(k)$  pour  $0 \leq k < d$ .

Ceci se justifie par le fait que (5.117) est toujours valable pour  $0 \leq k < d-1$ . On a  $p'(k) = \lim_{\epsilon \rightarrow 0} ((k+\epsilon)^d - k^d)/\epsilon = \lim_{\epsilon \rightarrow 0} (k+\epsilon)^d/\epsilon$ , donc cette astuce supprime un 0 du numérateur et du dénominateur de (5.117), tout comme la règle de L'Hospital. Maintenant, la méthode de Gosper donne une somme indéfinie.

**5.98**  $nS_{n+1} = 2nS_n$ . (Attention : cela ne dit rien sur  $S_1/S_0$ ).

**5.99** Soient  $p(n, k) = (n+1+k)\beta_0(n) + (n+1+a+b+c+k)\beta_1(n) = \hat{p}(n, k)$ ,  $\bar{t}(n, k) = t(n, k)/(n+1+k)$ ,  $q(n, k) = (n+1+a+b+c+k)(a-k)(b-k)$  et  $r(n, k) = (n+1+k)(c+k)k$ . Alors la solution de (5.129) est  $\beta_0(n) = (n+1+a+b+c)(n+1+a+b)$ ,  $\beta_1(n) = -(n+1+a)(n+1+b)$ ,  $\alpha_0(n) = s(n, k) = -1$ . On retrouve la formule (5.134) en observant qu'elle est vraie pour  $n = -a$  et en faisant une induction sur  $n$ .

**5.100** L'algorithme de Gosper-Zeilberger trouve facilement que

$$\frac{n+2}{\binom{n}{k}} - \frac{2n+2}{\binom{n+1}{k}} = \frac{n-k}{\binom{n}{k+1}} - \frac{n+1-k}{\binom{n}{k}}, \quad 0 \leq k < n.$$

En sommant de  $k = 0$  à  $n-1$ , on trouve  $(n+2)(S_n - 1) - (2n+2)(S_{n+1} - 1 - \frac{1}{n+1}) = -n$ . Par conséquent,  $(2n+2)S_{n+1} = (n+2)S_n + 2n+2$ . Il suffit alors d'appliquer un facteur de sommation pour aboutir à l'expression  $S_n = (n+1)2^{-n-1} \sum_{k=1}^{n+1} 2^k/k$ .

**5.101 (a)** Si on fixe  $m$ , l'algorithme de Gosper-Zeilberger découvre que  $(n+2)S_{m,n+2}(z) = (z-1)(n+1)S_{m,n}(z) + (2n+3-z(n-m+1))S_{m,n+1}(z)$ . On peut aussi appliquer la même méthode au terme

$$\beta_0(m, n)t(m, n, k) + \beta_1(m, n)t(m+1, n, k) + \beta_2(m, n)t(m, n+1, k),$$

auquel cas on trouve une récurrence plus simple,

$$(m+1)S_{m+1,n}(z) - (n+1)S_{m,n+1}(z) = (1-z)(m-n)S_{m,n}(z).$$

**(b)** Nous avons ici cinq équations et six inconnues, ce qui rend le travail un peu plus ardu. L'algorithme trouve

$$\begin{aligned} & (n+1)(z-1)^2 \binom{n}{k}^2 z^k - (2n+3)(z+1) \binom{n+1}{k}^2 z^k \\ & + (n+2) \binom{n+2}{k}^2 z^k = T(n, k+1) - T(n, k), \end{aligned}$$

$$T(n, k) = \binom{n+1}{k-1}^2 \frac{s(n, k)}{n+1} z^k,$$

$$s(n, k) = (z-1)k^2 - 2((n+2)z-2n-3)k + (n+2)((n+2)z-4n-5).$$

Par conséquent,  $(n+1)(z-1)^2 S_n(z) - (2n+3)(z+1)S_{n+1}(z) + (n+2)S_{n+2}(z) = 0$ . Remarquez que cette récurrence est valable aussi si  $n$  est négatif, et que  $S_{-n-1}(z) = S_n(z)/(1-z)^{2n+1}$ .

La somme  $S_n(z)$  peut être vue comme une variante du polynôme de Legendre  $P_n(z) = \sum_k \binom{n}{k}^2 (z-1)^{n-k} (z+1)^k / 2^n$ , car elle peut s'écrire  $S_n(z) = (1-z)^n P_n(\frac{1+z}{1-z})$ . De même,  $S_{m,n}(z) = (1-z)^n P_n^{(0,m-n)}(\frac{1+z}{1-z})$  est une variante du polynôme de Jacobi.

**5.102** Comme la somme est égale à  $F(a - \frac{1}{3}n, -n; b - \frac{4}{3}n; -z)$ , on n'a pas besoin de considérer le cas  $z = -1$ . Soit  $n = 3m$ . Nous cherchons des solutions de (5.129) avec

$$p(m, k) = (3m+3-k)^3(m+1-k)\beta_0 + (4m+4-b-k)^4\beta_1,$$

$$q(m, k) = (3m+3-k)(m+1-a-k)z,$$

$$r(m, k) = k(4m+1-b-k),$$

$$s(m, k) = \alpha_2 k^2 + \alpha_1 k + \alpha_0.$$

Le système de cinq équations homogènes qui en résulte a une solution non nulle  $(\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1)$  si et seulement si le déterminant des coefficients est nul. Ce déterminant, qui est un polynôme en  $m$ , ne s'annule que dans huit cas. L'un de ces cas est bien évidemment (5.113), mais nous pouvons maintenant calculer la somme pour tout entier positif ou nul  $n$ , et pas seulement pour  $n \not\equiv 2 \pmod{3}$  :

$$\sum_k \binom{n}{k} \binom{\frac{1}{3}n - \frac{1}{6}}{k} 8^k / \binom{\frac{4}{3}n - \frac{2}{3}}{k} = [1, 1, -\frac{1}{2}] \binom{2n}{n} / \binom{\frac{4}{3}n - \frac{2}{3}}{n}.$$

Dans cette expression, la notation  $[c_0, c_1, c_2]$  désigne une unique valeur,  $c_{n \bmod 3}$ . Voici une autre identité, correspondant au cas  $(a, b, z) = (\frac{1}{2}, 0, 8)$  :

$$\sum_k \binom{n}{k} \binom{\frac{1}{3}n - \frac{1}{2}}{k} 8^k / \binom{\frac{4}{3}n}{k} = [1, 0, 0] 16^{n/3} \binom{\frac{2}{3}n}{\frac{1}{3}n} / \binom{\frac{4}{3}n}{n}.$$

Cette somme est en fait nulle sauf si  $n$  est un multiple de 3, et dans ce cas elle peut s'écrire plus simplement :

$$\sum_k \binom{3m}{k} \binom{2m}{2k} \binom{2k}{k} 2^k / \binom{4m}{k} \binom{m}{k} = 16^m \frac{(3m)! (2m)!}{(4m)! m!}.$$

Et  $z = 0$  ?

Les six autres cas donnent lieu à des sommes plus bizarres encore :

$$\sum_k \binom{n}{k} \binom{\frac{1}{3}n - a}{k} z^k / \binom{\frac{4}{3}n - b}{k}$$

$$= [c_0, c_1, c_2] \frac{\binom{\frac{1}{3}n - a}{[n/3]} \binom{\frac{1}{3}n - a'}{[n/3]} x^{[n/3]}}{\binom{\frac{4}{3}n - b}{n} \binom{\frac{1}{3}n - b}{[n/3]} \binom{\frac{1}{3}n - b'}{[n/3]}}$$

dans lesquelles les valeurs respectives de  $(a, b, z, c_0, c_1, c_2, a', b', x)$  sont

$$(\frac{7}{12}, \frac{1}{3}, 8, 1, -1, 0, \frac{1}{4}, 0, 64); \quad (\frac{1}{4}, 0, 8, 1, 2, 0, \frac{7}{12}, \frac{1}{3}, 64);$$

$$(\frac{5}{12}, \frac{2}{3}, 8, 1, 0, -3, \frac{3}{4}, 0, 64); \quad (\frac{1}{12}, \frac{1}{3}, 8, 1, 3, 0, \frac{3}{4}, 0, 64);$$

$$(\frac{1}{2}, 0, -4, 1, 2, 0, \frac{1}{6}, \frac{1}{3}, -16); \quad (\frac{1}{6}, \frac{2}{3}, -4, 1, 0, -3, \frac{5}{6}, 0, -16).$$

**5.103** Nous supposerons qu'aucun des  $a'_i$  et de  $b'_i$  n'est nul, sinon les facteurs correspondants n'auraient aucune incidence sur les degrés en  $k$ . Soit

$$\hat{t}(n, k) = \hat{p}(n, k) \bar{t}(n, k),$$

où

$$\bar{t}(n, k) = \frac{\prod_{i=1}^p (a_i n + a'_i k + a_i l[a_i < 0] + a''_i)!}{\prod_{i=1}^q (b_i n + b'_i k + b_i l[b_i > 0] + b''_i)!} z^k.$$

On a  $\deg(\hat{p}) = \deg(f) + \max(\sum_{i=1}^q b_i [b_i > 0] - \sum_{i=1}^p a_i [a_i < 0]$ , et aussi  $\sum_{i=1}^p a_i [a_i > 0] - \sum_{i=1}^q b_i [b_i < 0] \geq \deg(f) + \frac{1}{2}l(|a_1| + \dots + |a_p| + |b_1| + \dots + |b_q|)$ , sauf dans les cas très particuliers où le coefficient directeur s'annule. On trouve aussi  $\deg(q) = \sum_{i=1}^p a'_i [a'_i > 0] - \sum_{i=1}^q b'_i [b'_i < 0]$ ,  $\deg(r) = \sum_{i=1}^q b'_i [b'_i > 0] - \sum_{i=1}^p a'_i [a'_i < 0]$ , sauf encore dans certains cas rares.

On peut utiliser ces expressions pour démontrer directement que lorsque la valeur de  $l$  augmente, le degré de  $\hat{p}$  finit par devenir assez grand pour qu'un polynôme  $s(n, k)$  puisse exister, et que le nombre d'inconnues  $\alpha_j$  et  $\beta_j$  devient supérieur au nombre d'équations linéaires homogènes à résoudre. On obtient ainsi une seconde preuve de l'algorithme de Gosper-Zeilberger, si on considère, comme dans le texte du chapitre, qu'il doit exister une solution telle que les  $\beta_0(n), \dots, \beta_l(n)$  ne soient pas tous nuls.

**5.104** Soit  $t(n, k) = (-1)^k (r-s-k)! (r-2k)! / ((r-s-2k)! (r-n-k+1)! (n-k)! k!)$ . Alors  $\beta_0(n)t(n, k) + \beta_1(n)t(n+1, k)$  n'est pas sommable en termes hypergéométriques car  $\deg(\hat{p}) = 1$ ,  $\deg(q-r) = 3$ ,  $\deg(q+r) = 4$ ,  $\lambda = -8$ ,  $\lambda' = -4$ ; Par contre,  $\beta_0(n)t(n, k) + \beta_1(n)t(n+1, k) + \beta_2(n)t(n+2, k)$  l'est,

car  $\lambda' = 0$  lorsque  $q(n, k) = -(r-s-2k)(r-s-2k-1)(n+2-k)(r-n-k+1)$  et  $r(k) = (r-s-k+1)(r-2k+2)(r-2k+1)k$ . La solution est

$$\begin{aligned}\beta_0(n) &= (s-n)(r-n+1)(r-2n+1), \\ \beta_1(n) &= (rs-s^2-2rn+2n^2-2r+2n)(r-2n-1), \\ \beta_2(n) &= (s-r+n+1)(n+2)(r-2n-3), \\ \alpha_0(n) &= r-2n-1,\end{aligned}$$

et on en conclut que  $\beta_0(n)S_n + \beta_1(n)S_{n+1} + \beta_2(n)S_{n+2} = 0$ , où  $S_n$  est la somme donnée. Cela suffit pour prouver l'identité par induction, après avoir vérifié les cas  $n = 0$  et  $n = 1$ . Cependant,  $S_n$  satisfait aussi la récurrence plus simple  $\bar{\beta}_0(n)S_n + \bar{\beta}_1(n)S_{n+1} = 0$ , où  $\bar{\beta}_0(n) = (s-n)(r-2n+1)$  et  $\bar{\beta}_1(n) = -(n+1)(r-2n-1)$ . Pourquoi l'algorithme ne l'a-t-il pas trouvée ? Eh bien, ce n'est pas parce qu'on a une récurrence de ce genre que les termes  $\bar{\beta}_0(n)t(n, k) + \bar{\beta}_1(n)t(n+1, k)$  sont forcément indéfiniment sommables. En fait, ce qui est surprenant, c'est que l'algorithme trouve la récurrence la plus simple dans la grande majorité des cas. Remarquez que la récurrence du second ordre que nous avons trouvée peut se factoriser :  $\beta_0(n) + \beta_1(n)N + \beta_2(n)N^2 = ((r-n+1)N + (r-s-n-1))(\bar{\beta}_0(n) + \bar{\beta}_1(n)N)$ , où  $N$  désigne l'opérateur d'incrémentation de (5.145).

**5.105** Posons  $a = 1$  et comparons les coefficients de  $z^{3n}$  dans les deux membres de l'identité "monstrueuse et néanmoins sympathique" de Henrici,

$$\begin{aligned}f(a, z)f(a, \omega z)f(a, \omega^2 z) \\ = F\left(\frac{1}{3}a, \frac{1}{3}a+\frac{1}{3}, \frac{1}{3}a+\frac{2}{3}, \frac{2}{3}a-\frac{1}{3}, \frac{2}{3}a, \frac{2}{3}a+\frac{1}{3}, a \mid \left(\frac{4z}{9}\right)^3\right),\end{aligned}$$

où  $f(a, z) = F(1; a, 1; z)$ . On peut prouver cette identité en montrant que les deux membres satisfont la même équation différentielle.

Peter Paule a trouvé une autre manière intéressante de calculer la somme :

$$\begin{aligned}\sum_{k,l} \binom{N}{k, l, N-k-l}^2 \omega^{k+2l} &= \sum_{k,l} \binom{N}{k-l, l, N-k}^2 \omega^{k+l} \\ &= \sum_{k,l} \binom{N}{k}^2 \binom{k}{l}^2 \omega^{k+l} \\ &= \sum_k \binom{N}{k}^2 \omega^k [z^k] ((1+z)(\omega+z))^k \\ &= [z^0] \sum_k \binom{N}{k}^2 \left(\frac{\omega(1+z)(\omega+z)}{z}\right)^k\end{aligned}$$

$$\begin{aligned}
&= [z^0] \sum_{k,j} \binom{N}{k}^2 \binom{k}{j} \left( \frac{\omega(1+z)(\omega+z)}{z} - 1 \right)^j \\
&= [z^0] \sum_{k,j} \binom{N}{k} \binom{N-j}{N-k} \binom{N}{j} \left( \frac{(\omega z - 1)^2}{\omega z} \right)^j \\
&= \sum_j \binom{2N-j}{N} \binom{N}{j} [z^j] (z-1)^{2j} \\
&= \sum_j \binom{2N-j}{N} \binom{N}{j} \binom{2j}{j} (-1)^j,
\end{aligned}$$

en utilisant la formule du binôme, la convolution de Vandermonde, et le fait que  $[z^0]g(az) = [z^0]g(z)$ . On peut maintenant poser  $N = 3n$  et appliquer l'algorithme de Gosper-Zeilberger à cette somme  $S_n$  pour obtenir, comme par miracle, la récurrence d'ordre un  $(n+1)^2 S_{n+1} = 4(4n+1)(4n+3)S_n$ . Le résultat s'ensuit par induction.

Si on remplace  $3n$  par  $3n+1$  ou  $3n+2$ , la somme donnée vaut zéro. En effet,  $\sum_{k+l+m=N} t(k,l,m)\omega^{l-m}$  est toujours nul si  $N \bmod 3 \neq 0$  et  $t(k,l,m) = t(l,m,k)$ .

**5.106** (Solution de Shalosh B Ekhad). Soit

$$\begin{aligned}
T(r,j,k) &= \frac{(1+n+s)(1+r) - (1+n+r)j + (s-r)k(j-l)j}{(l-m+n-r+s)(n+r+1)(j-r-1)(j+k)} t(r,j,k); \\
U(r,j,k) &= \frac{(s+n+1)(k+l)k}{(l-m+n-r+s)(n+r+1)(j+k)} t(r,j,k).
\end{aligned}$$

L'égalité donnée dans l'exercice est facilement vérifiable, et on en déduit (5.32) en sommant sur  $j$  et  $k$ . Plus précisément, on somme  $T(r,j+1,k) - T(r,j,k)$  d'abord sur  $j$ , puis sur  $k$  ; on somme les autres termes d'abord sur  $k$ , puis sur  $j$ .

Il nous faut aussi vérifier (5.32) lorsque  $r = 0$ . Dans ce cas, l'identité devient  $\sum_k (-1)^k \binom{n}{n+l} \binom{n+l}{k+l} \binom{s+n-k}{m} = (-1)^l \binom{n}{n+l} \binom{s}{m-n-l}$  après application d'une transformation trinomiale. On suppose que  $l, m$  et  $n$  sont des entiers et que  $n \geq 0$ . Clairement, les deux membres sont nuls sauf si  $n+l \geq 0$ . Dans le cas contraire, on peut remplacer  $k$  par  $n-k$  et appliquer (5.24).

**5.107** Supposons qu'il est propre. Alors il existe un opérateur de différence linéaire qui l'annule. En d'autres termes, nous avons une identité de la forme

$$\sum_{i=0}^I \sum_{j=0}^J \alpha_{i,j}(n) / ((n+i)(k+j)+1) = 0,$$

Notez que  $1/nk$  est propre, car c'est  $(n-1)!(k-1)!/n!k!$ . De même,  $1/(n^2 - k^2)$  est propre. Par contre,  $1/(n^2 + k^2)$  ne l'est pas.

dans laquelle les  $\alpha$  sont des polynômes en  $n$  non tous nuls. Choisissons les entiers  $i$ ,  $j$  et  $n$  de sorte que  $n > 1$  et  $\alpha_{i,j}(n) \neq 0$ . Alors, lorsque  $k = -1/(n+i) - j$ , le terme  $(i, j)$  de la somme est infini alors que les autres termes sont finis.

**5.108** Remplacez  $k$  par  $m - k$  dans la somme double, puis utilisez (5.28) pour sommer sur  $k$ , obtenant ainsi

$$A_{m,n} = \sum_j \binom{m}{j}^2 \binom{m+n-j}{m}^2.$$

En effectuant une transformation trinomiale (5.21), on obtient alors l'une des formules désirées.

Il semble difficile de trouver une preuve directe de l'égalité des deux sommes symétriques. On peut toutefois prouver l'équation indirectement, avec l'algorithme de Gosper-Zeilberger, en montrant que les deux sommes satisfont la récurrence

$$(n+1)^3 A_{m,n} - f(m, n) A_{m,n+1} + (n+2)^3 A_{m,n+2} = 0,$$

où  $f(m, n) = (2n+3)(n^2 + 3n + 2m^2 + 2m + 3)$ . En posant  $t_1(n, k) = \binom{m}{k} \binom{n}{k} \binom{m+k}{k} \binom{n+k}{k}$  et  $t_2(n, k) = \binom{m+n-k}{k}^2 \binom{m+n-2k}{m-k}^2$ , on trouve que

$$\begin{aligned} (n+1)^2 t_j(n, k) - f(m, n) t_j(n+1, k) + (n+2)^2 t_j(n+2, k) \\ = T_j(n, k+1) - T_j(n, k), \end{aligned}$$

où  $T_1(n, k) = -2(2n+3)k^4 t_1(n, k)/(n+1-k)(n+2-k)$  et  $T_2(n, k) = -((n+2)(4mn+n+3m^2+8m+2)-2(3mn+n+m^2+6m+2)k+(2m+1)k^2)k^2(m+n+1-k)^2 t_2(n, k)/(n+2-k)^2$ . Ceci prouve la récurrence, et il ne reste plus qu'à vérifier l'égalité pour  $n = 0$  et  $n = 1$ . (Nous aurions aussi pu utiliser la récurrence plus simple

$$m^3 A_{m,n-1} - n^3 A_{m-1,n} = (m-n)(m^2 + n^2 - mn) A_{m-1,n-1},$$

qui peut être obtenue avec la méthode de l'exercice 101).

Le fait que la première expression de  $A_{m,n}$  est égale à la troisième entraîne une remarquable identité qui concerne les fonctions génératrices  $\sum_{m,n} A_{m,n} w^m z^n$ :

$$\sum_k \frac{w^k S_k(z)^2}{(1-z)^{2k+1}} = \sum_k \binom{2k}{k}^2 \frac{w^k}{(1-w)^{2k+1}} \frac{z^k}{(1-z)^{2k+1}},$$

où  $S_k(z) = \sum_j \binom{k}{j}^2 z^j$ . En fait, on sait que

$$\sum_k \frac{w^k S_k(x) S_k(y)}{(1-x)^k (1-y)^k} = \sum_k \binom{2k}{k} \frac{w^k}{(1-w)^{2k+1}} \frac{\sum_j \binom{k}{j}^2 x^j y^{k-j}}{(1-x)^k (1-y)^k}.$$

C'est un cas particulier d'une identité due à Bailey [19].

**5.109** Soit  $X_n = \sum_k \binom{n}{k}^{a_0} \binom{n+k}{k}^{a_1} \dots \binom{n+l_k}{k}^{a_l} x^k$  pour tous entiers strictement positifs  $a_0, a_1, \dots, a_l$  et tout entier  $x$ . Alors, si  $0 \leq m < p$ , on a

$$X_{m+pn} = \sum_{j=0}^{p-1} \sum_k \binom{m+pn}{j+pk}^{a_0} \dots \binom{m+pn+l(j+pk)}{j+pk}^{a_l} x^{j+pk},$$

$$X_m X_n = \sum_{j=0}^{p-1} \sum_k \binom{m}{j}^{a_0} \binom{n}{k}^{a_0} \dots \binom{m+l_j}{j}^{a_l} \binom{n+l_k}{k}^{a_l} x^{j+k}.$$

Les termes qui sont en correspondance sont congrus modulo  $p$ , car l'exercice 36 entraîne qu'ils sont multiples de  $p$  lorsque  $lj + m \geq p$ , l'exercice 61 entraîne que les binomiaux sont congrus lorsque  $lj + m < p$ , et (4.48) entraîne que  $x^p \equiv x$ .

**5.110** C'est vrai si  $2n + 1$  est premier. Steven Skiena a aussi trouvé l'exemple  $n = 2953$ , donc  $2n + 1 = 3 \cdot 11 \cdot 179$ .

**5.111** Voir [96] pour des résultats partiels. Les calculs sur ordinateurs ont été faits par V. A. Vyssotsky.

**5.112** Si  $n$  n'est pas une puissance de 2, on sait, d'après l'exercice 36, que  $\binom{2n}{n}$  est un multiple de 4. Sinon, la proposition a été vérifiée pour toutes les valeurs de  $n \leq 2^{22000}$  par A. Granville et O. Ramaré. Ces derniers ont aussi ravivé un théorème de Sárközy [317], en montrant que  $\binom{2n}{n}$  est divisible par le carré d'un nombre premier pour tout  $n > 2^{22000}$ . Ils confirmaient ainsi une vieille conjecture selon laquelle  $\binom{2n}{n}$  n'est jamais sans carré lorsque  $n > 4$ .

Voici des conjectures analogues pour les cubes :  $\binom{2n}{n}$  est divisible par le cube d'un nombre premier pour tout  $n > 1056$ , et par soit  $2^3$  soit  $3^3$  pour tout  $n > 2^{29} + 2^{23}$ . Ceci a été vérifié pour tout  $n < 2^{10000}$ . Paul Erdős conjecturait que, en fait,  $\max_p \epsilon_p(\binom{2n}{n})$  tend vers l'infini lorsque  $n \rightarrow \infty$  ; ceci serait vrai même si  $p$  ne peut prendre que les valeurs 2 et 3.

**5.113** Peut-être le théorème sur les fonctions génératrices de l'exercice 7.20 pourrait-il aider à résoudre cette conjecture.

**5.114** Strehl [344] a montré que  $c_n^{(2)} = \sum_k \binom{n}{k}^3 = \sum_k \binom{n}{k}^2 \binom{2k}{n}$  est ce qu'on appelle un nombre de Franel [132], et que  $c_n^{(3)} = \sum_k \binom{n}{k}^2 \binom{2k}{k}^2 \binom{2k}{n-k}$ .

Han Vardi remarque que c'est vrai aussi si  $2n + 1 = p^2$ , avec  $p$  premier, à condition que

$2^{p-1} \bmod p^2 = 1$  ;  
ce qui donne deux exemples de plus :  
 $n = (1093^2 - 1)/2$  ;  
 $n = (3511^2 - 1)/2$ .

D'autre part, H. S. Wilf a montré que  $c_n^{(m)}$  est un entier pour tout  $m$  si  $n \leq 9$ .

**6.1** 2314, 2431, 3241, 1342, 3124, 4132, 4213, 1423, 2143, 3412, 4321.

**6.2**  $\binom{n}{k} m^k$ , car chacune de ces applications partitionne l'ensemble de départ en  $k$  sous-ensembles non vides, et il y a  $m^k$  façons possibles d'assigner une valeur à chaque partition. Remarquez qu'en sommant sur  $k$ , on obtient une preuve combinatoire de (6.10).

**6.3** Dans ce cas,  $d_{k+1} \leq (\text{centre de gravité}) - \epsilon = 1 - \epsilon + (d_1 + \dots + d_k)/k$ . Cette récurrence est identique à (6.55), sauf que  $1 - \epsilon$  remplace 1. La solution optimale est donc  $d_{k+1} = (1 - \epsilon)H_k$ . Ceci n'est pas borné, sauf si  $\epsilon \geq 1$ .

**6.4**  $H_{2n+1} - \frac{1}{2}H_n$ . Voici une autre identité similaire :  $\sum_{k=1}^{2n} (-1)^{k-1}/k = H_{2n} - H_n$ .

**6.5**  $U_n(x, y)$  est égal à

$$x \sum_{k \geq 1} \binom{n}{k} (-1)^{k-1} k^{-1} (x + ky)^{n-1} + y \sum_{k \geq 1} \binom{n}{k} (-1)^{k-1} (x + ky)^{n-1}$$

et la première somme vaut

$$U_{n-1}(x, y) + \sum_{k \geq 1} \binom{n-1}{k-1} (-1)^{k-1} k^{-1} (x + ky)^{n-1}.$$

Le  $k^{-1}$  qui reste peut être absorbé, et on obtient

$$\begin{aligned} \sum_{k \geq 1} \binom{n}{k} (-1)^{k-1} (x + ky)^{n-1} &= x^{n-1} + \sum_{k \geq 0} \binom{n}{k} (-1)^{k-1} (x + ky)^{n-1} \\ &= x^{n-1}, \end{aligned}$$

ce qui prouve (6.75). Soit  $R_n(x, y) = x^{-n} U_n(x, y)$ . Alors  $R_0(x, y) = 0$  et  $R_n(x, y) = R_{n-1}(x, y) + 1/n + y/x$ , donc  $R_n(x, y) = H_n + ny/x$ .

Remarquez que la somme d'origine  $U_n = U_n(n, -1)$  ne permet pas d'arriver à une récurrence de ce type. Il est plus facile de résoudre par induction la somme plus générale, dans laquelle  $x$  est indépendant de  $n$ , que son cas particulier. Encore une fois, c'est la force de l'hypothèse d'induction qui fait la différence entre le succès et l'échec.

*La récurrence des Fibonacci est additive, alors que les lapins se multiplient.*

**6.6** Chaque couple de bébés **bb** nés à la fin d'un mois devient un couple d'adultes **aa** à la fin du mois suivant. Pendant le même temps, chaque couple **aa** reste un **aa** et donne naissance à un **bb**. On voit donc que les **bb** se comportent exactement comme les faux bourdons et les **aa** comme les reines dans l'arbre des abeilles. La seule différence, c'est que le temps s'écoule en sens inverse. Par conséquent, il y a  $F_{n+1}$  couples de lapins après  $n$  mois, dont  $F_n$  adultes et  $F_{n-1}$  bébés. C'est dans ce contexte que Fibonacci introduisit ses nombres.

**6.7** (a) Posez  $k = 1 - n$  puis appliquez (6.107). (b) Posez  $m = 1$ ,  $k = n - 1$ , puis appliquez (6.128).

**6.8**  $55 + 8 + 2$  se convertit en  $89 + 13 + 3 = 105$ , et la vraie valeur est 104,607361.

**6.9** 21, et la vraie réponse est à peu près 20,72. Lorsqu'il s'agit d'unités de surface, on convertit en passant de  $F_n$  à  $F_{n+2}$ .

**6.10** Les quotients partiels  $a_0, a_1, a_2, \dots$  sont tous égaux à 1 car  $\phi = 1 + 1/\phi$  (et donc la représentation de Stern–Brocot s'écrit RLRLRLRLRL...).

**6.11**  $(-1)^{\bar{n}} = [n=0] - [n=1]$ ; voir (6.11).

**6.12** C'est une conséquence de (6.31) et de son dual dans la table 280.

**6.13** D'après l'exercice 12, les deux formules sont équivalentes. On peut les démontrer par induction. On peut aussi observer qu'en appliquant  $z^n D^n$  à  $f(z) = z^x$  on obtient  $x^n z^x$  tandis qu'en appliquant  $\vartheta^n$  à la même fonction on obtient  $x^n z^x$ . Par conséquent, la suite  $\langle \vartheta^0, \vartheta^1, \vartheta^2, \dots \rangle$  est liée à la suite  $\langle z^0 D^0, z^1 D^1, z^2 D^2, \dots \rangle$ , tout comme la suite  $\langle x^0, x^1, x^2, \dots \rangle$  est liée à la suite  $\langle x^0, x^1, x^2, \dots \rangle$ .

**6.14** On a

$$x \binom{x+k}{n} = (k+1) \binom{x+k}{n+1} + (n-k) \binom{x+k+1}{n+1},$$

car  $(n+1)x = (k+1)(x+k-n) + (n-k)(x+k+1)$  (il suffit de vérifier cette dernière identité pour  $k=0$ ,  $k=-1$  et  $k=n$ ).

**6.15** Comme  $\Delta(\binom{x+k}{n}) = \binom{x+k}{n-1}$ , on a la formule générale

$$\sum_k \binom{n}{k} \binom{x+k}{n-m} = \Delta^m(x^n) = \sum_j \binom{m}{j} (-1)^{m-j} (x+j)^n.$$

Posez  $x=0$  et faites appel à (6.16).

**6.16**  $A_{n,k} = \sum_{j \geq 0} a_j \binom{n-j}{k}$ . Cette somme est toujours finie.

**6.17** (a)  $\binom{n}{k} = \binom{n+1}{n+1-k}$ . (b)  $\binom{n}{k} = n \frac{n-k}{k!} = n! [n \geq k]/k!$ . (c)  $\binom{n}{k} = k! \binom{n}{k}$ .

**6.18** C'est équivalent à (6.3) et à (6.8).

**6.19** Servez-vous de la table 289.

**6.20**  $\sum_{1 \leq j \leq k \leq n} 1/j^2 = \sum_{1 \leq j \leq n} (n+1-j)/j^2 = (n+1)H_n^{(2)} - H_n$ .

**6.21** Le nombre suggéré est une somme de fractions ayant des dénominateurs impairs, donc il est de la forme  $a/b$  où  $a$  et  $b$  sont impairs. Notons en passant que le postulat de Bertrand implique que, si  $n > 2$ , alors  $b_n$  est aussi divisible par au moins un nombre premier impair.

Cette "vraie valeur" est la longueur de 65 miles internationaux. Or, un mile international ne vaut que 0,999998 mile américain légal. Il y a exactement 6336 kilomètres dans 3937 miles américains légaux, et la méthode de Fibonacci convertit 3937 en 6370.

**6.22**  $|z/k(k+z)| \leq 2|z|/k^2$  lorsque  $k > 2|z|$ , donc la somme est bien définie lorsque les dénominateurs sont non nuls. Si  $z = n$ , on a  $\sum_{k=1}^m (1/k - 1/(k+n)) = H_m - H_{m+n} + H_n$ , ce qui est très proche de  $H_n$  lorsque  $m \rightarrow \infty$ . (La quantité  $H_{z-1} - \gamma$  est souvent appelée la “fonction psi”  $\psi(z)$ ).

$$\text{6.23 } z/(e^z + 1) = z/(e^z - 1) - 2z/(e^{2z} - 1) = \sum_{n \geq 0} (1 - 2^n) B_n z^n / n!.$$

**6.24** Lorsque  $n$  est impair,  $T_n(x)$  est un polynôme en  $x^2$ . Par conséquent ses coefficients se trouvent multipliés par des nombres pairs lorsqu'on le dérive et qu'on calcule  $T_{n+1}(x)$  selon (6.95). On peut même prouver plus encore : d'après l'exercice 54, le nombre de Bernoulli  $B_{2n}$  a toujours un 2 dans son dénominateur. Par conséquent,  $2^{2n-k} \nmid T_{2n+1} \iff 2^k \nmid (n+1)$ . Les entiers positifs impairs  $(n+1)T_{2n+1}/2^{2n}$  sont appelés les nombres de Genocchi [145] :  $\langle 1, 1, 3, 17, 155, 2073, \dots \rangle$ .

**6.25**  $100n - nH_n < 100(n-1) - (n-1)H_{n-1} \iff H_{n-1} > 99$ . Le plus petit  $n$  possible est à peu près égal à  $e^{99-\gamma}$ , alors que le ver arrivera à  $N \approx e^{100-\gamma}$ , au bout d'un temps à peu près  $e$  fois plus long. Il se rapprochera donc du but pendant les derniers 63% de son périple.

**6.26** Soient  $u(k) = H_{k-1}$  et  $\Delta v(k) = 1/k$ , de sorte que  $u(k) = v(k)$ . Alors  $S_n - H_n^{(2)} = \sum_{k=1}^n H_{k-1}/k = H_{k-1}^2 |_{n+1} - S_n = H_n^2 - S_n$ .

**6.27** Remarquez que lorsque  $m > n$ , on a, d'après (6.108),  $\text{pgcd}(F_m, F_n) = \text{pgcd}(F_{m-n}, F_n)$ . On en déduit une preuve par induction.

**6.28** (a)  $Q_n = \alpha(L_n - F_n)/2 + \beta F_n$  (ou, de façon équivalente,  $Q_n = \alpha F_{n-1} + \beta F_n$ ). (b)  $L_n = \Phi^n + \bar{\Phi}^n$ .

**6.29** Lorsque  $k = 0$ , l'identité se ramène à (6.133). Lorsque  $k = 1$ , elle se ramène à

$$\begin{aligned} K(x_1, \dots, x_n) x_m &= K(x_1, \dots, x_m) K(x_m, \dots, x_n) \\ &\quad - K(x_1, \dots, x_{m-2}) K(x_{m+2}, \dots, x_n). \end{aligned}$$

Selon l'interprétation par le code de Morse, le second produit du membre droit soustrait les cas où le premier produit donne lieu à des traits qui se recouvrent. Pour  $k > 1$ , une preuve par induction suffit, en utilisant (6.127) et (6.132). Notons que l'identité est vraie aussi si un ou plusieurs indices de  $K$  sont égaux à  $-1$ , pourvu que l'on convienne que  $K_{-1} = 0$ . Si on suppose que la multiplication n'est pas commutative, l'identité d'Euler reste valable pour  $k = n-1$  lorsqu'on l'exprime sous la forme

$$\begin{aligned} K_{m+n}(x_1, \dots, x_{m+n}) K_{n-1}(x_{m+n-1}, \dots, x_{m+1}) \\ = K_{m+n-1}(x_1, \dots, x_{m+n-1}) K_n(x_{m+n}, \dots, x_{m+1}) \\ - (-1)^n K_{m-1}(x_1, \dots, x_{m-1}). \end{aligned}$$

On obtient ainsi des factorisations quelque peu surprenantes, comme

$$(abc + a + c)(1 + ba) = (ab + 1)(cba + a + c)$$

pour le cas  $m = 0, n = 3$ .

**6.30** La dérivée de  $K(x_1, \dots, x_n)$  par rapport à  $x_m$  est égale à

$$K(x_1, \dots, x_{m-1}) K(x_{m+1}, \dots, x_n),$$

et la dérivée seconde est nulle. Voici par conséquent la réponse :

$$K(x_1, \dots, x_n) + K(x_1, \dots, x_{m-1}) K(x_{m+1}, \dots, x_n) y.$$

**6.31** Comme  $x^n = (x + n - 1)^n = \sum_k \binom{n}{k} x^k (n-1)^{n-k}$ , on a  $\left| \binom{n}{k} \right| = \binom{n}{k} (n-1)^{n-k}$ . Remarquons que ces coefficients satisfont la récurrence  $\left| \binom{n}{k} \right| = \left| \binom{-k}{-n} \right|$ .

$$\left| \binom{n}{k} \right| = (n-1+k) \left| \binom{n-1}{k} \right| + \left| \binom{n-1}{k-1} \right|, \quad n, k > 0 \text{ entiers.}$$

**6.32**  $\sum_{k \leq m} k \left\{ \begin{smallmatrix} n+k \\ k \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} m+n+1 \\ m \end{smallmatrix} \right\}$  et  $\sum_{0 \leq k \leq n} \left\{ \begin{smallmatrix} k \\ m \end{smallmatrix} \right\} (m+1)^{n-k} = \left\{ \begin{smallmatrix} n+1 \\ m+1 \end{smallmatrix} \right\}$ . Ces deux formules apparaissent dans la table 281.

**6.33** Si  $n > 0$ , on a  $\left[ \begin{smallmatrix} n \\ 3 \end{smallmatrix} \right] = \frac{1}{2}(n-1)!(H_{n-1}^2 - H_{n-1}^{(2)})$ , d'après (6.71) ; et  $\left\{ \begin{smallmatrix} n \\ 3 \end{smallmatrix} \right\} = \frac{1}{6}(3^n - 3 \cdot 2^n + 3)$ , d'après (6.16).

**6.34** On a  $\langle \begin{smallmatrix} -1 \\ k \end{smallmatrix} \rangle = 1/(k+1)$ ,  $\langle \begin{smallmatrix} -2 \\ k \end{smallmatrix} \rangle = H_{k+1}^{(2)}$ , et, plus généralement,  $\langle \begin{smallmatrix} n \\ k \end{smallmatrix} \rangle$  est donné par (6.38) pour tout entier  $n$ .

**6.35** Prendre le plus petit entier  $n > 1/\epsilon$  tel que  $\lfloor H_n \rfloor > \lfloor H_{n-1} \rfloor$ .

**6.36** Ici,  $d_{k+1} = (100 + (1+d_1) + \dots + (1+d_k))/(100+k)$ , et la solution est  $d_{k+1} = H_{k+100} - H_{101} + 1$  pour  $k \geq 1$ . C'est supérieur à 2 lorsque  $k \geq 176$ .

**6.37** En sommant par parties, on trouve  $H_{mn} - \left( \frac{m}{m} + \frac{m}{2m} + \dots + \frac{m}{mn} \right) = H_{mn} - H_n$ . La somme infinie vaut donc  $\ln m$ . Remarquez qu'on peut en déduire que

$$\sum_{k \geq 1} \frac{\nu_m(k)}{k(k+1)} = \frac{m}{m-1} \ln m,$$

car  $\nu_m(k) = (m-1) \sum_{j \geq 1} (k \bmod m^j)/m^j$ .

**6.38**  $(-1)^k ((r-1 \choose k)r^{-1} - (r-1 \choose k-1)H_k) + C$  (sommation par parties, en utilisant (5.16)).

**6.39** Réécrivez la somme en  $\sum_{1 \leq j \leq n} j^{-1} \sum_{j \leq k \leq n} H_k$  et sommez d'abord sur  $k$  avec (6.67), pour obtenir

$$(n+1)H_n^2 - (2n+1)H_n + 2n.$$

**6.40** Si  $6n-1$  est premier, le numérateur de

$$\sum_{k=1}^{4n-1} \frac{(-1)^{k-1}}{k} = H_{4n-1} - H_{2n-1}$$

est divisible par  $6n-1$ , car la somme est égale à

$$\sum_{k=2n}^{4n-1} \frac{1}{k} = \sum_{k=2n}^{3n-1} \left( \frac{1}{k} + \frac{1}{6n-1-k} \right) = \sum_{k=2n}^{3n-1} \frac{6n-1}{k(6n-1-k)}.$$

De même, si  $6n+1$  est premier, le numérateur de  $\sum_{k=1}^{4n} (-1)^{k-1}/k = H_{4n} - H_{2n}$  est un multiple de  $6n+1$ . Pour 1987, on somme jusqu'à  $k=1324$ .

**6.41**  $S_{n+1} = \sum_k \binom{\lfloor (n+1+k)/2 \rfloor}{k} = \sum_k \binom{\lfloor (n+k)/2 \rfloor}{k-1}$ , donc  $S_{n+1} + S_n = \sum_k \binom{\lfloor (n+k)/2+1 \rfloor}{k} = S_{n+2}$ . La réponse est  $F_{n+2}$ .

**6.42**  $F_n$ .

**6.43** Posez  $z = \frac{1}{10}$  dans  $\sum_{n \geq 0} F_n z^n = z/(1-z-z^2)$  pour obtenir  $\frac{10}{89}$ . La représentation décimale de la somme est de période 44 :

$$0,11235\,95505\,61797\,75280\,89887\,64044\,94382\,02247\,19101\,12359\,55+.$$

**6.44** Remplacez  $(m, k)$  par  $(-m, -k)$ , ou  $(k, -m)$ , ou encore  $(-k, m)$ , si nécessaire, pour que  $m \geq k \geq 0$ . Le résultat est évident si  $m = k$ . Si  $m > k$ , on peut remplacer  $(m, k)$  par  $(m - k, m)$  et procéder par induction.

**6.45**  $X_n = A(n)\alpha + B(n)\beta + C(n)\gamma + D(n)\delta$ , avec  $B(n) = F_n$ ,  $A(n) = F_{n-1}$ ,  $A(n) + B(n) - D(n) = 1$  et  $B(n) - C(n) + 3D(n) = n$ .

**6.46**  $\phi/2$  et  $\phi^{-1}/2$ . Soient  $u = \cos 72^\circ$  et  $v = \cos 36^\circ$ ; alors  $u = 2v^2 - 1$  et  $v = 1 - 2\sin^2 18^\circ = 1 - 2u^2$ . Par conséquent,  $u + v = 2(u + v)(v - u)$  et  $4v^2 - 2v - 1 = 0$ . On peut poursuivre cette investigation pour trouver les cinq racines 5ièmes de l'unité :

$$1, \quad \frac{\phi^{-1} \pm i\sqrt{2+\phi}}{2}, \quad \frac{-\phi \pm i\sqrt{3-\phi}}{2}.$$

**6.47**  $2^n \sqrt{5} F_n = (1 + \sqrt{5})^n - (1 - \sqrt{5})^n$ , et les puissances paires de  $\sqrt{5}$  s'annulent. Soit maintenant  $p$  un nombre premier impair. Alors  $\binom{p}{2k+1} \equiv 0$  sauf lorsque  $k = (p-1)/2$ , et  $\binom{p+1}{2k+1} \equiv 0$  sauf lorsque  $k = 0$  ou  $k = (p-1)/2$ . Par conséquent,  $F_p \equiv 5^{(p-1)/2}$  et  $2F_{p+1} \equiv 1 + 5^{(p-1)/2} \pmod{p}$ . On peut montrer que  $5^{(p-1)/2} \equiv 1$  lorsque  $p$  est de la forme  $10k \pm 1$ , et que  $5^{(p-1)/2} \equiv -1$  lorsque  $p$  est de la forme  $10k \pm 3$ .

**6.48** Soit  $K_{i,j} = K_{j-i+1}(x_i, \dots, x_j)$ . Si on applique plusieurs fois (6.133), les deux membres donnent  $(K_{1,m-2}(x_{m-1} + x_{m+1}) + K_{1,m-3})K_{m+2,n} + K_{1,m-2}K_{m+3,n}$ .

**6.49** Posez  $z = \frac{1}{2}$  dans (6.146) pour trouver les quotients partiels 0,  $2^{F_0}$ ,  $2^{F_1}, 2^{F_2}, \dots$  (Knuth [206] fait remarquer que ce nombre est transcendant).

**6.50** (a)  $f(n)$  est pair si et seulement si  $3 \nmid n$ . (b) Si la représentation binaire de  $n$  s'écrit  $(1^{a_1}0^{a_2}\dots1^{a_{m-1}}0^{a_m})_2$ , où  $m$  est pair, alors  $f(n) = K(a_1, a_2, \dots, a_{m-1})$ .

**6.51** (a) En voici une preuve combinatoire : l'ensemble des décompositions de  $\{1, 2, \dots, p\}$  en  $k$  sous-ensembles ou cycles peut être subdivisé en "orbites", en ajoutant 1 modulo  $p$  à chaque élément. Par exemple,

$$\begin{aligned} \{1, 2, 4\} \cup \{3, 5\} &\rightarrow \{2, 3, 5\} \cup \{4, 1\} \rightarrow \{3, 4, 1\} \cup \{5, 2\} \\ &\rightarrow \{4, 5, 2\} \cup \{1, 3\} \rightarrow \{5, 1, 3\} \cup \{2, 4\} \rightarrow \{1, 2, 4\} \cup \{3, 5\}. \end{aligned}$$

Chaque orbite contient soit une, soit  $p$  décompositions. Une orbite ne peut être de taille 1 que si la transformation décrite ne modifie pas la décomposition ; cela n'est possible que si  $k = 1$  ou  $k = p$ . Il y a aussi une preuve algébrique :  $x^p \equiv x^p + x^1$  et  $x^p \equiv x^p - x \pmod{p}$ , car, d'après le théorème de Fermat,  $x^p - x$  est divisible par  $(x - 0)(x - 1)\dots(x - (p-1))$ .

(b) C'est une conséquence de (a) et du théorème de Wilson. On peut aussi appliquer  $x^{p-1} \equiv x^p/(x-1) \equiv (x^p - x)/(x-1) = x^{p-1} + x^{p-2} + \dots + x$ .

(c) On a  $\binom{p+1}{k} \equiv \binom{p+1}{k} \equiv 0$  pour  $3 \leq k \leq p$ , donc  $\binom{p+2}{k} \equiv \binom{p+2}{p} \equiv 0$  pour  $4 \leq k \leq p$ , etc (de même, on a  $\binom{2p-1}{p} \equiv -\binom{2p-1}{p} \equiv 1$ ).

(d)  $p! = p^p = \sum_k (-1)^{p-k} p^k \binom{p}{k} = p^p \binom{p}{p} - p^{p-1} \binom{p}{p-1} + \dots + p^3 \binom{p}{3} - p^2 \binom{p}{2} + p \binom{p}{1}$ . Or  $p \binom{p}{1} = p!$ , donc

$$\binom{p}{2} = p \binom{p}{3} - p^2 \binom{p}{4} + \dots + p^{p-2} \binom{p}{p}$$

est un multiple de  $p^2$ . Ce résultat est appelé le théorème de Wolstenholme.

**6.52** (a) Remarquez que  $H_n = H_n^* + H_{\lfloor n/p \rfloor}/p$ , avec  $H_n^* = \sum_{k=1}^n [k \perp p]/k$ .

(b) En travaillant mod 5, on trouve  $H_r = \langle 0, 1, 4, 1, 0 \rangle$  pour  $0 \leq r \leq 4$ . La première solution est donc  $n = 4$ . D'après la partie (a), nous savons que  $5 \nmid a_n \implies 5 \nmid a_{\lfloor n/5 \rfloor}$  ; les prochaines valeurs possibles sont donc de la forme  $n = 20 + r$  pour  $0 \leq r \leq 4$ , lorsque  $H_n = H_n^* + \frac{1}{5}H_4 = H_{20}^* + \frac{1}{5}H_4 + H_r + \sum_{k=1}^r 20/k(20+k)$ . Le numérateur de  $H_{20}^*$ , comme celui de  $H_4$ , est divisible par 25. Les seules solutions de cet intervalle sont donc  $n = 20$  et  $n = 24$ . Les prochaines valeurs possibles sont de la forme  $n = 100 + r$  ; maintenant,  $H_n = H_n^* + \frac{1}{5}H_{20}$ , qui vaut  $\frac{1}{5}H_{20} + H_r$  plus une fraction dont le numérateur est un multiple de 5. Si  $\frac{1}{5}H_{20} \equiv m \pmod{5}$ , où  $m$  est un

entier, le numérateur du nombre harmonique  $H_{100+r}$  sera divisible par 5 si et seulement si  $m + H_r \equiv 0 \pmod{5}$  ; donc  $m$  est congru à 0, 1, ou 4. En travaillant modulo 5, on trouve  $\frac{1}{5}H_{20} = \frac{1}{5}H_{20}^* + \frac{1}{25}H_4 \equiv \frac{1}{25}H_4 = \frac{1}{12} \equiv 3$  ; donc il n'y a pas de solution pour  $100 \leq n \leq 104$ . Il n'y en a pas non plus pour  $120 \leq n \leq 124$ , et nous avons trouvé les trois seules solutions.

Remarques : d'après l'exercice 6.51(d), on a forcément  $p^2 \nmid a_{p-1}$ ,  $p \nmid a_{p^2-p}$  et  $p \nmid a_{p^2-1}$  si  $p$  est un nombre premier  $\geq 5$ . L'argument que nous venons de donner montre que ce sont les seules solutions de  $p \nmid a_n$  si et seulement s'il n'y a pas de solution à  $p^{-2}H_{p-1} + H_r \equiv 0 \pmod{p}$  pour  $0 \leq r < p$ . Cette dernière condition est vraie non seulement pour  $p = 5$ , mais aussi pour  $p = 13, 17, 23, 41$  et  $67$ , et peut-être pour une infinité de nombres premiers. Le numérateur de  $H_n$  n'est divisible par 3 que si  $n = 2, 7$  ou  $22$  ; il n'est divisible par 7 que si  $n = 6, 42, 48, 295, 299, 337, 341, 2096, 2390, 14675, 16731, 16735$  ou  $102728$ . Voyez la réponse de l'exercice 92.

### 6.53 En sommant par parties, on trouve

$$\frac{n+1}{(n+2)^2} \left( \frac{(-1)^m}{\binom{n+1}{m+1}} ((n+2)H_{m+1} - 1) - 1 \right).$$

**6.54** (a) Si  $m \geq p$ , alors  $S_m(p) \equiv S_{m-(p-1)}(p) \pmod{p}$  car  $k^{p-1} \equiv 1$  lorsque  $1 \leq k < p$ . On a aussi  $S_{p-1}(p) \equiv p-1 \equiv -1$ . Si  $0 < m < p-1$ , on peut écrire

$$S_m(p) = \sum_{j=0}^m \begin{bmatrix} m \\ j \end{bmatrix} (-1)^{m-j} \sum_{k=0}^{p-1} k^j = \sum_{j=0}^m \begin{bmatrix} m \\ j \end{bmatrix} (-1)^{m-j} \frac{p^{j+1}}{j+1} \equiv 0.$$

(b) La condition suggérée implique que le dénominateur de  $I_{2n}$  n'est divisible par aucun nombre premier  $p$  ; donc  $I_{2n}$  est un entier. Pour prouver la proposition suggérée, nous pouvons supposer que  $n > 1$ . Alors

$$B_{2n} + \frac{[(p-1)\setminus(2n)]}{p} + \sum_{k=0}^{2n-2} \binom{2n+1}{k} B_k \frac{p^{2n-k}}{2n+1}$$

est un entier d'après (6.78), (6.84) et la partie (a). Il nous faut donc vérifier qu'aucune des fractions  $\binom{2n+1}{k} B_k p^{2n-k}/(2n+1) = \binom{2n}{k} B_k p^{2n-k}/(2n-k+1)$  n'a un dénominateur divisible par  $p$ . Le dénominateur de  $\binom{2n}{k} B_k p$  n'est pas divisible par  $p$  car  $B_k$  n'a pas de  $p^2$  dans son dénominateur (par induction) ; le dénominateur  $p^{2n-k-1}/(2n-k+1)$  n'est pas non plus divisible par  $p$ , car  $2n-k+1 < p^{2n-k}$  lorsque  $k \leq 2n-2$  ; CQFD. Remarques : on trouve des tables des nombres  $I_{2n}$  dans [224]. Hermite les calcula jusqu'à  $I_{18}$  en 1875 [184]. On sait par exemple que  $I_2 = I_4 = I_6 = I_8 = I_{10} = I_{12} = 1$  ; par conséquent, il existe en fait une expression "simple" des premiers nombres de Bernoulli , y compris  $\frac{-691}{2730}(!)$ . Cependant, les nombres  $I_{2n}$  ne

(A l'attention des programmeurs : voici une condition à tester sur tous les nombres premiers que vous pouvez).

(Les numérateurs des nombres de Bernoulli jouèrent un rôle important dans les toutes premières études du Dernier Théorème de Fermat ; voir Ribenboim [308].)

semblent plus rien avoir de remarquable lorsque  $2n > 12$ . Par exemple,  $B_{24} = -86579 - \frac{1}{2} - \frac{1}{3} - \frac{1}{5} - \frac{1}{7} - \frac{1}{13}$ , et 86579 est premier.

(c) Les nombres  $2-1$  et  $3-1$  divisent toujours  $2n$ . Si  $n$  est premier, les seuls diviseurs de  $2n$  sont 1, 2,  $n$  et  $2n$ , donc le dénominateur de  $B_{2n}$  pour  $n > 2$  premier sera égal à 6, sauf si  $2n+1$  est premier aussi. Dans ce dernier cas, on peut essayer  $4n+3, 8n+7, \dots$ , pour finalement arriver à coup sûr à un nombre non premier (car  $n$  divise  $2^{n-1}n + 2^{n-1} - 1$ ). (Cette preuve ne nécessite pas l'utilisation du théorème, plus difficile, selon lequel il existe une infinité de nombres premiers de la forme  $6k+1$ ). Le dénominateur de  $B_{2n}$  peut être égal à 6 même si  $n$  n'est pas premier, par exemple si  $n = 49$ .

**6.55** D'après la convolution de Vandermonde, la somme donnée égale  $\frac{m+1}{x+m+1} \binom{x+n}{n} \binom{n}{m+1}$ . Pour obtenir (6.70), dérivez et posez  $x = 0$ .

**6.56** Remplacez  $k^{n+1}$  par  $((k-m)+m)^{n+1}$  et développez en puissances de  $k-m$ ; des simplifications se font, comme dans le calcul de (6.72). Si  $m > n$  ou  $m < 0$ , la réponse est  $(-1)^n n! - m^n / \binom{n-m}{n}$ . Sinon, il faut prendre la limite de (5.41) moins le terme correspondant à  $k = m$ , lorsque  $x \rightarrow -m$ ; la réponse est alors  $(-1)^n n! + (-1)^{m+1} \binom{n}{m} m^n (n+1+mH_{n-m} - mH_m)$ .

**6.57** Prouvez par induction que la  $n$ ième ligne contient au plus trois valeurs distinctes  $A_n \geq B_n \geq C_n$ ; si  $n$  est pair, elles apparaissent dans l'ordre cyclique  $[C_n, B_n, A_n, B_n, C_n]$ ; si  $n$  est impair, elles apparaissent dans l'ordre cyclique  $[C_n, B_n, A_n, A_n, B_n]$ . On a aussi

$$\begin{aligned} A_{2n+1} &= A_{2n} + B_{2n}; & A_{2n} &= 2A_{2n-1}; \\ B_{2n+1} &= B_{2n} + C_{2n}; & B_{2n} &= A_{2n-1} + B_{2n-1}; \\ C_{2n+1} &= 2C_{2n}; & C_{2n} &= B_{2n-1} + C_{2n-1}. \end{aligned}$$

Il s'ensuit que  $Q_n = A_n - C_n = F_{n+1}$ . On peut voir dans l'exercice 5.75 des coefficients binomiaux enroulés d'ordre 3.

**6.58** (a)  $\sum_{n \geq 0} F_n^2 z^n = z(1-z)/(1+z)(1-3z+z^2) = \frac{1}{5}((2-3z)/(1-3z+z^2) - 2/(1+z))$  (élevez au carré la formule de Binet (6.123) et sommez sur  $n$ , puis combinez les termes de façon que  $\phi$  et  $\bar{\phi}$  disparaissent). (b) De façon similaire,

$$\sum_{n \geq 0} F_n^3 z^n = \frac{z(1-2z-z^2)}{(1-4z-z^2)(1+z-z^2)} = \frac{1}{5} \left( \frac{2z}{1-4z-z^2} + \frac{3z}{1+z-z^2} \right).$$

Il s'ensuit que  $F_{n+1}^3 - 4F_n^3 - F_{n-1}^3 = 3(-1)^n F_n$ . Remarque : la récurrence correspondante pour les puissances mièmes fait apparaître les coefficients Fibonomiaux qu'on peut voir dans l'exercice 86 ; ils furent découverts par Jarden et Motzkin [194].

**6.59** Soit  $m$  fixé. On peut prouver par induction sur  $n$  qu'il est en fait possible de trouver un tel  $x$  avec la condition supplémentaire  $x \not\equiv 2 \pmod{4}$ . Si  $x$  est une telle solution, on peut se ramener à une solution modulo  $3^{n+1}$  du fait que

$$F_{8 \cdot 3^{n-1}} \equiv 3^n, \quad F_{8 \cdot 3^{n-1}-1} \equiv 3^n + 1 \pmod{3^{n+1}}.$$

Par conséquent, soit  $x$ , soit  $x + 8 \cdot 3^{n-1}$ , soit  $x + 16 \cdot 3^{n-1}$  conviendra.

**6.60** Les seuls cas possibles sont  $F_1 + 1$ ,  $F_2 + 1$ ,  $F_3 + 1$ ,  $F_4 - 1$  et  $F_6 - 1$ . Dans tous les autres cas, il existe une factorisation par les nombres de Lucas (vus dans l'exercice 28) :

$$\begin{aligned} F_{2m} + (-1)^m &= L_{m+1} F_{m-1}; & F_{2m+1} + (-1)^m &= L_m F_{m+1}; \\ F_{2m} - (-1)^m &= L_{m-1} F_{m+1}; & F_{2m+1} - (-1)^m &= L_{m+1} F_m. \end{aligned}$$

Plus généralement,  $F_{m+n} - (-1)^n F_{m-n} = L_m F_n$ .

**6.61**  $1/F_{2m} = F_{m-1}/F_m - F_{2m-1}/F_{2m}$  lorsque  $m$  est pair et strictement positif. La seconde somme vaut  $5/4 - F_{3 \cdot 2^n-1}/F_{3 \cdot 2^n}$  pour  $n \geq 1$ .

**6.62** (a)  $A_n = \sqrt{5} A_{n-1} - A_{n-2}$  et  $B_n = \sqrt{5} B_{n-1} - B_{n-2}$ . Notons en passant qu'on a aussi  $\sqrt{5} A_n + B_n = 2A_{n+1}$  et  $\sqrt{5} B_n - A_n = 2B_{n-1}$ . (b) En observant les premières valeurs, on découvre que

$$A_n = \begin{cases} L_n, & n \text{ pair;} \\ \sqrt{5} F_n, & n \text{ impair;} \end{cases} \quad B_n = \begin{cases} \sqrt{5} F_n, & n \text{ pair;} \\ L_n, & n \text{ impair.} \end{cases}$$

(c)  $B_n/A_{n+1} - B_{n-1}/A_n = 1/(F_{2n+1} + 1)$  car  $B_n A_n - B_{n-1} A_{n+1} = \sqrt{5}$  et  $A_n A_{n+1} = \sqrt{5}(F_{2n+1} + 1)$ . Remarquez que  $B_n/A_{n+1} = (F_n/F_{n+1})[n \text{ pair}] + (L_n/L_{n+1})[n \text{ impair}]$ . (d) De même,  $\sum_{k=1}^n 1/(F_{2k+1} - 1) = (A_0/B_1 - A_1/B_2) + \dots + (A_{n-1}/B_n - A_n/B_{n+1}) = 2 - A_n/B_{n+1}$ . La même quantité peut aussi s'écrire  $(5F_n/L_{n+1})[n \text{ pair}] + (L_n/F_{n+1})[n \text{ impair}]$ .

**6.63** (a)  $\binom{n}{k}$ . Il y en a  $\binom{n-1}{k-1}$  telles que  $\pi_n = n$  et  $(n-1)\binom{n-1}{k}$  telles que  $\pi_n < n$ . (b)  $\binom{n}{k}$ . Chaque permutation  $\rho_1 \dots \rho_{n-1}$  de  $\{1, \dots, n-1\}$  donne  $n$  permutations  $\pi_1 \pi_2 \dots \pi_n = \rho_1 \dots \rho_{j-1} n \rho_{j+1} \dots \rho_{n-1} \rho_j$ . Si  $\rho_1 \dots \rho_{n-1}$  contient  $k$  excédents, il y a  $k+1$  valeurs de  $j$  qui donnent  $k$  excédents dans  $\pi_1 \pi_2 \dots \pi_n$ ; les  $n-1-k$  valeurs qui restent en donnent  $k+1$ . Par conséquent, le nombre total de façons d'obtenir  $k$  excédents dans  $\pi_1 \pi_2 \dots \pi_n$  est  $(k+1)\binom{n-1}{k} + ((n-1)-(k-1))\binom{n-1}{k-1} = \binom{n}{k}$ .

**6.64** Le dénominateur de  $\binom{1/2}{2n}$  est égal à  $2^{4n-\nu_2(n)}$ , d'après la preuve de l'exercice 5.72. Le dénominateur de  $\binom{1/2}{1/2-n}$  est le même, selon (6.44), car  $\langle\langle n \rangle\rangle = 1$  et  $\langle\langle n \rangle\rangle$  est pair pour tout  $k > 0$ .

**6.65** Cela revient à dire que  $\binom{n}{k}/n!$  est la probabilité d'avoir  $\lfloor x_1 + \dots + x_n \rfloor = k$ , où  $x_1, \dots, x_n$  sont des nombres aléatoires indépendants et uniformément distribués entre 0 et 1. Soit  $y_j = (x_1 + \dots + x_j) \bmod 1$ . Alors  $y_1, \dots, y_n$  sont indépendants et uniformément distribués, et  $\lfloor x_1 + \dots + x_n \rfloor$  compte le nombre de descentes dans les  $y$ . La permutation des  $y$  est aléatoire, et la probabilité d'avoir  $k$  descentes est la même que celle d'avoir  $k$  montées.

**6.66**  $2^{n+1}(2^{n+1}-1)B_{n+1}/(n+1)$ , si  $n > 0$ . (Voir (7.57) et (6.92) ; les nombres cherchés sont les coefficients de  $1 - \tanh z$ ).

**6.67** D'après (6.3) et (6.40),  $\sum_k \left\{ \begin{smallmatrix} n \\ k+1 \end{smallmatrix} \right\} (k+1)! + \left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} k! \binom{n-k}{n-m} (-1)^{m-k} = \sum_k \left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} k! (-1)^{m-k} \left( \binom{n-k}{n-m} - \binom{n+1-k}{n-m} \right) = \sum_k \left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} k! (-1)^{m+1-k} \binom{n-k}{n-m-1} = \left\langle \begin{smallmatrix} n \\ n-m-1 \end{smallmatrix} \right\rangle$ . Utilisez alors (6.34). Il existe aussi une interprétation combinatoire de cette identité [59]).

**6.68** On a la formule générale

$$\left\langle \begin{smallmatrix} n \\ m \end{smallmatrix} \right\rangle = \sum_{k=0}^m \binom{2n+1}{k} \left\{ \begin{smallmatrix} n+m+1-k \\ m+1-k \end{smallmatrix} \right\} (-1)^k, \text{ pour } n > m \geq 0,$$

analogue à (6.38). Lorsque  $m = 2$ , ceci est égal à

$$\begin{aligned} \left\langle \begin{smallmatrix} n \\ 2 \end{smallmatrix} \right\rangle &= \left\{ \begin{smallmatrix} n+3 \\ 3 \end{smallmatrix} \right\} - (2n+1) \left\{ \begin{smallmatrix} n+2 \\ 2 \end{smallmatrix} \right\} + \binom{2n+1}{2} \left\{ \begin{smallmatrix} n+1 \\ 1 \end{smallmatrix} \right\} \\ &= \frac{1}{2}3^{n+2} - (2n+3)2^{n+1} + \frac{1}{2}(4n^2 + 6n + 3). \end{aligned}$$

**6.69**  $\frac{1}{3}n(n+\frac{1}{2})(n+1)(2H_{2n}-H_n) - \frac{1}{36}n(10n^2+9n-1)$ . (Ce serait bien si on pouvait trouver un moyen d'automatiser les calculs de ce genre).

**6.70**  $1/k - 1/(k+z) = z/k^2 - z^2/k^3 + \dots$ , ce qui converge pour  $|z| < 1$ .

**6.71** Remarquez que  $\prod_{k=1}^n (1+z/k)e^{-z/k} = \binom{n+z}{n} n^{-z} e^{(\ln n - H_n)z}$ . Si  $f(z) = \frac{d}{dz}(z!)$ , on trouve  $f(z)/z! + \gamma = H_z$ .

**6.72** Pour  $\tan z$ , on peut utiliser la formule  $\tan z = \cot z - 2 \cot 2z$  (qui est équivalente à l'identité de l'exercice 23). La fonction  $z/\sin z = z \cot z + z \tan \frac{1}{2}z$  se développe en  $\sum_{n \geq 0} (-1)^{n-1} (4^n - 2) B_{2n} z^{2n}/(2n)!$ . Pour finir,

$$\begin{aligned} \ln \frac{\tan z}{z} &= \ln \frac{\sin z}{z} - \ln \cos z \\ &= \sum_{n \geq 1} (-1)^n \frac{4^n B_{2n} z^{2n}}{(2n)(2n)!} - \sum_{n \geq 1} (-1)^n \frac{4^n (4^n - 1) B_{2n} z^{2n}}{(2n)(2n)!} \\ &= \sum_{n \geq 1} (-1)^{n-1} \frac{4^n (4^n - 2) B_{2n} z^{2n}}{(2n)(2n)!}, \end{aligned}$$

car  $\frac{d}{dz} \ln \sin z = \cot z$  et  $\frac{d}{dz} \ln \cos z = -\tan z$ .

**6.73**  $\cot(z + \pi) = \cot z$  et  $\cot(z + \frac{1}{2}\pi) = -\tan z$  ; donc l'identité est équivalente à

$$\cot z = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \cot \frac{z + k\pi}{2^n},$$

qui se prouve par induction à partir du cas  $n = 1$ . La limite s'ensuit du fait que  $z \cot z \rightarrow 1$  lorsque  $z \rightarrow 0$ . On peut montrer qu'on a le droit de calculer la limite terme à terme, ce qui justifie la formule (6.88). La formule générale

$$\cot z = \frac{1}{n} \sum_{k=0}^{n-1} \cot \frac{z + k\pi}{n}$$

est vraie aussi. Elle peut se déduire de (6.88) ou de

$$\frac{1}{e^{nz} - 1} = \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{e^{z+2k\pi i/n} - 1},$$

cette dernière formule étant équivalente au développement en éléments simples de  $1/(z^n - 1)$ .

**6.74** Comme  $\tan 2z + \sec 2z = (\sin z + \cos z)/(\cos z - \sin z)$ , en posant  $x = 1$  dans (6.94) on obtient  $T_n(1) = 2^n T_n$  si  $n$  est impair,  $T_n(1) = 2^n |E_n|$  si  $n$  est pair, avec  $1/\cos z = \sum_{n \geq 0} |E_{2n}| z^{2n}/(2n)!$ . Les coefficients  $|E_n|$  sont appelés  *nombres sécants* ; s'ils ont des signes alternés, on les appelle  *nombres d'Euler*, à ne pas confondre avec les nombres eulériens  $\langle n \rangle_k$ . On a  $\langle E_0, E_2, E_4, \dots \rangle = \langle 1, -1, 5, -61, 1385, -50521, 2702765, \dots \rangle$ .

**6.75** Soient  $G(w, z) = \sin z / \cos(w + z)$  et  $H(w, z) = \cos z / \cos(w + z)$ , et soit  $G(w, z) + H(w, z) = \sum_{m,n} A_{m,n} w^m z^n / m! n!$ . Alors les équations  $G(w, 0) = 0$  et  $(\frac{\partial}{\partial z} - \frac{\partial}{\partial w})G(w, z) = H(w, z)$  impliquent que  $A_{m,0} = 0$  si  $m$  est impair,  $A_{m,n+1} = A_{m+1,n} + A_{m,n}$  si  $m+n$  est pair ; les équations  $H(0, z) = 1$  et  $(\frac{\partial}{\partial w} - \frac{\partial}{\partial z})H(w, z) = G(w, z)$  impliquent que  $A_{0,n} = [n=0]$  si  $n$  est pair,  $A_{m+1,n} = A_{m,n+1} + A_{m,n}$  si  $m+n$  est impair. Par conséquent, la  $n$ ième ligne sous le sommet du triangle contient les nombres  $A_{n,0}, A_{n-1,1}, \dots, A_{0,n}$ . À gauche,  $A_{n,0}$  est le nombre sécant  $|E_n|$  ; à droite,  $A_{0,n} = T_n + [n=0]$ .

**6.76** Soit  $A_n$  cette somme. Si on regarde l'équation (7.50), on voit que  $\sum_n A_n z^n / n! = \sum_{n,k} (-1)^k \{n\}_k 2^{n-k} k! z^n / n! = \sum_k (-1)^k 2^{-k} (e^{2z} - 1)^k = 2/(e^{2z} + 1) = 1 - \tanh z$ . Donc, d'après l'exercice 23 ou l'exercice 72,

$$A_n = (2^{n+1} - 4^{n+1}) B_{n+1} / (n+1) = (-1)^{(n+1)/2} T_n + [n=0].$$

**6.77** Cela se prouve par induction sur  $m$ , en utilisant la récurrence de l'exercice 18. On peut aussi le démontrer à partir de l'exercice (6.46), en utilisant le fait que

$$\begin{aligned} \frac{(-1)^{m-1}(m-1)!}{(e^z - 1)^m} &= (D + 1)^{\overline{m-1}} \frac{1}{e^z - 1} \\ &= \sum_{k=0}^{m-1} \begin{Bmatrix} m \\ m-k \end{Bmatrix} \frac{d^{m-k-1}}{dz^{m-k-1}} \frac{1}{e^z - 1}, \quad m > 0 \text{ entier}. \end{aligned}$$

Notons que la dernière équation est équivalente à

$$\frac{d^m}{dz^m} \frac{1}{e^z - 1} = (-1)^m \sum_k \begin{Bmatrix} m+1 \\ k \end{Bmatrix} \frac{(k-1)!}{(e^z - 1)^k}, \quad m \geq 0 \text{ entier}.$$

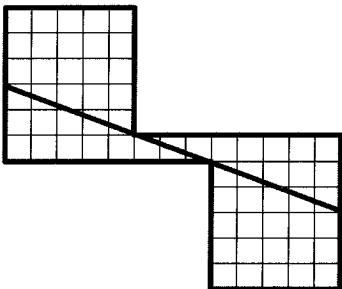
**6.78** Pour tout polynôme  $p(x)$  de degré  $\leq n$ , on a

$$p(x) = \sum_k p(-k) \binom{-x}{k} \binom{x+n}{n-k},$$

car cette équation est valable pour  $x = 0, -1, \dots, -n$ . L'identité donnée dans l'exercice constitue le cas particulier où  $p(x) = x\sigma_n(x)$  et  $x = 1$ . Notons en passant qu'on peut obtenir une expression plus simple des nombres de Bernoulli en fonction de nombres de Stirling en posant  $k = 1$  dans (6.99) :

$$\sum_{k \geq 0} \begin{Bmatrix} m \\ k \end{Bmatrix} (-1)^k \frac{k!}{k+1} = B_m.$$

**6.79** Sam Loyd [256, pages 288 et 378] donnait la construction



et déclarait qu'il avait découvert (sans le publier) l'assemblage  $64 = 65$  en 1858. Les paradoxes de ce genre remontent au moins au dix-huitième siècle, mais Loyd trouva une meilleure manière de les présenter.

**6.80** Nous savons que  $A_m/A_{m-1} \approx \phi$ , donc nous essayons  $A_{m-1} = 618034 + r$  et  $A_{m-2} = 381966 - r$ . Alors  $A_{m-3} = 236068 + 2r$ , etc., et nous finissons par trouver  $A_{m-18} = 144 - 2584r$  et  $A_{m-19} = 154 + 4181r$ . Par conséquent,  $r = 0$ ,  $x = 154$ ,  $y = 144$  et  $m = 20$ .

**6.81** Si  $P(F_{n+1}, F_n) = 0$  pour une infinité de  $n$  pairs, alors  $P(x, y)$  est divisible par  $U(x, y) - 1$ , où  $U(x, y) = x^2 - xy - y^2$ . En effet, si  $t$  est le degré total de  $P$ , on peut écrire

$$P(x, y) = \sum_{k=0}^t q_k x^k y^{t-k} + \sum_{j+k < t} r_{j,k} x^j y^k = Q(x, y) + R(x, y).$$

Alors

$$\frac{P(F_{n+1}, F_n)}{F_n^t} = \sum_{k=0}^t q_k \left( \frac{F_{n+1}}{F_n} \right)^k + O(1/F_n)$$

et on obtient  $\sum_{k=0}^t q_k \phi^k = 0$  en prenant la limite lorsque  $n \rightarrow \infty$ . Donc  $Q(x, y)$  est un multiple de  $U(x, y)$ , que nous noterons  $A(x, y)U(x, y)$ . Or,  $U(F_{n+1}, F_n) = (-1)^n$  et  $n$  est pair, donc  $P_0(x, y) = P(x, y) - (U(x, y) - 1)A(x, y)$  est un autre polynôme tel que  $P_0(F_{n+1}, F_n) = 0$ . Le degré total de  $P_0$  étant plus petit que  $t$ ,  $P_0$  est un multiple de  $U - 1$  par induction sur  $t$ .

De même,  $P(x, y)$  est divisible par  $U(x, y) + 1$  si  $P(F_{n+1}, F_n) = 0$  pour une infinité de  $n$  pairs. En combinant ces deux faits, on arrive à la condition nécessaire et suffisante désirée :  $P(x, y)$  est divisible par  $U(x, y)^2 - 1$ .

**6.82** Commencez par additionner les chiffres sans faire de retenue, pour obtenir ainsi des chiffres 0, 1 et 2. Puis utilisez les deux règles de retenue suivantes,

$$\begin{aligned} 0(d+1)(e+1) &\rightarrow 1 \text{ d } e, \\ 0(d+2)0e &\rightarrow 1 \text{ d } 0(e+1), \end{aligned}$$

en commençant toujours par la retenue la plus à gauche. Ce processus se termine forcément, parce que le nombre obtenu en lisant  $(b_m \dots b_2)_F$  comme  $(b_m \dots b_2)_2$  augmente à chaque fois que l'on fait une retenue. Cependant, il peut arriver qu'une retenue se propage à droite de la "virgule" : par exemple,  $(1)_F + (1)_F$  devient  $(10,01)_F$ . Cette propagation est toutefois limitée à deux chiffres, et on peut annuler ceux-ci en utilisant, si nécessaire, l'algorithme d'addition de 1 vu dans le chapitre.

Il existe aussi une opération de "multiplication" des entiers positifs ou nuls : si  $m = F_{j_1} + \dots + F_{j_q}$  et  $n = F_{k_1} + \dots + F_{k_r}$  dans le système de numération de Fibonacci, soit  $m \circ n = \sum_{b=1}^q \sum_{c=1}^r F_{j_b+k_c}$ , par analogie avec la multiplication des nombres binaires (cette définition implique que  $m \circ n \approx \sqrt{5} mn$  lorsque  $m$  et  $n$  sont grands, bien que  $1 \circ n \approx \phi^2 n$ ). On peut prouver, en utilisant les propriétés de l'addition de Fibonacci, que  $l \circ (m \circ n) = (l \circ m) \circ n$ .

Exercice :  $m \circ n =$   
 $mn +$   
 $\lfloor (m+1)/\phi \rfloor n +$   
 $m \lfloor (n+1)/\phi \rfloor$ .

**6.83** Oui. Par exemple, on peut prendre

$$\begin{aligned} A_0 &= 331635635998274737472200656430763; \\ A_1 &= 1510028911088401971189590305498785. \end{aligned}$$

La suite qui en résulte est telle que  $A_n$  est divisible par (mais différent de)  $p_k$  lorsque  $n \bmod m_k = r_k$ , où les nombres  $(p_k, m_k, r_k)$  prennent respectivement les 18 valeurs suivantes :

(3, 4, 1)	(2, 3, 2)	(5, 5, 1)
(7, 8, 3)	(17, 9, 4)	(11, 10, 2)
(47, 16, 7)	(19, 18, 10)	(61, 15, 3)
(2207, 32, 15)	(53, 27, 16)	(31, 30, 24)
(1087, 64, 31)	(109, 27, 7)	(41, 20, 10)
(4481, 64, 63)	(5779, 54, 52)	(2521, 60, 60)

Pour tout entier  $n$ , l'un de ces triplets s'applique. Par exemple, les six triplets de la première colonne couvrent tous les  $n$  impairs et ceux de la deuxième colonne couvrent tous les  $n$  pairs qui ne sont pas divisibles par 6. Le reste de la preuve est basé sur le fait que  $A_{m+n} = A_m F_{n-1} + A_{m+1} F_n$  ainsi que sur les deux congruences  $A_0 \equiv F_{m_k-r_k} \pmod{p_k}$  et  $A_1 \equiv F_{m_k-r_k+1} \pmod{p_k}$ , pour chacun des triplets  $(p_k, m_k, r_k)$ . Il existe aussi une meilleure solution, dans laquelle  $A_0$  et  $A_1$  ne contiennent "que" 17 chiffres [218].

**6.84** Les suites de l'exercice 62 satisfont  $A_{-m} = A_m$ ,  $B_{-m} = -B_m$  et

$$\begin{aligned} A_m A_n &= A_{m+n} + A_{m-n}; \\ A_m B_n &= B_{m+n} - B_{m-n}; \\ B_m B_n &= A_{m+n} - A_{m-n}. \end{aligned}$$

Soient  $f_k = B_{mk}/A_{mk+l}$  et  $g_k = A_{mk}/B_{mk+l}$ , où  $l = \frac{1}{2}(n-m)$ . Alors  $f_{k+1} - f_k = A_l B_m / (A_{2mk+n} + A_m)$  et  $g_k - g_{k+1} = A_l B_m / (A_{2mk+n} - A_m)$ . Par conséquent,

$$\begin{aligned} S_{m,n}^+ &= \frac{\sqrt{5}}{A_l B_m} \lim_{k \rightarrow \infty} (f_k - f_0) = \frac{\sqrt{5}}{\phi^l A_l L_m}; \\ S_{m,n}^- &= \frac{\sqrt{5}}{A_l B_m} \lim_{k \rightarrow \infty} (g_0 - g_k) = \frac{\sqrt{5}}{A_l L_m} \left( \frac{2}{B_l} - \frac{1}{\phi^l} \right) \\ &= \frac{2}{F_l L_l L_m} - S_{m,n}^+. \end{aligned}$$

**6.85** La propriété est vraie si et seulement si  $N$  est de l'une des sept formes suivantes :  $5^k$ ,  $2 \cdot 5^k$ ,  $4 \cdot 5^k$ ,  $3^j \cdot 5^k$ ,  $6 \cdot 5^k$ ,  $7 \cdot 5^k$ ,  $14 \cdot 5^k$ .

**6.86** Pour tout entier strictement positif  $m$ , soit  $r(m)$  le plus petit indice  $j$  tel que  $C_j$  soit divisible par  $m$ . Si un tel  $j$  n'existe pas, on pose  $r(m) = \infty$ . Alors  $C_n$  est divisible par  $m$  si et seulement si  $\text{pgcd}(C_n, C_{r(m)})$  est divisible par  $m$  si et seulement si  $C_{\text{pgcd}(n, r(m))}$  est divisible par  $m$  si et seulement si  $\text{pgcd}(n, r(m)) = r(m)$  si et seulement si  $n$  est divisible par  $r(m)$ .

(Réciproquement, il est facile de voir que la condition sur le pgcd est une implication de la condition que la suite  $C_1, C_2, \dots$  admet une fonction  $r(m)$ , à valeur éventuellement infinie, telle que  $C_n$  soit divisible par  $m$  si et seulement si  $n$  est divisible par  $r(m)$ ).

Soit maintenant  $\Pi(n) = C_1 C_2 \dots C_n$ , de sorte que

$$\binom{m+n}{m}_c = \frac{\Pi(m+n)}{\Pi(m)\Pi(n)}.$$

Si  $p$  est premier, le nombre de fois que  $p$  divise  $\Pi(n)$  est égal à  $f_p(n) = \sum_{k \geq 1} \lfloor n/r(p^k) \rfloor$ , car  $\lfloor n/p^k \rfloor$  est le nombre d'éléments  $\{C_1, \dots, C_n\}$  qui sont divisibles par  $p^k$ . Par conséquent,  $f_p(m+n) \geq f_p(m) + f_p(n)$  pour tout  $p$ , et  $\binom{m+n}{m}_c$  est un entier.

**6.87** Le produit matriciel est égal à

$$\begin{pmatrix} K_{n-2}(x_2, \dots, x_{n-1}) & K_{n-1}(x_2, \dots, x_{n-1}, x_n) \\ K_{n-1}(x_1, x_2, \dots, x_{n-1}) & K_n(x_1, x_2, \dots, x_{n-1}, x_n) \end{pmatrix}.$$

Il est lié, comme dans (6.137), aux produits de  $L$  et  $R$ , car

$$R^a \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & a \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} L^a.$$

Le déterminant est égal à  $K_n(x_1, \dots, x_n)$ . Le déterminant tridiagonal plus général suivant

$$\det \begin{pmatrix} x_1 & 1 & 0 & \dots & 0 \\ y_2 & x_2 & 1 & & 0 \\ 0 & y_3 & x_3 & 1 & \vdots \\ \vdots & & & \ddots & 1 \\ 0 & 0 & \dots & y_n & x_n \end{pmatrix},$$

satisfait la récurrence  $D_n = x_n D_{n-1} - y_n D_{n-2}$ .

**6.88** Soit  $\alpha^{-1} = a_0 + 1/(a_1 + 1/(a_2 + \dots))$  la représentation en fraction

continue de  $\alpha^{-1}$ . Alors on a

$$\frac{a_0}{z} + \frac{1}{A_0(z) + \frac{1}{A_1(z) + \frac{1}{A_2(z) + \ddots}}} = \frac{1-z}{z} \sum_{n \geq 1} z^{\lfloor n\alpha \rfloor},$$

où

$$A_m(z) = \frac{z^{-q_{m+1}} - z^{-q_{m-1}}}{z^{-q_m} - 1}, \quad q_m = K_m(a_1, \dots, a_m).$$

Il existe une preuve analogue à celle de (6.146), qui fait appel à une généralisation du théorème de Zeckendorf (Fraenkel [129, §4]). Si  $z = 1/b$ , où  $b$  est un entier  $\geq 2$ , la formule fournit la représentation en fraction continue du nombre transcendant  $(b-1) \sum_{n \geq 1} b^{-\lfloor n\alpha \rfloor}$ , comme dans l'exercice 49.

**6.89** Soit  $p = K(0, a_1, a_2, \dots, a_m)$ , de sorte que  $p/n$  est la  $m$ ième fraction de la suite qui converge vers la fraction continue. Alors  $\alpha = p/n + (-1)^m/nq$ , où  $q = K(a_1, \dots, a_m, \beta)$  et  $\beta > 1$ . Par conséquent, l'ensemble des points  $\{k\alpha\}$  pour  $0 \leq k < n$  peut s'écrire

$$\frac{0}{n}, \frac{1}{n} + \frac{(-1)^m \pi_1}{nq}, \dots, \frac{n-1}{n} + \frac{(-1)^m \pi_{n-1}}{nq},$$

où  $\pi_1 \dots \pi_{n-1}$  est une permutation de  $\{1, \dots, n-1\}$ . Soit  $f(v)$  le nombre de ces points qui sont  $< v$ ; alors  $f(v)$  et  $v_n$  augmentent tous deux de 1 lorsque  $v$  passe de  $k/n$  à  $(k+1)/n$ , sauf si  $k = 0$  ou  $k = n-1$ ; donc ils ne diffèrent jamais de 2 ou plus.

**6.90** D'après (6.139) et (6.136), il nous faut maximiser  $K(a_1, \dots, a_m)$  sur toutes les suites d'entiers strictement positifs dont la somme est  $\leq n+1$ . Le maximum est atteint lorsque tous les  $a$  sont égaux à 1, car si  $j \geq 1$  et  $a \geq 1$  on a

$$\begin{aligned} K_{j+k+1}(1, \dots, 1, a+1, b_1, \dots, b_k) \\ = K_{j+k+1}(1, \dots, 1, a, b_1, \dots, b_k) + K_j(1, \dots, 1) K_k(b_1, \dots, b_k) \\ \leq K_{j+k+1}(1, \dots, 1, a, b_1, \dots, b_k) + K_{j+k}(1, \dots, 1, a, b_1, \dots, b_k) \\ = K_{j+k+2}(1, \dots, 1, a, b_1, \dots, b_k). \end{aligned}$$

(Motzkin et Straus [278] montrent comment résoudre des problèmes plus généraux de maximisation sur des continuants).

**6.91** Un candidat pour le cas  $n \bmod 1 = \frac{1}{2}$  est présenté dans [213, §6], bien qu'il semblerait plus intéressant de multiplier les entiers dont il est question dans l'article par une constante où apparaîtrait  $\sqrt{\pi}$ . D'autre part, Renzo

Sprugnoli remarque qu'on peut définir  $\left\{ \begin{smallmatrix} n \\ m \end{smallmatrix} \right\} = \sum_k \binom{m}{k} k^n (-1)^{m-k} / m!$  pour  $m \geq 0$  entier et  $n \geq 0$  quelconque ; alors (6.3) est vraie pour tout  $n \geq 1$ .

*Une autre raison de retenir 1066 ?*

**6.92** (a) David Boyd a montré qu'il n'y a qu'un nombre fini de solutions pour tout  $p < 500$ , sauf peut-être pour  $p = 83, 127, 397$ . (b) Le comportement de  $b_n$  est plutôt étrange :  $b_n = \text{ppcm}(1, \dots, n)$  pour  $968 \leq n \leq 1066$  ; d'autre part,  $b_{600} = \text{ppcm}(1, \dots, 600) / (3^3 \cdot 5^2 \cdot 43)$ . Andrew Odlyzko remarque que  $p$  divise  $\text{ppcm}(1, \dots, n) / b_n$  si et seulement si  $k p^m \leq n < (k+1)p^m$  pour un certain  $m \geq 1$  et un certain  $k < p$  tel que  $p$  divise le numérateur de  $H_k$ . Par conséquent, il existera une infinité de  $n$  si on peut montrer, par exemple, que presque tous les nombres premiers n'ont qu'une telle valeur de  $k$  (à savoir  $k = p - 1$ ).

**6.93** (Brent [38] a trouvé le quotient partiel étonnamment grand 1568705 dans  $e^Y$ , mais ce n'est probablement qu'une coïncidence. Par exemple, Gosper a trouvé des quotients partiels plus grands encore dans  $\pi$ : Le 453294ième est 12996958 et le 11504931ième est 878783625).

**6.94** Considérez la fonction génératrice  $\sum_{m,n \geq 0} \frac{|w^{m+n}|}{m} w^m z^n$ , qui est de la forme  $\sum_n (wF(a, b, c) + zF(a', b', c'))^n$ , où  $F(a, b, c)$  est l'opérateur différentiel  $a + b\vartheta_w + c\vartheta_z$ .

**6.95** Il semble difficile d'envisager un succès complet, car les nombres de Stirling ne sont pas "holonomes" au sens de [382].

**7.1** Remplacez  $\square$  par  $z^4$  et  $\square$  par  $z$  dans la fonction génératrice, pour obtenir  $1/(1 - z^4 - z^2)$ . C'est la même formule que pour  $T$ , sauf que  $z$  est remplacé par  $z^2$ . La réponse est donc zéro si  $m$  est impair et  $F_{m/2+1}$  sinon.

**7.2**  $G(z) = 1/(1 - 2z) + 1/(1 - 3z)$  ;  $\widehat{G}(z) = e^{2z} + e^{3z}$ .

**7.3** Posez  $z = 1/10$  dans la fonction génératrice ; on trouve  $\frac{10}{9} \ln \frac{10}{9}$ .

**7.4** Divisez  $P(z)$  par  $Q(z)$  pour obtenir un quotient  $T(z)$  et un reste  $P_0(z)$  de degré plus petit que celui de  $Q$ . Il faut ajouter les coefficients de  $T(z)$  à ceux de  $[z^n] P_0(z)/Q(z)$  pour les petites valeurs de  $n$  (ce polynôme  $T(z)$  est le même que celui de (7.28)).

**7.5** C'est la convolution de  $(1 + z^2)^r$  et  $(1 + z)^r$ . Par conséquent,

$$S(z) = (1 + z + z^2 + z^3)^r.$$

On ne connaît pas de forme close pour les coefficients de cette fonction génératrice. La somme donnée n'a donc probablement pas de forme close simple. (Ainsi, les fonctions génératrices permettent d'aboutir à des résultats négatifs aussi bien que positifs).

**7.6** Soit  $g_n = A(n)\alpha + B(n)\beta + C(n)\gamma$  la solution de  $g_0 = \alpha$ ,  $g_1 = \beta$ ,  $g_n = g_{n-1} + 2g_{n-2} + (-1)^n\gamma$ . La fonction  $2^n$  marche pour  $\alpha = 1$ ,  $\beta = 2$ ,  $\gamma = 0$ ; la fonction  $(-1)^n$  marche pour  $\alpha = 1$ ,  $\beta = -1$ ,  $\gamma = 0$ ; la fonction  $(-1)^n n$  marche pour  $\alpha = 0$ ,  $\beta = -1$ ,  $\gamma = 3$ . Par conséquent,  $A(n) + 2B(n) = 2^n$ ,  $A(n) - B(n) = (-1)^n$  et  $-B(n) + 3C(n) = (-1)^n n$ .

**7.7**  $G(z) = (z/(1-z)^2)G(z) + 1$ , donc

$$G(z) = \frac{1-2z+z^2}{1-3z+z^2} = 1 + \frac{z}{1-3z+z^2};$$

on a  $g_n = F_{2n} + [n=0]$ .

**7.8** Dérivez  $(1-z)^{-x-1}$  deux fois par rapport à  $x$ , obtenant ainsi

$$\binom{x+n}{n} ((H_{x+n} - H_x)^2 - (H_{x+n}^{(2)} - H_x^{(2)}));$$

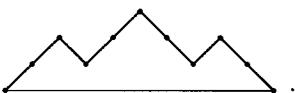
puis posez  $x = m$ .

**7.9**  $(n+1)(H_n^2 - H_n^{(2)}) - 2n(H_n - 1)$ .

**7.10** L'identité  $H_{k-1/2} - H_{-1/2} = \frac{2}{2k-1} + \dots + \frac{2}{1} = 2H_{2k} - H_k$  entraîne que  $\sum_k \binom{2k}{n-k} (2H_{2k} - H_k) = 4^n H_n$ .

**7.11** (a)  $C(z) = A(z)B(z^2)/(1-z)$ . (b)  $zB'(z) = A(2z)e^z$ , donc  $A(z) = \frac{z}{2}e^{-z/2}B'(\frac{z}{2})$ . (c)  $A(z) = B(z)/(1-z)^{r+1}$ , donc  $B(z) = (1-z)^{r+1}A(z)$  et on a  $f_k(r) = \binom{r+1}{k}(-1)^k$ .

**7.12**  $C_n$ . Les nombres de la ligne du haut correspondent aux positions des  $+1$  dans une suite de  $+1$  et de  $-1$  qui définissent une "chaîne de montagnes". Les nombres de la ligne du bas correspondent aux positions des  $-1$ . Par exemple, le tableau donné en exemple est associé à



**7.13** Prolongez la suite de façon périodique (en posant  $x_{m+k} = x_k$ ) et définissez  $s_n = x_1 + \dots + x_n$ . On a  $s_m = l$ ,  $s_{2m} = 2l$ , etc. Il existe forcément un plus grand indice  $k_j$  tel que  $s_{k_j} = j$ ,  $s_{k_j+m} = l+j$ , etc. Ces indices  $k_1, \dots, k_l$  (modulo  $m$ ) spécifient les décalages circulaires en question. Par exemple, dans la suite  $\langle -2, 1, -1, 0, 1, 1, -1, 1, 1, 1 \rangle$ , où  $m = 10$  et  $l = 2$ , on a  $k_1 = 17$  et  $k_2 = 24$ .

**7.14**  $\widehat{G}(z) = -2z\widehat{G}(z) + \widehat{G}(z)^2 + z$  (attention au dernier terme) implique que

$$\widehat{G}(z) = \frac{1+2z-\sqrt{1+4z^2}}{2}.$$

*Je parle que  
l'éventail d'ordre  
zéro, celui qui est  
tellement contro-  
versé, admet un  
arbre couvrant.*

Par conséquent,  $g_{2n+1} = 0$  et  $g_{2n} = (-1)^n (2n)! C_{n-1}$ , pour tout  $n > 0$ .

**7.15** Il existe  $\binom{n}{k} \omega_{n-k}$  partitions telles que le sous-ensemble contenant  $n+1$  contienne aussi  $k$  autres éléments. Par conséquent,  $\widehat{P}'(z) = e^z \widehat{P}(z)$ . La solution de cette équation différentielle est  $\widehat{P}(z) = e^{e^z+c}$ , et  $c = -1$  car  $\widehat{P}(0) = 1$ . On aurait aussi pu obtenir ce résultat en sommant (7.50) sur  $m$ , du fait que  $\omega_n = \sum_m \binom{n}{m}$ .

**7.16** On peut prendre le logarithme de

$$B(z) = 1 / ((1-z)^{a_1} (1-z^2)^{a_2} (1-z^3)^{a_3} (1-z^4)^{a_4} \dots),$$

puis appliquer la formule qui donne  $\ln \frac{1}{1-z}$  et changer l'ordre de sommation.

**7.17** Cela se déduit du fait que  $\int_0^\infty t^n e^{-t} dt = n!$ . Il existe aussi une formule duale :

$$\widehat{G}(z) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} G(ze^{-i\theta}) e^{e^{i\theta}} d\theta.$$

**7.18** (a)  $\zeta(z - \frac{1}{2})$ ; (b)  $-\zeta'(z)$ ; (c)  $\zeta(z)/\zeta(2z)$ . Tout entier strictement positif peut s'écrire de façon unique  $m^2 q$ , où  $q$  est sans carré.

**7.19** Si  $n > 0$ , le coefficient  $[z^n] \exp(x \ln F(z))$  est un polynôme de degré  $n$  en  $x$ , multiple de  $x$ . On trouve la première formule en mettant en équations les coefficients de  $z^n$  dans  $F(z)^x F(z)^y = F(z)^{x+y}$ . On trouve la seconde en mettant en équations les coefficients de  $z^{n-1}$  dans  $F'(z) F(z)^{x-1} F(z)^y = F'(z) F(z)^{x+y-1}$ , car on a

$$F'(z) F(z)^{x-1} = x^{-1} \frac{\partial}{\partial z} (F(z)^x) = x^{-1} \sum_{n \geq 0} n f_n(x) z^{n-1}.$$

(On peut trouver d'autres convolutions en dérivant par rapport à  $x$ , comme dans (7.44)).

On peut prouver encore plus, comme indiqué dans [221] :

$$\sum_{k=0}^n \frac{x f_k(x + tk)}{x + tk} \frac{y f_{n-k}(y + t(n-k))}{y + t(n-k)} = \frac{(x+y) f_n(x+y+tn)}{x+y+tn},$$

pour tous  $x, y$  et  $t$ . En fait,  $x f_n(x+tn)/(x+tn)$  est la suite de polynômes correspondant aux coefficients de  $\mathcal{F}_t(z)^x$ , où

$$\mathcal{F}_t(z) = F(z \mathcal{F}_t(z)^t).$$

(Nous en avons vu des cas particuliers en (5.59) et (6.52)).

**7.20** Soit  $G(z) = \sum_{n \geq 0} g_n z^n$ . Alors

$$z^l G^{(k)}(z) = \sum_{n \geq 0} n^k g_n z^{n-k+l} = \sum_{n \geq 0} (n+k-l)^k g_{n+k-l} z^n$$

pour tous  $k, l \geq 0$ , si on considère que  $g_n = 0$  pour tout  $n < 0$ . Donc, si  $P_0(z), \dots, P_m(z)$  sont des polynômes non tous nuls de degré maximal  $d$ , alors il existe des polynômes  $p_0(n), \dots, p_{m+d}(n)$  tels que

$$P_0(z)G(z) + \dots + P_m(z)G^{(m)}(z) = \sum_{n \geq 0} \sum_{j=0}^{m+d} p_j(n) g_{n+j-d} z^n.$$

Par conséquent, si  $G(z)$  est  $d$ -finie, alors

$$\sum_{j=0}^{m+d} p_j(n+d) g_{n+j} = 0, \quad \text{pour tout } n \geq 0.$$

La réciproque est similaire. (On peut en déduire que  $G(z)$  est  $d$ -finie si et seulement si la fge correspondante  $\tilde{G}(z)$  est  $d$ -finie).

**7.21** C'est le problème de la monnaie avec des pièces de valeurs 10 et 20. Par conséquent,  $G(z) = 1/(1-z^{10})(1-z^{20}) = \check{G}(z^{10})$ , et  $\check{G}(z) = 1/(1-z)(1-z^2)$ . (a) La décomposition en éléments simples de  $\check{G}(z)$  est  $\frac{1}{2}(1-z)^{-2} + \frac{1}{4}(1-z)^{-1} + \frac{1}{4}(1+z)^{-1}$ , donc  $[z^n] \check{G}(z) = \frac{1}{4}(2n+3+(-1)^n)$ . En posant  $n = 50$ , on trouve 26 façons possibles de payer. (b)  $\check{G}(z) = (1+z)/(1-z^2)^2 = (1+z)(1+2z^2+3z^4+\dots)$ , donc  $[z^n] \check{G}(z) = \lfloor n/2 \rfloor + 1$ .

Cette expression est à comparer avec la valeur  $N_n = \lfloor n/5 \rfloor + 1$  du problème de monnaie présenté dans le chapitre.

**7.22** Tout polygone a une "base", matérialisée par le côté horizontal le plus bas. Si A et B sont des polygones triangulés, soit  $A \Delta B$  la figure obtenue en collant la base de A sur le côté diagonal gauche de  $\Delta$  et la base de B sur son côté diagonal droit. Ainsi, par exemple,



(pour effectuer cette opération, on a le droit de déformer un peu les polygones). Toutes les triangulations peuvent être construites de cette façon, car la base de tout polygone est le côté d'un unique triangle et il existe toujours des polygones A et B à gauche et à droite de ce triangle.

En remplaçant chaque triangle par  $z$ , on obtient une série dans laquelle le coefficient de  $z^n$  est le nombre de triangulations comprenant  $n$  triangles, donc le nombre de façons de décomposer un polygone à  $n+2$  côtés en triangles. Comme  $P = 1 + zP^2$ , on trouve la fonction génératrice des nombres de Catalan,  $C_0 + C_1 z + C_2 z^2 + \dots$ . Le nombre de façons de triangular un polygone à  $n$  côtés est donc égal à  $C_{n-2} = \binom{2n-4}{n-2}/(n-1)$ .

*C'est une façon plutôt lente de trouver la réponse. Ça permet au caissier de gagner du temps en attendant que la police arrive.*

**7.23** Soit  $a_n$  le nombre recherché et  $b_n$  le nombre de façons de construire un pilier ayant une encoche de dimensions  $2 \times 1 \times 1$  à son sommet. En considérant tous les motifs possibles sur la surface supérieure du pilier, on trouve

$$\begin{aligned} a_n &= 2a_{n-1} + 4b_{n-1} + a_{n-2} + [n=0]; \\ b_n &= a_{n-1} + b_{n-1}. \end{aligned}$$

Les fonctions génératrices satisfont donc  $A = 2zA + 4zB + z^2A + 1$ ,  $B = zA + zB$ , et on trouve

$$A(z) = \frac{1-z}{(1+z)(1-4z+z^2)}.$$

Cette formule est liée aux pavages par des rectangles  $3 \times n$  par des dominos. En effet,  $a_n = \frac{1}{3}(U_{2n} + V_{2n+1} + (-1)^n) = \frac{1}{6}(2 + \sqrt{3})^{n+1} + \frac{1}{6}(2 - \sqrt{3})^{n+1} + \frac{1}{3}(-1)^n$ , ce qui est égal à  $(2 + \sqrt{3})^{n+1}/6$  arrondi à l'entier le plus proche.

**7.24**  $n \sum_{k_1+\dots+k_m=n} k_1 \cdot \dots \cdot k_m/m = F_{2n+1} + F_{2n-1} - 2$ . (Considérez le coefficient  $[z^{n-1}] \frac{d}{dz} \ln(1/(1-G(z)))$ , avec  $G(z) = z/(1-z)^2$ ).

**7.25** La fonction génératrice est  $P(z)/(1-z^m)$ , où  $P(z) = z + 2z^2 + \dots + (m-1)z^{m-1} = ((m-1)z^{m+1} - mz^m + z)/(1-z)^2$ . Le dénominateur est  $Q(z) = 1 - z^m = (1 - \omega^0 z)(1 - \omega^1 z) \dots (1 - \omega^{m-1} z)$ . D'après le théorème de développement rationnel pour des racines distinctes, on obtient

$$n \bmod m = \frac{m-1}{2} + \sum_{k=1}^{m-1} \frac{\omega^{-kn}}{\omega^k - 1}.$$

**7.26**  $(1-z-z^2)\mathfrak{F}(z) = F(z)$  entraîne que  $\mathfrak{F}_n = (2(n+1)F_n + nF_{n+1})/5$ , comme dans l'équation (7.61).

**7.27** Tout motif à cycles orientés commence par  ou par  ou encore par un cycle  $2 \times k$  (pour un certain  $k \geq 2$ ) orienté dans un sens ou dans l'autre. Par conséquent,

$$Q_n = Q_{n-1} + Q_{n-2} + 2Q_{n-2} + 2Q_{n-3} + \dots + 2Q_0$$

pour tout  $n \geq 2$ , et  $Q_0 = Q_1 = 1$ . La fonction génératrice est donc

$$\begin{aligned} Q(z) &= zQ(z) + z^2Q(z) + 2z^2Q(z)/(1-z) + 1 \\ &= 1/(1-z-z^2-2z^2/(1-z)) \\ &= \frac{(1-z)}{(1-2z-2z^2+z^3)} \\ &= \frac{\phi^2/5}{1-\phi^2z} + \frac{\phi^{-2}/5}{1-\phi^{-2}z} + \frac{2/5}{1+z}, \end{aligned}$$

"Curiously,  $a_{2n}$  is equal to  $U_{2n}^2$ , the square of the number of ways to tile a  $3 \times 2n$  rectangle with dominoes; and  $a_{2n+1} = 2V_{2n+1}^2$ ."

—I. Kaplansky

et  $Q_n = (\phi^{2n+2} + \phi^{-2n-2} + 2(-1)^n)/5 = ((\phi^{n+1} - \hat{\phi}^{n+1})/\sqrt{5})^2 = F_{n+1}^2$ .

**7.28** Plus généralement, si  $A(z) = (1+z+\dots+z^{m-1})B(z)$ , alors  $A_r + A_{r+m} + A_{r+2m} + \dots = B(1)$  pour tout  $0 \leq r < m$ . Dans notre cas,  $m = 10$  et  $B(z) = (1+z+\dots+z^9)(1+z^2+z^4+z^6+z^8)(1+z^5)$ .

**7.29**  $F(z) + F(z)^2 + F(z)^3 + \dots = z/(1-z-z^2-z) = (1/(1-(1+\sqrt{2})z) - (1/(1-(1-\sqrt{2})z))/\sqrt{8}$ , donc la réponse est  $((1+\sqrt{2})^n - (1-\sqrt{2})^n)/\sqrt{8}$ .

**7.30**  $\sum_{k=1}^n \binom{2n-1-k}{n-1} (a^n b^{n-k}/(1-\alpha z)^k + a^{n-k} b^n/(1-\beta z)^k)$ , d'après l'exercice 5.39.

**7.31** La fgd est  $\zeta(z)^2/\zeta(z-1)$ ; donc  $g(n)$  est le produit de  $(k+1-kp)$  sur toutes les puissances de nombres premiers  $p^k$  qui divisent exactement  $n$ .

**7.32** Nous pouvons supposer que  $b_k \geq 0$  pour tout  $k$ . Un ensemble de progressions arithmétiques forme une partition de l'ensemble des entiers naturels si et seulement si

$$\frac{1}{1-z} = \frac{z^{b_1}}{1-z^{a_1}} + \dots + \frac{z^{b_m}}{1-z^{a_m}}.$$

Soustrayez  $z^{b_m}/(1-z^{a_m})$  aux deux membres et posez  $z = e^{2\pi i/a_m}$ . Le membre gauche est infini ; le membre droit est fini, sauf si  $a_{m-1} = a_m$ .

**7.33**  $(-1)^{n-m+1}[n > m]/(n-m)$ .

**7.34** On a aussi  $G_n(z) = \sum_{k_1+(m+1)k_{m+1}=n} \binom{k_1+k_{m+1}}{k_{m+1}} (z^m)^{k_{m+1}}$ . Plus généralement, si

$$G_n = \sum_{k_1+2k_2+\dots+rk_r=n} \binom{k_1+k_2+\dots+k_r}{k_1, k_2, \dots, k_r} z_1^{k_1} z_2^{k_2} \dots z_r^{k_r},$$

alors  $G_n = z_1 G_{n-1} + z_2 G_{n-2} + \dots + z_r G_{n-r} + [n=0]$ , et la fonction génératrice est  $1/(1-z_1w-z_2w^2-\dots-z_rw^r)$ . Dans le cas particulier qui nous préoccupe, la réponse est  $1/(1-w-z^mw^{m+1})$  (voir (5.74) pour le cas  $m = 1$ ).

**7.35** (a)  $\frac{1}{n} \sum_{0 < k < n} (1/k + 1/(n-k)) = \frac{2}{n} H_{n-1}$ . (b)  $[z^n] (\ln \frac{1}{1-z})^2 = \frac{2!}{n!} \binom{n}{2} = \frac{2}{n} H_{n-1}$ , d'après (7.51) et (6.58). Une autre manière de résoudre la partie (b) consiste à utiliser la règle  $[z^n] F(z) = \frac{1}{n} [z^{n-1}] F'(z)$  avec  $F(z) = (\ln \frac{1}{1-z})^2$ .

**7.36**  $\frac{1-z^m}{1-z} A(z^m)$ .

**7.37** (a) On remarque dans la table que  $a_{2n} = a_{2n+1} = b_n$ .

$n$	0	1	2	3	4	5	6	7	8	9	10
$a_n$	1	1	2	2	4	4	6	6	10	10	14
$b_n$	1	2	4	6	10	14	20	26	36	46	60

(b)  $A(z) = 1/((1-z)(1-z^2)(1-z^4)(1-z^8)\dots)$ . (c)  $B(z) = A(z)/(1-z)$ , et nous voulons montrer que  $A(z) = (1+z)B(z^2)$ . On le déduit de  $A(z) = A(z^2)/(1-z)$ .

**7.38**  $(1-wz)M(w,z) = \sum_{m,n \geq 1} (\min(m,n) - \min(m-1,n-1))w^m z^n = \sum_{m,n \geq 1} w^m z^n = wz/(1-w)(1-z)$ . Plus généralement,

$$M(z_1, \dots, z_m) = \frac{z_1 \dots z_m}{(1-z_1) \dots (1-z_m)(1-z_1 \dots z_m)}.$$

**7.39** Voici les réponses à la suggestion :

$$\begin{aligned} & \sum_{\substack{1 \leq k_1 < k_2 < \dots < k_m \leq n}} a_{k_1} a_{k_2} \dots a_{k_m}; \\ & \sum_{\substack{1 \leq k_1 \leq k_2 \leq \dots \leq k_m \leq n}} a_{k_1} a_{k_2} \dots a_{k_m}. \end{aligned}$$

Nous en déduisons ce qui suit. (a) Nous cherchons le coefficient de  $z^m$  dans le produit  $(1+z)(1+2z)\dots(1+nz)$ . Ce polynôme est le polynôme réciproque de  $(z+1)^{\overline{n}}$ , donc il est égal à  $\left[\begin{smallmatrix} n+1 \\ n+1 \end{smallmatrix}\right] + \left[\begin{smallmatrix} n+1 \\ n \end{smallmatrix}\right]z + \dots + \left[\begin{smallmatrix} n+1 \\ 1 \end{smallmatrix}\right]z^n$  et la réponse est  $\left[\begin{smallmatrix} n+1 \\ n+1-m \end{smallmatrix}\right]$ . (b) D'après (7.48), le coefficient de  $z^m$  dans  $1/((1-z)(1-2z)\dots(1-nz))$  est  $\left\{\begin{smallmatrix} m+n \\ n \end{smallmatrix}\right\}$ .

**7.40** La fge de  $\langle nF_{n-1} - F_n \rangle$  est  $(z-1)\widehat{F}(z)$ , où  $\widehat{F}(z) = \sum_{n \geq 0} F_n z^n / n! = (e^{\phi z} - e^{\widehat{\phi} z})/\sqrt{5}$ . La fge de  $\langle n_i \rangle$  est  $e^{-z}/(1-z)$ . Le produit vaut

$$5^{-1/2} (e^{(\widehat{\phi}-1)z} - e^{(\phi-1)z}) = 5^{-1/2} (e^{-\phi z} - e^{-\widehat{\phi} z}).$$

On a

$$\widehat{F}(z)e^{-z} = -\widehat{F}(-z).$$

La réponse est donc  $(-1)^n F_n$ .

**7.41** Le nombre de permutations alternantes dont le plus grand élément  $n$  est en position  $2k$  est égal à  $\binom{n-1}{2k-1} A_{2k-1} A_{n-2k}$ . De même, le nombre de permutations alternantes dont le plus petit élément  $1$  est en position  $2k+1$  est égal à  $\binom{n-1}{2k} A_{2k} A_{n-2k-1}$ . En sommant sur tous les cas possibles, on trouve

$$2A_n = \sum_k \binom{n-1}{k} A_k A_{n-1-k} + 2[n=0] + [n=1].$$

La fge  $\widehat{A}$  satisfait donc  $2\widehat{A}'(z) = \widehat{A}(z)^2 + 1$  et  $\widehat{A}(0) = 1$ . La fonction donnée dans l'énoncé est la solution de cette équation différentielle. (Par conséquent,  $A_n = |\mathbb{E}_n| + T_n$  est un nombre sécant lorsque  $n$  est pair et un nombre tangent lorsque  $n$  est impair).

**7.42** Soit  $a_n$  le nombre de chaînes d'ADN martien qui ne finissent ni par  $c$  ni par  $e$ , et  $b_n$  le nombre de chaînes qui finissent par  $c$  ou par  $e$ . Alors

$$\begin{aligned} a_n &= 3a_{n-1} + 2b_{n-1} + [n=0], & b_n &= 2a_{n-1} + b_{n-1}; \\ A(z) &= 3zA(z) + 2zB(z) + 1, & B(z) &= 2zA(z) + zB(z); \\ A(z) &= \frac{1-z}{1-4z-z^2}, & B(z) &= \frac{2z}{1-4z-z^2}; \end{aligned}$$

et le nombre total est égal à  $[z^n](1+z)/(1-4z-z^2) = F_{3n+2}$ .

**7.43** D'après (5.45),  $g_n = \Delta^n \dot{G}(0)$ . La  $n$ ième différence d'un produit peut s'écrire

$$\Delta^n A(z)B(z) = \sum_k \binom{n}{k} (\Delta^k E^{n-k} A(z)) (\Delta^{n-k} B(z)),$$

et  $E^{n-k} = (1+\Delta)^{n-k} = \sum_j \binom{n-k}{j} \Delta^j$ . Par conséquent, on trouve

$$h_n = \sum_{j,k} \binom{n}{k} \binom{n-k}{j} f_{j+k} g_{n-k}.$$

C'est une somme sur tous les coefficients binomiaux. On peut l'écrire de façon plus symétrique

$$h_n = \sum_{j+k+l=n} \binom{n}{j, k, l} f_{j+k} g_{k+l}.$$

**7.44** Chaque partition en  $k$  sous-ensembles non vides peut être ordonnée de  $k!$  manières, donc  $b_k = k!$ . Ainsi,  $\widehat{Q}(z) = \sum_{n,k \geq 0} \left\{ \begin{matrix} n \\ k \end{matrix} \right\} k! z^n / n! = \sum_{k \geq 0} (e^z - 1)^k = 1/(2 - e^z)$ . Comme ceci n'est rien d'autre que la série géométrique  $\sum_{k \geq 0} e^{kz}/2^{k+1}$ , on a  $a_k = 1/2^{k+1}$ . Pour finir,  $c_k = 2^k$ ; considérez toutes les permutations telles que les  $x_k$  sont tous distincts, transformez chaque " $>$ " en " $<$ " et autorisez chaque " $<$ " à se transformer en " $<$ " ou en " $=$ ". Par exemple, comme  $1 < 3 > 2$ , la permutation  $x_1 x_3 x_2$  engendre  $x_1 < x_3 < x_2$  et  $x_1 = x_3 < x_2$ .

**7.45** Cette somme est égale à  $\sum_{n \geq 1} r(n)/n^2$ , où  $r(n)$  désigne le nombre de façons d'écrire  $n$  comme un produit de deux facteurs premiers entre eux. Si  $n$  est divisible par  $t$  nombres premiers distincts, alors  $r(n) = 2^t$ . Par conséquent,  $r(n)/n^2$  est multiplicative et la somme est égale à

$$\begin{aligned} \prod_p \left( 1 + \frac{2}{p^2} + \frac{2}{p^4} \dots \right) &= \prod_p \left( 1 + \frac{2}{p^2 - 1} \right) \\ &= \prod_p \left( \frac{p^2 + 1}{p^2 - 1} \right) = \zeta(2)^2 / \zeta(4) = \frac{5}{2}. \end{aligned}$$

**7.46** Soit  $S_n = \sum_{0 \leq k \leq n/2} \binom{n-2k}{k} \alpha^k$ . Alors  $S_n = S_{n-1} + \alpha S_{n-3} + [n=0]$ , et la fonction génératrice est  $1/(1-z-\alpha z^3)$ . Lorsque  $\alpha = -\frac{4}{27}$ , la suggestion nous indique qu'elle se factorise en  $1/(1 + \frac{1}{3}z)(1 - \frac{2}{3}z)^2$ . Alors, d'après le théorème de développement des fonctions rationnelles, on a  $S_n = (\frac{2}{3}n + c)(\frac{2}{3})^n + \frac{1}{9}(-\frac{1}{3})^n$ , et la constante  $c$  est égale à  $\frac{8}{9}$ .

**7.47** La représentation de Stern–Brocot de  $\sqrt{3}$  est  $R(LR^2)^\infty$ , car

$$\sqrt{3} + 1 = 2 + \cfrac{1}{1 + \cfrac{1}{\sqrt{3} + 1}}.$$

La suite de fractions correspondante commence par  $\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \frac{12}{7}, \frac{19}{11}, \frac{26}{15}, \dots$ . Elle finit par présenter le motif cyclique

$$\frac{V_{2n-1}+V_{2n+1}}{U_{2n}}, \frac{U_{2n}+V_{2n+1}}{V_{2n+1}}, \frac{U_{2n+2}+V_{2n-1}}{U_{2n}+V_{2n+1}}, \frac{V_{2n+1}+V_{2n+3}}{U_{2n+2}}, \dots$$

**7.48** On a  $g_0 = 0$ , et si  $g_1 = m$  la fonction génératrice satisfait

$$aG(z) + bz^{-1}G(z) + cz^{-2}(G(z) - mz) + \frac{d}{1-z} = 0.$$

Il existe donc un polynôme  $P(z)$  tel que  $G(z) = P(z)/(az^2 + bz + c)(1-z)$ . Soient  $\rho_1$  et  $\rho_2$  les racines de  $cz^2 + bz + a$ , avec  $|\rho_1| \geq |\rho_2|$ . Si  $b^2 - 4ac \leq 0$ , alors  $|\rho_1|^2 = \rho_1\rho_2 = a/c$  est rationnel, ce qui contredit le fait que  $\sqrt{g_n}$  tend vers  $1 + \sqrt{2}$ . Par conséquent,  $\rho_1 = (-b + \sqrt{b^2 - 4ca})/2c = 1 + \sqrt{2}$ ; ceci implique que  $a = -c$ ,  $b = -2c$  et  $\rho_2 = 1 - \sqrt{2}$ . La fonction génératrice est maintenant de la forme

$$\begin{aligned} G(z) &= \frac{z(m - (r + m)z)}{(1 - 2z - z^2)(1 - z)} \\ &= \frac{-r + (2m + r)z}{2(1 - 2z - z^2)} + \frac{r}{2(1 - z)} = mz + (2m - r)z^2 + \dots, \end{aligned}$$

avec  $r = d/c$ . Comme  $g_2$  est un entier,  $r$  est aussi un entier. Nous avons aussi

$$g_n = \alpha(1 + \sqrt{2})^n + \hat{\alpha}(1 - \sqrt{2})^n + \frac{1}{2}r = \lfloor \alpha(1 + \sqrt{2})^n \rfloor,$$

ce qui n'est possible que si  $r = -1$ , car l'expression  $(1 - \sqrt{2})^n$  tend vers zéro en prenant des valeurs successivement positives et négatives. Par conséquent,  $(a, b, c, d) = \pm(1, 2, -1, 1)$ . Maintenant, nous trouvons que  $\alpha = \frac{1}{4}(1 + \sqrt{2}m)$ , ce qui ne peut être entre 0 et 1 que si  $0 \leq m \leq 2$ . Comme chacune de ces valeurs donne lieu à une solution, il y a trois suites  $\langle g_n \rangle$  possibles, qui sont  $\langle 0, 0, 1, 3, 8, \dots \rangle$ ,  $\langle 0, 1, 3, 8, 20, \dots \rangle$  et  $\langle 0, 2, 5, 13, 32, \dots \rangle$ .

**7.49** (a) Le dénominateur de  $(1/(1-(1+\sqrt{2})z) + 1/(1-(1-\sqrt{2})z))$  est  $1-2z-z^2$ , donc  $a_n = 2a_{n-1} + a_{n-2}$  pour  $n \geq 2$ . (b) C'est vrai car  $a_n$  est pair et  $-1 < 1-\sqrt{2} < 0$ . (c) Soit

$$b_n = \left(\frac{p+\sqrt{q}}{2}\right)^n + \left(\frac{p-\sqrt{q}}{2}\right)^n.$$

Nous aimerais bien que  $b_n$  soit impair pour tout  $n > 0$ , et que  $-1 < (p-\sqrt{q})/2 < 0$ . En opérant comme dans la partie (a), on trouve que  $b_0 = 2$ ,  $b_1 = p$  et  $b_n = pb_{n-1} + \frac{1}{4}(q-p^2)b_{n-2}$  pour  $n \geq 2$ . Les valeurs  $p = 3$  et  $q = 17$  donnent une solution satisfaisante.

**7.50** En généralisant l'idée de la multiplication de l'exercice 22, on trouve

$$Q = \dots + Q \begin{array}{c} Q \\ \triangle \end{array} Q + Q \begin{array}{c} Q \\ \square \end{array} Q + Q \begin{array}{c} Q \\ \text{pentagon} \\ Q \end{array} Q + \dots$$

Remplacez chaque polygone à  $n$  côtés par  $z^{n-2}$ . Cette substitution est bien compatible avec la multiplication car l'opération de collage prend un polygone à  $m$  côtés et un polygone à  $n$  côtés pour en faire un polygone à  $m+n-2$  côtés. Par conséquent, la fonction génératrice est

$$Q = 1 + zQ^2 + z^2Q^3 + z^3Q^4 + \dots = 1 + \frac{zQ^2}{1 - zQ},$$

ce qui se résout en  $Q = (1+z-\sqrt{1-6z+z^2})/4z$ . Le coefficient de  $z^{n-2}$  dans cette série est égal au nombre de façons de tracer des diagonales qui ne se coupent pas dans un polygone convexe à  $n$  côtés. Ces coefficients ne semblent pas admettre de forme close en fonction d'autres quantités que nous avons vues dans ce livre ; cependant, leur comportement asymptotique est connu [207, exercice 2.2.1-12].

Notons en passant que, si on remplace chaque polygone à  $n$  côtés dans  $Q$  par  $wz^{n-2}$ , on obtient

$$Q = \frac{1+z-\sqrt{1-(4w+2)z+z^2}}{2(1+w)z},$$

une formule dans laquelle le coefficient de  $w^mz^{n-2}$  représente le nombre de façons de diviser un polygone à  $n$  côtés en  $m$  polygones en traçant des diagonales qui ne se croisent pas.

**7.51** Il est important de remarquer tout d'abord que le carré du nombre de pavages possibles est égal au nombre de configurations de cycles d'une certaine sorte ; c'est une généralisation de l'exercice 27. On peut dénombrer ces configurations en calculant le déterminant d'une matrice dont les valeurs propres sont faciles à déterminer. Lorsque  $m = 3$  et  $n = 4$ , le fait que  $\cos 36^\circ = \phi/2$  peut vous être utile (exercice 6.46).

*Donnez-moi les polynômes de Legendre et je vous donnerai une forme close.*

**7.52** Les toutes premières valeurs sont  $p_0(y) = 1$ ,  $p_1(y) = y$ ,  $p_2(y) = y^2 + y$ ,  $p_3(y) = y^3 + 3y^2 + 3y$ . Soit  $p_n(y) = q_{2n}(x)$ , avec  $y = x(1-x)$ , et cherchons une fonction génératrice qui définit  $q_{2n+1}(x)$  de façon pratique. La fonction  $\sum_n q_n(x)z^n/n! = 2e^{ixz}/(e^{iz} + 1)$  convient, et il s'ensuit que  $q_n(x) = i^n E_n(x)$ , où  $E_n(x)$  est ce qu'on appelle un polynôme d'Euler. On a  $\sum (-1)^x x^n \delta x = \frac{1}{2}(-1)^{x+1} E_n(x)$ ; par conséquent, les polynômes d'Euler sont analogues aux polynômes de Bernoulli, et ils se factorisent de façon similaire à (6.98). Selon l'exercice 6.23, nous avons  $n!E_{n-1}(x) = \sum_{k=0}^n \binom{n}{k} B_k x^{n-k} (2 - 2^{k+1})$ . D'après l'exercice 6.54, ce polynôme est à coefficients entiers. Par conséquent, les coefficients de  $q_{2n}(x)$ , qui sont des puissances de 2, sont aussi entiers; donc  $p_n(y)$  est à coefficients entiers. Pour finir, la relation  $(4y-1)p_n''(y) + 2p_n'(y) = 2n(2n-1)p_{n-1}(y)$  entraîne que

$$2m(2m-1)\binom{n}{m} = m(m+1)\binom{n}{m+1} + 2n(2n-1)\binom{n-1}{m-1},$$

et il s'ensuit que les  $\binom{n}{m}$  sont strictement positifs.

On peut écrire une preuve similaire pour montrer que l'expression  $(-1)^n(2n+2)E_{2n+1}(x)/(2x-1)$ , lorsqu'on l'exprime comme un polynôme de degré  $n$  en  $y$ , est à coefficients entiers. On peut montrer aussi que  $\binom{n}{1}$  est égal au nombre de Genocchi  $(-1)^{n-1}(2^{2n+1}-2)B_{2n}$  (voir l'exercice 6.24), et que  $\binom{n}{n-1} = \binom{n}{2}$ ,  $\binom{n}{n-2} = 2\binom{n+1}{4} + 3\binom{n}{4}$ , etc.

**7.53** Il est égal à  $P_{(1+\nu_{4n+1}+\nu_{4n+3})/6}$ . Ainsi, par exemple,  $T_{20} = P_{12} = 210$ ;  $T_{285} = P_{165} = 40755$ .

**7.54** Soit  $E_k$  l'opération qui annule tous les coefficients d'une série sauf ceux de  $z^n$  tels que  $n \bmod m = k$ . La construction décrite dans l'énoncé équivaut à appliquer l'opération

$$E_0 S E_0 S (E_0 + E_1) S \dots S (E_0 + E_1 + \dots + E_{m-1})$$

à  $1/(1-z)$ , où  $S$  signifie "multiplier par  $1/(1-z)$ ". Il y a  $m!$  termes

$$E_0 S E_{k_1} S E_{k_2} S \dots S E_{k_m},$$

avec  $0 \leq k_j < j$ , et chacun de ces termes vaut  $z^{rm}/(1-z^m)^{m+1}$ , si on convient que  $r$  est le nombre de valeurs  $j$  telles que  $k_j < k_{j+1}$ . Comme exactement  $\binom{m}{r}$  termes ont une valeur donnée de  $r$ , on en déduit, d'après (6.37), que le coefficient de  $z^{mn}$  est  $\sum_{r=0}^{m-1} \binom{m}{r} \binom{n+m-r}{m} = (n+1)^m$ . (Le fait que l'opération  $E_k$  peut s'exprimer en fonction des racines complexes de l'unité semble n'être d'aucune utilité dans ce problème).

**7.55** Supposons que  $P_0(z)F(z) + \dots + P_m(z)F^{(m)}(z) = Q_0(z)G(z) + \dots + Q_n(z)G^{(n)}(z) = 0$ , où  $P_m(z)$  et  $Q_n(z)$  sont non nuls. (a) Soit  $H(z) =$

$F(z) + G(z)$ . Alors il existe des fonctions rationnelles  $R_{k,l}(z)$ , pour  $0 \leq l < m+n$ , telles que  $H^{(k)}(z) = R_{k,0}(z)F^{(0)}(z) + \cdots + R_{k,m-1}(z)F^{(m-1)}(z) + R_{k,m}(z)G^{(0)}(z) + \cdots + R_{k,m+n-1}(z)G^{(n-1)}(z)$ . Les  $m+n+1$  vecteurs  $(R_{k,0}(z), \dots, R_{k,m+n-1}(z))$  sont linéairement dépendants dans l'espace vectoriel de dimension  $m+n$  dont les éléments sont des fonctions rationnelles. Par conséquent, il existe des fonctions rationnelles  $S_l(z)$ , non toutes nulles, telles que  $S_0(z)H^{(0)}(z) + \cdots + S_{m+n}(z)H^{(m+n)}(z) = 0$ . (b) De même, soit  $H(z) = F(z)G(z)$ . Il existe des fonctions rationnelles  $R_{k,l}(z)$ , pour  $0 \leq l < mn$ , telles que  $H^{(k)}(z) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} R_{k,ni+j}(z)F^{(i)}(z)G^{(j)}(z)$ . Par conséquent, il existe des fonctions rationnelles  $S_l(z)$ , non toutes nulles, telles que  $S_0(z)H^{(0)}(z) + \cdots + S_{mn}(z)H^{(mn)}(z) = 0$ . (Avec une preuve similaire, on peut montrer que si  $\langle f_n \rangle$  et  $\langle g_n \rangle$  sont p-récurrentes, alors  $\langle f_n + g_n \rangle$  et  $\langle f_n g_n \rangle$  le sont aussi ; par contre, il n'existe pas de résultat similaire pour les quotients : par exemple,  $\cos z$  est d-finie mais  $1/\cos z$  ne l'est pas).

**7.56** Euler [113] a démontré que ce nombre est égal à  $[z^n] 1/\sqrt{1-2z-3z^2}$  et a donné la formule  $t_n = \sum_{k \geq 0} n^{2k}/k!^2 = \sum_k \binom{n}{k} \binom{n-k}{k}$ . Il a aussi découvert, en étudiant ces nombres, une “défaillance de l'induction” :  $3t_n - t_{n+1}$  est égal à  $F_{n-1}(F_{n-1} + 1)$  pour  $0 \leq n \leq 8$ , mais cette règle empirique faillit pour  $n \geq 8$ . George Andrews [12] a donné l'explication de ce mystère en montrant que la somme  $\sum_k [z^{n+10k}] (1+z+z^2)^n$  admet une forme close en fonction des nombres de Fibonacci.

H. S. Wilf fait remarquer que  $[z^n](a+bz+cz^2)^n = [z^n] 1/f(z)$ , où  $f(z) = \sqrt{1-2bz+(b^2-4ac)z^2}$  (voir [373, page 159]), et il s'ensuit que les coefficients satisfont

$$(n+1)A_{n+1} - (2n+1)bA_n + n(b^2 - 4ac)A_{n-1} = 0.$$

On peut utiliser l'algorithme de Petkovšek [291] pour prouver que cette récurrence admet une solution qui peut s'exprimer en forme close comme une somme finie de termes hypergéométriques si et seulement si  $abc(b^2-4ac) = 0$ . On en déduit en particulier que les coefficients trinomiaux centraux n'ont pas de forme close de ce type. L'étape suivante consisterait vraisemblablement à étendre ce résultat à une plus large classe de formes closes (comprenant par exemple les nombres harmoniques et/ou les nombres de Stirling).

*Donnez-moi les polynômes de Legendre et je vous donnerai une forme close.*

**7.57** (Paul Erdős offrait une récompense de \$500 pour la solution).

*Paul Erdős est décédé en 1996 (N.d.T.).*

**8.1**  $\frac{1}{24} + \frac{1}{48} + \frac{1}{48} + \frac{1}{48} + \frac{1}{48} + \frac{1}{24} = \frac{1}{6}$  (en fait, la probabilité d'un double est toujours égale à  $\frac{1}{6}$  si l'un au moins des dés est juste). Comme toutes les faces dont la somme est 7 ont la même probabilité dans la distribution  $Pr_1$ , la probabilité que  $S = 7$  est la même que celle de faire un double.

**8.2** Il y a 12 façons possibles de choisir les deux cartes dont il est question

et 50! façons de placer les autres. La probabilité est donc  $12 \cdot 50!/52! = 12/(51 \cdot 52) = \frac{1}{17 \cdot 13} = \frac{1}{221}$ .

**8.3**  $\frac{1}{10}(3+2+\dots+9+2) = 4.8$ ;  $\frac{1}{9}(3^2+2^2+\dots+9^2+2^2-10(4.8)^2) = \frac{388}{45}$ , ce qui fait à peu près 8,6. Les vraies moyenne et variance, pour une pièce juste, sont respectivement de 6 et 22. Les étudiants de Stanford ne tombent pas pile dessus, loin s'en faut. A Princeton, on trouve respectivement 6,4 et  $\frac{562}{45} \approx 12,5$ . (On a aussi  $\kappa_4 = 2974$ , ce qui est plutôt élevé. Par conséquent, l'écart-type de cette estimation de la variance pour  $n = 10$  est aussi assez grand :  $\sqrt{2974/10 + 2(22)^2/9} \approx 20,1$  d'après l'exercice 54. On ne peut pas dire que les étudiants ont triché).

**8.4** Cela se déduit de (8.38) et (8.39), car  $F(z) = G(z)H(z)$  (la même formule est valable pour tous les cumulants, même si  $F(z)$  et  $G(z)$  ont des coefficients négatifs).

**8.5** Remplacez  $P$  par  $p$  et  $F$  par  $q = 1 - p$ . Si  $S_A = S_B = \frac{1}{2}$ , alors  $p^2qN = \frac{1}{2}$  et  $pq^2N = \frac{1}{2}q + \frac{1}{2}$ . La solution est  $p = 1/\phi^2$ ,  $q = 1/\phi$ .

**8.6** Dans ce cas,  $X|y$  a la même distribution que  $X$  pour tout  $y$ , donc  $E(X|Y) = EX$  est une constante et  $V(E(X|Y)) = 0$ . De même,  $V(X|Y)$  est constante et égale à sa valeur moyenne.

**8.7** D'après l'inégalité monotone de Tchebychev du chapitre 2, on a  $1 = (p_1 + p_2 + \dots + p_6)^2 \leqslant 6(p_1^2 + p_2^2 + \dots + p_6^2)$ .

**8.8** Soient  $p = \Pr(\omega \in A \cap B)$ ,  $q = \Pr(\omega \notin A)$  et  $r = \Pr(\omega \notin B)$ . Alors  $p + q + r = 1$ , et l'identité à prouver est  $p = (p + r)(p + q) - qr$ .

**8.9** C'est vrai (pourvu, bien évidemment, que  $F$  et  $G$  soient définies sur les valeurs respectives de  $X$  et  $Y$ ), car

$$\begin{aligned}\Pr(F(X)=f \text{ et } G(Y)=g) &= \sum_{\substack{x \in F^{-1}(f) \\ y \in G^{-1}(g)}} \Pr(X=x \text{ et } Y=y) \\ &= \sum_{\substack{x \in F^{-1}(f) \\ y \in G^{-1}(g)}} \Pr(X=x) \cdot \Pr(Y=y) \\ &= \Pr(F(X)=f) \cdot \Pr(G(Y)=g).\end{aligned}$$

**8.10** Deux. Soient  $x_1 < x_2$  deux médianes ; alors  $1 \leqslant \Pr(X \leqslant x_1) + \Pr(X \geqslant x_2) \leqslant 1$ , donc il y a égalité. (Il existe des distributions discrètes qui n'ont pas de médiane ; par exemple, lorsque  $\Omega$  est l'ensemble de toutes les fractions de la forme  $\pm 1/n$ , avec  $\Pr(+1/n) = \Pr(-1/n) = \frac{\pi^2}{12}n^{-2}$ ).

**8.11** Prenons par exemple  $K = k$  avec probabilité  $4/(k+1)(k+2)(k+3)$ , pour tout entier  $k \geq 0$ . Alors  $EK = 1$ , mais  $E(K^2) = \infty$  (de même, on peut construire des variables aléatoires avec des cumulants finis jusqu'à  $\kappa_m$ , mais telles que  $\kappa_{m+1} = \infty$ ).

**8.12** (a) Soit  $p_k = \Pr(X=k)$ . Si  $0 < x \leq 1$ , alors  $\Pr(X \leq r) = \sum_{k \leq r} p_k \leq \sum_{k \leq r} x^{k-r} p_k \leq \sum_k x^{k-r} p_k = x^{-r} P(x)$ . L'autre inégalité se prouve de manière similaire. (b) Posez  $x = \alpha/(1-\alpha)$  pour minimiser le membre droit (on obtient dans l'exercice 9.42 une meilleure estimation de la somme donnée).

**8.13** (Solution de Boris Pittel). Posons  $Y = (X_1 + \dots + X_n)/n$  et  $Z = (X_{n+1} + \dots + X_{2n})/n$ . Alors

$$\begin{aligned} \Pr\left(\left|\frac{Y+Z}{2} - \alpha\right| \leq |Y - \alpha|\right) \\ \geq \Pr\left(\left|\frac{Y - \alpha}{2}\right| + \left|\frac{Z - \alpha}{2}\right| \leq |Y - \alpha|\right) \\ = \Pr(|Z - \alpha| \leq |Y - \alpha|) \geq \frac{1}{2}. \end{aligned}$$

Le " $\geq$ " de la dernière inégalité est en fait un " $>$ " pour toute distribution de probabilité discrète, car  $\Pr(Y = Z) > 0$ .

**8.14**  $Moy(H) = p Moy(F) + q Moy(G)$  ;  $Var(H) = p Var(F) + q Var(G) + pq(Moy(F) - Moy(G))^2$ . En fait, un mélange n'est rien d'autre qu'un cas particulier de probabilité conditionnelle : soient  $Y$  la pièce de monnaie,  $X|P$  la probabilité associée à  $F(z)$  et  $X|F$  la probabilité associée à  $G(z)$  ; alors  $VX = EV(X|Y) + VE(X|Y)$ , où  $EV(X|Y) = pV(X|P) + qV(X|F)$  et  $VE(X|Y)$  est la variance de  $p z^{Moy(F)} + q z^{Moy(G)}$ .

**8.15** D'après la règle de dérivation des fonctions composées, on a  $H'(z) = G'(z)F'(G(z))$  et  $H''(z) = G''(z)F'(G(z)) + G'(z)^2 F''(G(z))$ . Par conséquent,

$$\begin{aligned} Moy(H) &= Moy(F) Moy(G); \\ Var(H) &= Var(F) Moy(G)^2 + Moy(F) Var(G). \end{aligned}$$

On peut se représenter comme suit la variable aléatoire correspondant à la distribution de probabilité  $H$  : tirer un entier positif ou nul  $n$  selon la distribution  $F$  ; puis ajouter les valeurs de  $n$  variables aléatoires indépendantes qui ont la distribution  $G$ . L'identité qui donne la variance dans cet exercice est un cas particulier de (8.106), lorsque  $X$  admet la distribution  $H$  et  $Y$  admet la distribution  $F$ .

**8.16**  $e^{w(z-1)} / (1-w)$ .

**8.17**  $\Pr(Y_{n,p} \leq m) = \Pr(Y_{n,p} + n \leq m + n)$  = probabilité d'avoir besoin de  $m + n$  lancers au plus pour obtenir  $n$  piles = probabilité d'obtenir au moins  $n$  piles en  $m + n$  lancers =  $\Pr(X_{m+n,p} \geq n)$ . Ainsi,

$$\begin{aligned}\sum_{k \leq m} \binom{n+k-1}{k} p^n q^k &= \sum_{k \geq n} \binom{m+n}{k} p^k q^{m+n-k} \\ &= \sum_{k \leq m} \binom{m+n}{k} p^{m+n-k} q^k;\end{aligned}$$

et on retrouve (5.19) avec  $n = r$ ,  $x = q$  et  $y = p$ .

**8.18** (a)  $G_X(z) = e^{\mu(z-1)}$ . (b) Pour tout  $m \geq 1$ , le  $m$ ième cumulant est égal à  $\mu$  (dans (8.55), le cas  $\mu = 1$  est appelé  $F_\infty$ ).

**8.19** (a)  $G_{X_1+X_2}(z) = G_{X_1}(z)G_{X_2}(z) = e^{(\mu_1+\mu_2)(z-1)}$ . Par conséquent, la probabilité est égale à  $e^{-\mu_1-\mu_2}(\mu_1 + \mu_2)^n/n!$ ; la somme de deux lois de Poisson est une loi de Poisson. (b) Si  $K_m X$  désigne le  $m$ ième cumulant d'une variable aléatoire quelconque  $X$ , alors  $K_m(aX_1+bX_2) = a^m(K_m X_1) + b^m(K_m X_2)$  pour  $a, b \geq 0$ . La réponse est donc  $2^m \mu_1 + 3^m \mu_2$ .

**8.20** La fgp correspondante est  $G(z) = z^m/F(z)$ , avec

$$\begin{aligned}F(z) &= z^m + (1-z) \sum_{k=1}^m \tilde{A}_{(k)}[A^{(k)} = A_{(k)}] z^{m-k}, \\ F'(1) &= m - \sum_{k=1}^m \tilde{A}_{(k)}[A^{(k)} = A_{(k)}], \\ F''(1) &= m(m-1) - 2 \sum_{k=1}^m (m-k) \tilde{A}_{(k)}[A^{(k)} = A_{(k)}].\end{aligned}$$

**8.21**  $N$  est égal à  $\sum_{n \geq 0} q_n$ , où  $q_n$  désigne la probabilité que le jeu ne soit pas terminé après  $n$  lancers. Soit  $p_n$  la probabilité que le jeu s'arrête au  $n$ ième lancer ; alors  $p_n + q_n = q_{n-1}$ . Donc, la durée moyenne du jeu est égale à  $\sum_{n \geq 1} np_n = (q_0 - q_1) + 2(q_1 - q_2) + 3(q_2 - q_3) + \dots = q_0 + q_1 + q_2 + \dots = N$ , du fait que  $\lim_{n \rightarrow \infty} nq_n = 0$ .

On peut aussi résoudre le problème différemment, en remplaçant  $P$  et  $F$  par  $\frac{1}{2}z$ . Alors, en dérivant la première équation de (8.78), on trouve que  $N(1) + N'(1) = N'(1) + S'_A(1) + S'_B(1)$ .

Dans notre cas,  $N = \frac{16}{3}$ .

**8.22** Par définition, on a  $V(X|Y) = E(X^2|Y) - (E(X|Y))^2$  et  $V(E(X|Y)) = E((E(X|Y))^2) - (E(E(X|Y)))^2$ ; donc  $E(V(X|Y)) + V(E(X|Y)) = E(E(X^2|Y)) - (E(E(X|Y)))^2$ . Or,  $E(E(X|Y)) = \sum_y \Pr(Y=y)E(X|y) = \sum_{x,y} \Pr(Y=y) \times \Pr((X|y)=x)x = EX$  et  $E(E(X^2|Y)) = E(X^2)$ , donc on trouve bien  $VX$ .

**8.23** Soient  $\Omega_0 = \{\square, \blacksquare\}^2$ ,  $\Omega_1 = \{\square\cdot, \square\circ, \square\square, \square\blacksquare\}^2$ , et  $\Omega_2$  l'ensemble des 16 autres éléments de  $\Omega$ . Alors  $\Pr_{11}(\omega) - \Pr_{00}(\omega) = \frac{+20}{576}, \frac{-7}{576}$  ou  $\frac{+2}{576}$  selon que  $\omega \in \Omega_0$ ,  $\Omega_1$  ou  $\Omega_2$ . Les événements A doivent donc contenir  $k_j$  éléments de  $\Omega_j$ , où le triplet  $(k_0, k_1, k_2)$  est l'un des suivants :  $(0, 0, 0)$ ,  $(0, 2, 7)$ ,  $(0, 4, 14)$ ,  $(1, 4, 4)$ ,  $(1, 6, 11)$ ,  $(2, 6, 1)$ ,  $(2, 8, 8)$ ,  $(2, 10, 15)$ ,  $(3, 10, 5)$ ,  $(3, 12, 12)$ ,  $(4, 12, 2)$ ,  $(4, 14, 9)$ ,  $(4, 16, 16)$ . Par exemple, il y a  $\binom{4}{2} \binom{16}{6} \binom{16}{1}$  événements de type  $(2, 6, 1)$ . Le nombre total de ces événements est  $[z^0](1 + z^{20})^4(1 + z^{-7})^{16}(1 + z^2)^{16}$ , ce qui donne 1304872090. Si on se restreint aux événements qui ne dépendent que de S, on trouve 40 solutions  $S \in A$ , où  $A = \emptyset, \{\frac{2}{12}, \frac{4}{10}, \frac{6}{8}\}, \{\frac{2}{12}, 5, 9\}, \{2, 12, \frac{4}{10}, \frac{6}{8}, 5, 9\}, \{2, 4, 6, 8, 10, 12\}, \{\frac{3}{11}, 7, \frac{5}{9}, 4, 10\}$ , et les complémentaires de ces ensembles (la notation " $\frac{2}{12}$ " signifie "soit 2, soit 12, mais pas les deux").

**8.24** (a) Chacun des dés se retrouve entre les mains de J avec probabilité  $p = \frac{1}{6} + \left(\frac{5}{6}\right)^2 p$ ; donc  $p = \frac{6}{11}$ . Soit  $q = \frac{5}{11}$ . Alors la fgp du gain total de J est  $(q + pz)^{2n+1}$ ; sa moyenne est égale à  $(2n+1)p$  et sa variance à  $(2n+1)pq$  d'après (8.61). (b)  $\binom{5}{3}p^3q^2 + \binom{5}{4}p^4q + \binom{5}{5}p^5 = \frac{94176}{161051} \approx 0,585$ .

**8.25** Soit  $G_n(z)$  la fgp de la somme restante après n lancers ; alors

$$\begin{aligned} G_0(z) &= z^A; \\ G_n(z) &= \sum_{k=1}^6 G_{n-1}(z^{2(k-1)/5})/6, \quad \text{pour } n > 0 \end{aligned}$$

(les exposants non entiers ne posent pas de problème). Il s'ensuit donc que  $\text{Moy}(G_n) = \text{Moy}(G_{n-1})$  et  $\text{Var}(G_n) + \text{Moy}(G_n)^2 = \frac{22}{15}(\text{Var}(G_{n-1}) + \text{Moy}(G_{n-1})^2)$ . La moyenne est donc toujours égale à A, mais la variance, qui vaut  $((\frac{22}{15})^n - 1)A^2$ , croît avec n.

Ce problème est peut-être plus facile à résoudre sans les fonctions génératrices.

**8.26** La fgp  $F_{l,n}(z)$  satisfait  $F'_{l,n}(z) = F_{l,n-1}(z)/l$  ; donc  $\text{Moy}(F_{l,n}) = F'_{l,n}(1) = [n \geq l]/l$  et  $F''_{l,n}(1) = [n \geq 2l]/l^2$  ; la variance se calcule aisément. (En fait,

$$F_{l,n}(z) = \sum_{0 \leq k \leq n/l} \frac{1}{k!} \left(\frac{z-1}{l}\right)^k,$$

et la distribution tend vers une loi de Poisson de moyenne  $1/l$  lorsque  $n \rightarrow \infty$ ).

**8.27**  $(n^2\Sigma_3 - 3n\Sigma_2\Sigma_1 + 2\Sigma_1^3)/n(n-1)(n-2)$ , où  $\Sigma_k = X_1^k + \dots + X_n^k$ , convient. On déduit cette formule des identités

$$\begin{aligned} E\Sigma_3 &= n\mu_3; \\ E(\Sigma_2\Sigma_1) &= n\mu_3 + n(n-1)\mu_2\mu_1; \\ E(\Sigma_1^3) &= n\mu_3 + 3n(n-1)\mu_2\mu_1 + n(n-1)(n-2)\mu_1^3. \end{aligned}$$

Notons en passant que le troisième cumulant est  $\kappa_3 = E((X - EX)^3)$ , mais que le quatrième n'a pas une expression aussi simple :  $\kappa_4 = E((X - EX)^4) - 3(VX)^2$ .

**8.28** (Bien qu'on suppose implicitement dans l'énoncé que  $p = q = \frac{1}{2}$ , nous allons résoudre l'exercice dans un cadre général). Remplacez P par  $pz$  et F par  $qz$  pour obtenir  $S_A(z) = p^2qz^3/(1-pz)(1-qz)(1-pqz^2)$  et  $S_B(z) = pq^2z^3/(1-qz)(1-pqz^2)$ . La fgp de la probabilité conditionnelle qu'Alice gagne au  $n$ ième lancer, sachant qu'elle gagne, est

$$\frac{S_A(z)}{S_A(1)} = z^3 \cdot \frac{q}{1-pz} \cdot \frac{p}{1-qz} \cdot \frac{1-pq}{1-pqz^2}.$$

C'est un produit de pseudo-fgp, dont la moyenne est égale à  $3 + p/q + q/p + 2pq/(1-pq)$ . La fgp concernant Bill est la même sans le facteur  $q/(1-pz)$ ; la moyenne de Bill vaut donc  $3 + q/p + 2pq/(1-pq)$ . Si  $p = q = \frac{1}{2}$ , la réponse à la question (a) est  $\frac{17}{3}$ , et, pour la question (b), on trouve  $\frac{14}{3}$ . Bien que Bill gagne deux fois moins souvent qu'Alice, il gagne plus rapidement qu'elle en général. Le nombre moyen total de lancers est  $\frac{2}{3} \cdot \frac{17}{3} + \frac{1}{3} \cdot \frac{14}{3} = \frac{16}{3}$ , ce qui concorde avec l'exercice 21. Pour chacun des deux motifs, la durée moyenne du jeu en solitaire est 8.

**8.29** Posez  $P = F = \frac{1}{2}$  dans

$$\begin{aligned} 1 + N(P+F) &= N + S_A + S_B + S_C \\ NPPFP &= S_A(PFP+1) + S_B(PFP+FP) + S_C(PFP+FP) \\ NPFPP &= S_A(FPP+P) + S_B(FPP+1) + S_C(FPP) \\ NFPPP &= S_A(PP) + S_B(P) + S_C \end{aligned}$$

pour obtenir les probabilités de victoire. Plus généralement, on a  $S_A + S_B + S_C = 1$  et

$$\begin{aligned} S_A(A:A) + S_B(B:A) + S_C(C:A) &= S_A(A:B) + S_B(B:B) + S_C(C:B) \\ &= S_A(A:B) + S_B(B:C) + S_C(C:C). \end{aligned}$$

En particulier, le fait que  $9S_A + 3S_B + 3S_C = 5S_A + 9S_B + S_C = 2S_A + 4S_B + 8S_C$  entraîne que  $S_A = \frac{16}{52}$ ,  $S_B = \frac{17}{52}$  et  $S_C = \frac{19}{52}$ .

**8.30** La variance de  $P(h_1, \dots, h_n; k)|k$  est celle de la loi binomiale décalée  $((m-1+z)/m)^{k-1}z$ ; cette variance est égale, d'après (8.61), à  $(k-1)(\frac{1}{m})(1-\frac{1}{m})$ . Par conséquent, la moyenne de la variance vaut  $Moy(S)(m-1)/m^2$ . La variance de la moyenne est la variance de  $(k-1)/m$ , soit  $Var(S)/m^2$ . Selon (8.106), la somme de ces deux quantités doit être égale à VP, et c'est bien le cas. En effet, nous avons simplement refait, sous une forme un peu différente, le calcul de (8.96) (voir l'exercice 15).

**8.31** (a) Une solution brutale consisterait à poser cinq équations à cinq inconnues :

$$\begin{aligned} A &= \frac{1}{2}zB + \frac{1}{2}zE; & B &= \frac{1}{2}zC; & C &= 1 + \frac{1}{2}zB + \frac{1}{2}zD; \\ D &= \frac{1}{2}zC + \frac{1}{2}zE; & E &= \frac{1}{2}zD. \end{aligned}$$

Cependant, comme les sommets C et D sont à égale distance du but, on peut les réunir ; c'est aussi le cas de A et E. En posant  $X = B + E$  et  $Y = C + D$ , on se ramène à trois équations :

$$A = \frac{1}{2}zX; \quad X = \frac{1}{2}zY; \quad Y = 1 + \frac{1}{2}zX + \frac{1}{2}zY.$$

Par conséquent,  $A = z^2/(4 - 2z - z^2)$  ; on trouve  $\text{Moy}(A) = 6$  et  $\text{Var}(A) = 22$ . Cela ne vous rappelle rien ? Ce problème est équivalent à celui qui consiste à tirer à pile ou face jusqu'à obtenir deux piles à la suite : "pile" correspond à "s'approcher de la pomme", tandis que "face" correspond à "s'en éloigner". (b) L'inégalité de Tchebychev entraîne que  $\Pr(S \geq 100) = \Pr((S - 6)^2 \geq 94^2) \leq 22/94^2 \approx 0,0025$ . (c) La seconde inégalité implique que  $\Pr(S \geq 100) \leq 1/x^{98}(4 - 2x - x^2)$  pour tout  $x \geq 1$ , et on trouve le majorant  $0,00000005$  pour  $x = (\sqrt{49001} - 99)/100$  (la probabilité réelle vaut à peu près  $0,000000009$  selon l'exercice 37).

**8.32** En raison des symétries du graphe, on peut réduire les situations envisageables à quatre possibilités :

- D, les deux états sont diagonalement opposés ;
- A, les deux états sont adjacents et aucun n'est le Kansas ;
- K, l'un des états est le Kansas, l'autre est différent ;
- S, les deux états sont égaux.

*"Toto, j'ai  
l'impression que  
nous ne sommes plus  
au Kansas."*

— Dorothy

Si on considère les transitions de ce processus de Markov, on obtient quatre équations

$$\begin{aligned} D &= 1 + z\left(\frac{2}{9}D + \frac{2}{12}K\right) \\ A &= z\left(\frac{4}{9}A + \frac{4}{12}K\right) \\ K &= z\left(\frac{4}{9}D + \frac{4}{9}A + \frac{4}{12}K\right) \\ S &= z\left(\frac{3}{9}D + \frac{1}{9}A + \frac{2}{12}K\right) \end{aligned}$$

dont la somme donne  $D + K + A + S = 1 + z(D + A + K)$ . On trouve

$$S = \frac{81z - 45z^2 - 4z^3}{243 - 243z + 24z^2 + 8z^3},$$

mais la façon la plus simple de trouver la moyenne et la variance semble être d'écrire  $z = 1 + w$  et de développer en série en  $w$ , en ignorant les multiples

de  $w^2$  :

$$\begin{aligned} D &= \frac{27}{16} + \frac{1593}{512}w + \dots; \\ A &= \frac{9}{8} + \frac{2115}{256}w + \dots; \\ K &= \frac{15}{8} + \frac{2661}{256}w + \dots. \end{aligned}$$

Alors  $S'(1) = \frac{27}{16} + \frac{9}{8} + \frac{15}{8} = \frac{75}{16}$  et  $\frac{1}{2}S''(1) = \frac{1593}{512} + \frac{2115}{256} + \frac{2661}{256} = \frac{11145}{512}$ . La moyenne est égale à  $\frac{75}{16}$  et la variance vaut  $\frac{105}{4}$ . Quelqu'un a-t-il une solution plus simple ?

**8.33** Première réponse : bien sûr que oui, car les valeurs de la fonction de hachage  $h_1, \dots, h_n$  sont indépendantes. Seconde réponse : absolument pas, bien que les valeurs de la fonction de hachage  $h_1, \dots, h_n$  soient indépendantes. On a  $\Pr(X_j=0) = \sum_{k=1}^n s_k ([j \neq k] (m-1)/m) = (1-s_j)(m-1)/m$ , mais  $\Pr(X_1=X_2=0) = \sum_{k=1}^n s_k [k>2] (m-1)^2/m^2 = (1-s_1-s_2)(m-1)^2/m^2 \neq \Pr(X_1=0)\Pr(X_2=0)$ .

**8.34** Soit  $[z^n] S_m(z)$  la probabilité que Gina ait avancé de strictement moins de  $m$  unités après  $n$  tours. Alors  $S_m(1)$  désigne son score moyen sur un trou par- $m$ ,  $[z^m] S_m(z)$  désigne la probabilité qu'elle perde sur un tel trou contre un joueur régulier, et  $1 - [z^{m-1}] S_m(z)$  est sa probabilité de gagner sur le même trou. On trouve la récurrence

$$\begin{aligned} S_0(z) &= 0; \\ S_m(z) &= (1 + pzS_{m-2}(z) + qzS_{m-1}(z))/(1 - rz), \quad \text{pour } m > 0. \end{aligned}$$

Pour résoudre la question (a), il suffit de calculer les coefficients pour  $m, n \leq 4$ . Pour ne manipuler que des entiers, on peut remplacer  $z$  par 100. On obtient ainsi le tableau

$S_0$	0	0	0	0	0
$S_1$	1	4	16	64	256
$S_2$	1	95	744	4432	23552
$S_3$	1	100	9065	104044	819808
$S_4$	1	100	9975	868535	12964304

Ainsi, Gina gagne avec probabilité  $1 - 0,868535 = 0,131465$  et elle perd avec probabilité  $0,12964304$ . (b) Pour trouver le nombre moyen de coups, on calcule

$$S_1(1) = \frac{25}{24}; \quad S_2(1) = \frac{4675}{2304}; \quad S_3(1) = \frac{667825}{221184}; \quad S_4(1) = \frac{85134475}{21233664}.$$

Notons en passant que  $S_5(1) \approx 4,9995$ . Par conséquent, sur des trous par-5, en moyenne, Gina gagnera quel que soit le mode de décompte (score ou nombre de trous), mais, sur des trous par-3, elle perdra dans les deux cas.

**8.35** Selon le théorème des restes chinois, la condition sera vraie pour tout  $n$  si et seulement si elle est vraie pour  $n = 1$ . L'identité suivante constitue une condition nécessaire et suffisante :

$$(p_2 + p_4 + p_6 + (p_1 + p_3 + p_5)w)(p_3 + p_6 + (p_1 + p_4)z + (p_2 + p_5)z^2) \\ = (p_1 w z + p_2 z^2 + p_3 w + p_4 z + p_5 w z^2 + p_6).$$

Cependant, ce n'est en fait qu'une reformulation du problème. La condition suivante est plus simple et ne fait appel qu'à deux coefficients du produit ci-dessus :

$$(p_2 + p_4 + p_6)(p_3 + p_6) = p_6, \quad (p_1 + p_3 + p_5)(p_2 + p_5) = p_5.$$

Voici comment construire la distribution : se donner  $a_0 + a_1 = b_0 + b_1 + b_2 = 1$  et poser  $p_1 = a_1 b_1$ ,  $p_2 = a_0 b_2$ ,  $p_3 = a_1 b_0$ ,  $p_4 = a_0 b_1$ ,  $p_5 = a_1 b_2$ ,  $p_6 = a_0 b_0$ .

**8.36** (a)       . (b) Supposons que le  $k$ ème dé a des faces de valeurs  $s_1, \dots, s_6$ , et soit  $p_k(z) = z^{s_1} + \dots + z^{s_6}$ . Nous voulons que ces polynômes satisfassent  $p_1(z) \dots p_n(z) = (z + z^2 + z^3 + z^4 + z^5 + z^6)^n$ . Ce dernier polynôme à coefficients rationnels se décompose en facteurs irréductibles  $z^n(z+1)^n(z^2+z+1)^n(z^2-z+1)^n$ . Par conséquent,  $p_k(z)$  doit être de la forme  $z^{a_k}(z+1)^{b_k}(z^2+z+1)^{c_k}(z^2-z+1)^{d_k}$ . On a forcément  $a_k \geq 1$ , du fait que  $p_k(0) = 0$ ; en fait,  $a_k = 1$  car  $a_1 + \dots + a_n = n$ . De plus, la condition  $p_k(1) = 6$  implique que  $b_k = c_k = 1$ . Il est maintenant facile de voir que  $0 \leq d_k \leq 2$ , car on obtiendrait des coefficients négatifs si on posait  $d_k > 2$ . Lorsque  $d = 0$  et  $d = 2$ , on retrouve les deux dés de la question (a); donc, les seules solutions possibles sont constituées de  $k$  paires de dés de la question (a), auxquelles on ajoute  $n - 2k$  dés ordinaires, pour  $k \leq \frac{1}{2}n$ .

**8.37** Pour tout  $n > 0$ , le nombre de suites de piles et de faces de longueur  $n$  qui ne contiennent pas deux piles à la suite sauf juste à la fin et égal à  $F_{n-1}$  en raison de la relation qui lie ces suites aux pavages par des dominos. Par conséquent, la probabilité  $p_n$  est égale à  $F_{n-1}/2^n$ . On a aussi  $q_n = F_{n+1}/2^{n-1}$ , car  $\sum_{k \geq n} F_k z^k = (F_n z^n + F_{n-1} z^{n+1})/(1 - z - z^2)$ . (On peut aussi, bien sûr, résoudre ce problème avec des fonctions génératrices).

**8.38** Si on a déjà vu  $k$  faces, le processus amenant à en voir une nouvelle équivaut à tirer à pile ou face avec une probabilité de succès égale à  $p_k = (m - k)/m$ . Par conséquent, la fgp est  $\prod_{k=0}^{l-1} p_k z / (1 - q_k z) = \prod_{k=0}^{l-1} (m - k)z / (m - kz)$ . La moyenne est  $\sum_{k=0}^{l-1} p_k^{-1} = m(H_m - H_{m-l})$ ; la variance est  $m^2(H_m^{(2)} - H_{m-l}^{(2)}) - m(H_m - H_{m-l})$ ; et l'équation (7.48) nous fournit une forme close pour la probabilité recherchée,  $m^{-n} m! \{ \begin{smallmatrix} n-1 \\ l-1 \end{smallmatrix} \} / (m-l)!$ . (Cet exercice est connu sous le nom du "problème du collectionneur de coupons").

**8.39**  $E(X) = P(-1)$  ;  $V(X) = P(-2) - P(-1)^2$  ;  $E(\ln X) = -P'(0)$ .

**8.40** (a) On a  $\kappa_m = n(0!\{^m_1\}p - 1!\{^m_2\}p^2 + 2!\{^m_3\}p^3 - \dots)$  d'après (7.50). Remarquons que le troisième cumulant est  $n p q (q - p)$  et que le quatrième est  $n p q (1 - 6pq)$ . L'identité  $q + pe^t = (p + qe^{-t})e^t$  montre que  $f_m(p) = (-1)^m f_m(q) + [m=1]$ ; on peut donc écrire  $f_m(p) = g_m(pq)(q-p)^{[m \text{ impair}]}$ , où  $g_m$  est un polynôme de degré  $[m/2]$ , pour tout  $m > 1$ . (b) Soient  $p = \frac{1}{2}$  et  $F(t) = \ln(\frac{1}{2} + \frac{1}{2}e^t)$ . Alors  $\sum_{m \geq 1} \kappa_m t^{m-1}/(m-1)! = F'(t) = 1/(e^t+1)$ , et on peut appliquer le résultatat de l'exercice 6.23.

**8.41** Si  $G(z)$  est la fgp d'une variable aléatoire  $X$  à valeurs entières strictement positives, alors  $\int_0^1 G(z) dz/z = \sum_{k \geq 1} \Pr(X=k)/k = E(X^{-1})$ . Si  $X$  est la distribution du nombre de jets nécessaires pour obtenir  $n+1$  piles, alors  $G(z) = (pz/(1-qz))^{n+1}$  d'après (8.59), et l'intégrale devient

$$\int_0^1 \left( \frac{pz}{1-qz} \right)^{n+1} \frac{dz}{z} = \int_0^1 \frac{w^n dw}{1 + (q/p)w}$$

si on pose  $w = pz/(1-qz)$ . Lorsque  $p = q$ , l'expression à intégrer peut s'écrire  $(-1)^n ((1+w)^{-1} - 1 + w - w^2 + \dots + (-1)^n w^{n-1})$ , et donc l'intégrale vaut  $(-1)^n (\ln 2 - 1 + \frac{1}{2} - \frac{1}{3} + \dots + (-1)^n/n)$ . On a  $H_{2n} - H_n = \ln 2 - \frac{1}{4}n^{-1} + \frac{1}{16}n^{-2} + O(n^{-4})$  d'après (9.28), et il s'ensuit que  $E(X_{n+1}^{-1}) = \frac{1}{2}n^{-1} - \frac{1}{4}n^{-2} + O(n^{-4})$ .

**8.42** Soient  $F_n(z)$  et  $G_n(z)$  les fgp respectives du nombre d'après-midi travaillés, selon que notre homme est embauché ou non le premier matin. Soient aussi  $q_h = 1 - p_h$  et  $q_f = 1 - p_f$ . Alors  $F_0(z) = G_0(z) = 1$  et

$$\begin{aligned} F_n(z) &= p_h z G_{n-1}(z) + q_h F_{n-1}(z); \\ G_n(z) &= p_f F_{n-1}(z) + q_f z G_{n-1}(z). \end{aligned}$$

La solution est donnée par la fonction génératrice double

$$G(w, z) = \sum_{n \geq 0} G_n(z) w^n = A(w)/(1 - zB(w)),$$

où  $B(w) = w(q_f - (q_f - p_h)w)/(1 - q_h w)$  et  $A(w) = (1 - B(w))/(1 - w)$ . Alors  $\sum_{n \geq 0} G'_n(1) w^n = \alpha w/(1-w)^2 + \beta/(1-w) - \beta/(1-(q_f - p_h)w)$ , avec

$$\alpha = \frac{p_h}{p_h + p_f}, \quad \beta = \frac{p_f(q_f - p_h)}{(p_h + p_f)^2};$$

donc  $G'_n(1) = \alpha n + \beta(1 - (q_f - p_h)^n)$  (on a aussi  $G''_n(1) = \alpha^2 n^2 + O(n)$ , donc la variance est en  $O(n)$ ).

**8.43**  $G_n(z) = \sum_{k \geq 0} [n]_k z^k / n! = z^n / n!$ , d'après (6.11). C'est un produit de fgp binomiales,  $\prod_{k=1}^n ((k-1+z)/k)$ , dont la kième a une moyenne de  $1/k$  et une variance de  $(k-1)/k^2$ ; donc  $\text{Moy}(G_n) = H_n$  et  $\text{Var}(G_n) = H_n - H_n^{(2)}$ .

**8.44** (a) Le champion doit rester invaincu pendant  $n$  tours, donc la réponse est  $p^n$ . (b,c) Les joueurs  $x_1, \dots, x_{2^k}$  doivent être "éparpillés" (par le hasard) dans des sous-tournois différents et chacun d'eux doit gagner ses  $2^k(n-k)$  premiers matches. Il y a  $2^n!$  façons d'affecter des joueurs aux  $2^n$  feuilles de l'arbre du tournoi; il y a  $2^k!(2^{n-k})^{2^k}$  façons d'"éparpiller" de façon adéquate les  $2^k$  joueurs considérés et  $(2^n - 2^k)!$  façons de placer les autres. La probabilité cherchée est donc égale à  $(2p)^{2^k(n-k)} / \binom{2^n}{2^k}$ . Pour  $k=1$ , on trouve  $(2p^2)^{n-1} / (2^n - 1)$ . (d) A chaque déroulement possible du tournoi correspond une permutation des joueurs: soit  $y_1$  le vainqueur,  $y_2$  l'autre finaliste,  $y_3$  et  $y_4$  ceux qui ont perdu contre  $y_1$  et  $y_2$  respectivement en demi-finale,  $(y_5, \dots, y_8)$  ceux qui ont perdu contre  $(y_1, \dots, y_4)$  respectivement en quart de finale, etc. Autre preuve possible: le premier tour admet  $2^n! / 2^{n-1}!$  déroulements possibles, le second tour en admet  $2^{n-1}! / 2^{n-2}!$ , et ainsi de suite. (e) Soit  $S_k$  l'ensemble des  $2^{k-1}$  adversaires possibles de  $x_2$  au kième tour. La probabilité que  $x_2$  gagne sachant que  $x_1$  appartient à  $S_k$  est

$$\begin{aligned} \Pr(x_1 \text{ joue contre } x_2) \cdot p^{n-1}(1-p) + \Pr(x_1 \text{ ne joue pas contre } x_2) \cdot p^n \\ = p^{k-1}p^{n-1}(1-p) + (1-p^{k-1})p^n. \end{aligned}$$

La probabilité que  $x_1 \in S_k$  est égale à  $2^{k-1} / (2^n - 1)$ ; on trouve la réponse en sommant sur  $k$ :

$$\sum_{k=1}^n \frac{2^{k-1}}{2^n - 1} (p^{k-1}p^{n-1}(1-p) + (1-p^{k-1})p^n) = p^n - \frac{(2p)^n - 1}{2^n - 1} p^{n-1}.$$

(f) Chacun des  $2^n!$  déroulements possibles du tournoi s'effectue avec une certaine probabilité. La probabilité que  $x_j$  gagne est égale à la somme des probabilités des  $(2^n - 1)!$  déroulements dans lesquels il gagne. Supposons qu'on permute  $x_j$  et  $x_{j+1}$  dans tous ces déroulements. Cela n'affecte pas la probabilité si  $x_j$  et  $x_{j+1}$  ne se rencontrent pas, mais cela la multiplie par  $(1-p)/p < 1$  dans le cas contraire.

**8.45** (a)  $A(z) = 1/(3-2z)$ ;  $B(z) = zA(z)^2$ ;  $C(z) = z^2A(z)^3$ . La fgp du xérès lorsqu'il est mis en bouteilles est  $z^3A(z)^3$ ; c'est  $z^3$  fois la loi binomiale négative de paramètres  $n=3$  et  $p=\frac{1}{3}$ . (b)  $\text{Moy}(A) = 2$ ,  $\text{Var}(A) = 6$ ;  $\text{Moy}(B) = 5$ ,  $\text{Var}(B) = 2\text{Var}(A) = 12$ ;  $\text{Moy}(C) = 8$ ,  $\text{Var}(C) = 18$ . L'âge moyen du xérès est neuf ans. La fraction qui a 25 ans est  $\binom{-3}{22}(-2)^{22}3^{-25} = \binom{24}{22}2^{22}3^{-25} = 23 \cdot (\frac{2}{3})^{24} \approx 0,00137$ . (c) Dans les fgp doubles suivantes, le

coefficient de  $w^n$  constitue la fgp du début de l'année  $n$  :

$$A = (1 + \frac{1}{3}w/(1-w))/(1 - \frac{2}{3}zw);$$

$$B = (1 + \frac{1}{3}zwA)/(1 - \frac{2}{3}zw);$$

$$C = (1 + \frac{1}{3}zwB)/(1 - \frac{2}{3}zw).$$

En dérivant par rapport à  $z$  et en posant  $z = 1$ , on trouve

$$C' = \frac{8}{1-w} - \frac{1/2}{(1 - \frac{2}{3}w)^3} - \frac{3/2}{(1 - \frac{2}{3}w)^2} - \frac{6}{1 - \frac{2}{3}w}.$$

L'âge moyen du xérès mis en bouteilles  $n$  années après le début du processus est égal à 1 plus le coefficient de  $w^{n-1}$ , donc  $9 - (\frac{2}{3})^n(3n^2 + 21n + 72)/8$  (c'est déjà plus grand que 8 pour  $n = 11$ ).

**8.46** (a)  $P(w, z) = 1 + \frac{1}{2}(wP(w, z) + zP(w, z)) = (1 - \frac{1}{2}(w+z))^{-1}$ , donc  $p_{mn} = 2^{-m-n} \binom{m+n}{n}$ . (b)  $P_k(w, z) = \frac{1}{2}(w^k + z^k)P(w, z)$ ; donc

$$p_{k,m,n} = 2^{k-1-m-n} \left( \binom{m+n-k}{m} + \binom{m+n-k}{n} \right).$$

(c)  $\sum_k kp_{k,n,n} = \sum_{k=0}^n k2^{k-2n} \binom{2n-k}{n} = \sum_{k=0}^n (n-k)2^{-n-k} \binom{n+k}{n}$ ; pour sommer, on utilise (5.20) :

$$\begin{aligned} & \sum_{k=0}^n 2^{-n-k} \left( (2n+1) \binom{n+k}{n} - (n+1) \binom{n+1+k}{n+1} \right) \\ &= (2n+1) - (n+1)2^{-n} \left( 2^{n+1} - 2^{-n-1} \binom{2n+2}{n+1} \right) \\ &= \frac{2n+1}{2^{2n}} \binom{2n}{n} - 1. \end{aligned}$$

(Avec les méthodes du chapitre 9, on peut montrer que cela fait  $2\sqrt{n/\pi} - 1 + O(n^{-1/2})$ ).

**8.47** Après  $n$  irradiations, il y a  $n+2$  récepteurs. Soit  $X_n$  la variable aléatoire qui représente le nombre de diphages. Alors  $X_{n+1} = X_n + Y_n$ , où  $Y_n = -1$  si la  $(n+1)$ ième particule est absorbée par un diphage (avec probabilité conditionnelle  $2X_n/(n+2)$ ), et  $Y_n = +2$  sinon. Par conséquent,

$$EX_{n+1} = EX_n + EY_n = EX_n - 2EX_n/(n+2) + 2(1 - 2EX_n/(n+2)).$$

Pour résoudre la récurrence  $(n+2)EX_{n+1} = (n-4)EX_n + 2n+4$ , on peut multiplier les deux membres par le facteur de sommation  $(n+1)^5$ ; on peut aussi deviner la réponse et la démontrer par induction :  $EX_n = (2n+4)/7$  pour tout  $n > 4$ . (Remarquez que, quelle que soit la configuration après la

quatrième particule, il y a toujours deux diphages et un triphage après la cinquième).

**8.48** (a) La distance entre les deux frisbees (mesurée du côté où elle est paire) vaut soit 0, soit 2, soit 4 ; au départ, elle est égale à 4. Les fonctions génératrices correspondantes A, B et C (où, par exemple,  $[z^n] C$  est la probabilité que la distance soit égale à 4 après n étapes) satisfont

$$A = \frac{1}{4}zB, \quad B = \frac{1}{2}zB + \frac{1}{4}zC, \quad C = 1 + \frac{1}{4}zB + \frac{3}{4}zC.$$

Il s'ensuit que  $A = z^2/(16 - 20z + 5z^2) = z^2/F(z)$ , et on a  $Moy(A) = 2 - Moy(F) = 12$ ,  $Var(A) = -Var(F) = 100$ .

Voici une autre solution, plus difficile mais plus intéressante. La fgp A se factorise comme suit :

$$A = \frac{p_1 z}{1 - q_1 z} \cdot \frac{p_2 z}{1 - q_2 z} = \frac{p_2}{p_2 - p_1} \frac{p_1 z}{1 - q_1 z} + \frac{p_1}{p_1 - p_2} \frac{p_2 z}{1 - q_2 z},$$

avec

$$p_1 = \phi^2/4 = (3 + \sqrt{5})/8,$$

$$p_2 = \bar{\phi}^2/4 = (3 - \sqrt{5})/8,$$

$$p_1 + q_1 = p_2 + q_2 = 1.$$

Ainsi, jouer à ce jeu équivaut à jouer à pile ou face avec deux pièces dont les probabilités de faire pile sont respectivement égales à  $p_1$  et  $p_2$ , en les lançant l'une après l'autre jusqu'à ce qu'elles donnent pile toutes les deux. La distribution du nombre total de jets de pièces est exactement la même que celle du nombre de lancers de frisbee. La moyenne et la variance du nombre de jets des deux pièces valent respectivement  $6 \mp 2\sqrt{5}$  et  $50 \mp 22\sqrt{5}$ , donc la moyenne et la variance totales valent bien 12 et 100.

(b) En développant la fonction génératrice en éléments simples, on peut sommer les probabilités (notez que, comme  $\sqrt{5}/(4\phi) + \phi^2/4 = 1$ , la réponse peut s'exprimer en fonction de puissances de  $\phi$ ). Le jeu durera plus de n étapes avec la probabilité  $5^{(n-1)/2} 4^{-n} (\phi^{n+2} - \phi^{-n-2})$  ; lorsque n est pair, cela donne  $5^{n/2} 4^{-n} F_{n+2}$ . La réponse est donc  $5^{50} 4^{-100} F_{102} \approx 0,00006$ .

**8.49** (a) Si  $n > 0$ ,  $P_N(0, n) = \frac{1}{2}[N=0] + \frac{1}{4}P_{N-1}(0, n) + \frac{1}{4}P_{N-1}(1, n-1)$  ; l'expression de  $P_N(m, 0)$  est similaire ;  $P_N(0, 0) = [N=0]$ . Par conséquent,

$$g_{m,n} = \frac{1}{4}zg_{m-1,n+1} + \frac{1}{2}zg_{m,n} + \frac{1}{4}zg_{m+1,n-1};$$

$$g_{0,n} = \frac{1}{2} + \frac{1}{4}zg_{0,n} + \frac{1}{4}g_{1,n-1}; \quad \text{etc.}$$

(b)  $g'_{m,n} = 1 + \frac{1}{4}g'_{m-1,n+1} + \frac{1}{2}g'_{m,n} + \frac{1}{4}g'_{m+1,n-1}$  ;  $g'_{0,n} = \frac{1}{2} + \frac{1}{4}g'_{0,n} + \frac{1}{4}g'_{1,n-1}$  ; etc. Par induction sur m, on a  $g'_{m,n} = (2m+1)g'_{0,m+n} - 2m^2$

pour tous  $m, n \geq 0$ . Comme  $g'_{m,0} = g'_{0,m}$ , on a aussi  $g'_{m,n} = m + n + 2mn$ .

(c) La récurrence est satisfaite lorsque  $mn > 0$  parce que

$$\begin{aligned}\sin(2m+1)\theta &= \frac{1}{\cos^2 \theta} \left( \frac{\sin(2m-1)\theta}{4} \right. \\ &\quad \left. + \frac{\sin(2m+1)\theta}{2} + \frac{\sin(2m+3)\theta}{4} \right).\end{aligned}$$

C'est une conséquence de l'identité  $\sin(x-y) + \sin(x+y) = 2 \sin x \cos y$ . Il ne reste donc plus qu'à vérifier les conditions aux bornes.

**8.50** (a) En suivant le conseil donné, on trouve

$$\begin{aligned}3(1-z)^2 \sum_k \binom{1/2}{k} \left(\frac{8}{9}z\right)^k (1-z)^{2-k} \\ = 3(1-z)^2 \sum_k \binom{1/2}{k} \left(\frac{8}{9}\right)^k \sum_j \binom{k+j-3}{j} z^{j+k};\end{aligned}$$

regardez alors le coefficient de  $z^{3+1}$ . (b)  $H(z) = \frac{2}{3} + \frac{5}{27}z + \frac{1}{2} \sum_{l \geq 0} c_{3+l} z^{2+l}$ .

(c) Soit  $r = \sqrt{(1-z)(9-z)}$ . On peut montrer que  $(z-3+r)(z-3-r) = 4z$ ,

et donc que  $(r/(1-z)+2)^2 = (13-5z+4r)/(1-z) = (9-H(z))/(1-H(z))$ .

(d) Le calcul de la dérivée en  $z = 1$  montre que  $\text{Moy}(H) = 1$ . Comme la dérivée seconde diverge en  $z = 1$ , on en conclut que la variance est infinie.

**8.51** (a) Soit  $H_n(z)$  la fgp de la somme que vous possédez après  $n$  tours, avec  $H_0(z) = z$ . Comme la distribution pour  $n$  tours satisfait

$$H_{n+1}(z) = H_n(H(z)),$$

on peut prouver le résultat par induction (en utilisant l'étonnante identité du problème précédent). (b)  $g_n = H_n(0) - H_{n-1}(0) = 4/n(n+1)(n+2) = 4(n-1)\underline{-3}$ . La moyenne vaut 2 et la variance est infinie. (c) D'après l'exercice 15, le nombre moyen de tickets que vous aurez achetés au  $n$ ième tour est  $\text{Moy}(H_n) = 1$ . Par conséquent, le nombre moyen total de tickets achetés est infini. (Ainsi, vous êtes pratiquement sûr de perdre un jour, en moyenne au bout de 2 tours seulement, mais vous achèterez en moyenne un nombre infini de tickets). (d) Maintenant, la fgp après  $n$  tours est  $H_n(z)^2$ , et la méthode de la question (b) aboutit à une moyenne de  $16 - \frac{4}{3}\pi^2 \approx 2,8$  (la somme  $\sum_{k \geq 1} 1/k^2 = \pi^2/6$  apparaît dans ce calcul).

**8.52** Si  $\omega$  et  $\omega'$  sont des événements tels que  $\Pr(\omega) > \Pr(\omega')$ , alors la probabilité est grande de trouver  $\omega$  plus souvent que  $\omega'$  dans une suite de  $n$  expériences indépendantes, car  $\omega$  apparaît à peu près  $n\Pr(\omega)$  fois. Par conséquent, lorsque  $n \rightarrow \infty$ , la probabilité que la médiane (resp. le

mode) des valeurs de  $X$  dans une suite de valeurs aléatoires soit égal(e) à la médiane (resp. au mode) de la variable aléatoire  $X$  tend vers 1.

**8.53** On peut montrer que c'est faux même dans le cas où chaque variable ne peut prendre que les valeurs 0 et 1. Soient  $p_0 = \Pr(X=Y=Z=0)$ ,  $p_1 = \Pr(X=Y=\bar{Z}=0)$ , ...,  $p_7 = \Pr(\bar{X}=\bar{Y}=\bar{Z}=0)$ , où  $\bar{X} = 1 - X$ . Alors  $p_0 + p_1 + \dots + p_7 = 1$ , et les trois variables sont indépendantes deux à deux si et seulement si

$$\begin{aligned}(p_4 + p_5 + p_6 + p_7)(p_2 + p_3 + p_6 + p_7) &= p_6 + p_7, \\ (p_4 + p_5 + p_6 + p_7)(p_1 + p_3 + p_5 + p_7) &= p_5 + p_7, \\ (p_2 + p_3 + p_6 + p_7)(p_1 + p_3 + p_5 + p_7) &= p_3 + p_7.\end{aligned}$$

Or,  $\Pr(X+Y=Z=0) \neq \Pr(X+Y=0)\Pr(Z=0) \iff p_0 \neq (p_0 + p_1)(p_0 + p_2 + p_4 + p_6)$ . Voici une solution :

$$p_0 = p_3 = p_5 = p_6 = 1/4; \quad p_1 = p_2 = p_4 = p_7 = 0.$$

Cela revient à tirer à pile ou face avec deux pièces et prendre  $X =$  (la première tombe sur pile),  $Y =$  (la seconde tombe sur pile),  $Z =$  (les résultats diffèrent). Voici un autre exemple tel que toutes les probabilités sont non nulles :

$$\begin{aligned}p_0 &= 4/64, \quad p_1 = p_2 = p_4 = 5/64, \\ p_3 &= p_5 = p_6 = 10/64, \quad p_7 = 15/64.\end{aligned}$$

C'est pour cela que, par définition, on dit que  $n$  variables  $X_1, \dots, X_n$  sont indépendantes si

$$\Pr(X_1=x_1 \text{ et } \dots \text{ et } X_n=x_n) = \Pr(X_1=x_1) \dots \Pr(X_n=x_n);$$

l'indépendance deux à deux ne suffit pas pour garantir ce résultat.

**8.54** (Voir l'exercice 27 pour les notations). On a

$$\begin{aligned}\mathbb{E}(\Sigma_2^2) &= n\mu_4 + n(n-1)\mu_2^2; \\ \mathbb{E}(\Sigma_2\Sigma_1^2) &= n\mu_4 + 2n(n-1)\mu_3\mu_1 + n(n-1)\mu_2^2 + n(n-1)(n-2)\mu_2\mu_1^2; \\ \mathbb{E}(\Sigma_1^4) &= n\mu_4 + 4n(n-1)\mu_3\mu_1 + 3n(n-1)\mu_2^2 \\ &\quad + 6n(n-1)(n-2)\mu_2\mu_1^2 + n(n-1)(n-2)(n-3)\mu_1^4;\end{aligned}$$

il s'ensuit que  $V(\widehat{V}X) = \kappa_4/n + 2\kappa_2^2/(n-1)$ .

**8.55** Il y a  $A = \frac{1}{17} \cdot 52!$  permutations telles que  $X = Y$ , et  $B = \frac{16}{17} \cdot 52!$  permutations telles que  $X \neq Y$ . A la fin du processus, chaque permutation telle que  $X = Y$  apparaît avec probabilité  $\frac{1}{17}/((1 - \frac{16}{17}p)A)$ , car on revient

à l'étape S1 avec probabilité  $\frac{16}{17}p$ . De même, chaque permutation telle que  $X \neq Y$  apparaît avec probabilité  $\frac{16}{17}(1-p)/((1-\frac{16}{17}p)B)$ . En prenant  $p = \frac{1}{4}$ , on a  $\Pr(X=x \text{ et } Y=y) = \frac{1}{169}$  pour tout  $x$  et tout  $y$  (on pourrait donc tirer deux fois à pile ou face avec une pièce juste et revenir en S1 si on obtenait deux fois pile).

**8.56** Si  $m$  est pair, les frisbees restent toujours à distance impaire l'un de l'autre et le jeu ne s'arrête jamais. Si  $m = 2l + 1$ , voici les fonctions génératrices qui nous intéressent :

$$\begin{aligned} G_m &= \frac{1}{4}zA_1; \\ A_1 &= \frac{1}{2}zA_1 + \frac{1}{4}zA_2, \\ A_k &= \frac{1}{4}zA_{k-1} + \frac{1}{2}zA_k + \frac{1}{4}zA_{k+1}, \quad \text{pour } 1 < k < l, \\ A_l &= \frac{1}{4}zA_{l-1} + \frac{3}{4}zA_l + 1. \end{aligned}$$

(Le coefficient  $[z^n]A_k$  représente la probabilité que la distance entre les frisbees soit égale à  $2k$  après  $n$  lancers). Inspirons-nous de l'exercice 49 et posons  $z = 1/\cos^2 \theta$  et  $A_1 = X \sin 2\theta$ , où la valeur de  $X$  reste à trouver. On en déduit par induction (sans utiliser l'équation de  $A_l$ ) que  $A_k = X \sin 2k\theta$ . Par conséquent, il nous faut choisir  $X$  de sorte que

$$\left(1 - \frac{3}{4 \cos^2 \theta}\right)X \sin 2l\theta = 1 + \frac{1}{4 \cos^2 \theta}X \sin(2l-2)\theta.$$

On trouve  $X = 2 \cos^2 \theta / \sin \theta \cos(2l+1)\theta$ , donc

$$G_m = \frac{\cos \theta}{\cos m\theta}.$$

Le dénominateur s'annule lorsque  $\theta$  est un multiple impair de  $\pi/(2m)$  ; ainsi, pour tout  $1 \leq k \leq l$ ,  $1 - q_k z$  est une racine du dénominateur et la fonction génératrice peut bien s'exprimer comme le produit donné dans l'énoncé. Pour trouver la moyenne et la variance, on peut écrire

$$\begin{aligned} G_m &= (1 - \frac{1}{2}\theta^2 + \frac{1}{24}\theta^4 - \dots)/(1 - \frac{1}{2}m^2\theta^2 + \frac{1}{24}m^4\theta^4 - \dots) \\ &= 1 + \frac{1}{2}(m^2 - 1)\theta^2 + \frac{1}{24}(5m^4 - 6m^2 + 1)\theta^4 + \dots \\ &= 1 + \frac{1}{2}(m^2 - 1)(\tan \theta)^2 + \frac{1}{24}(5m^4 - 14m^2 + 9)(\tan \theta)^4 + \dots \\ &= 1 + G'_m(1)(\tan \theta)^2 + \frac{1}{2}G''_m(1)(\tan \theta)^4 + \dots, \end{aligned}$$

car  $\tan^2 \theta = z - 1$  et  $\tan \theta = \theta + \frac{1}{3}\theta^3 + \dots$ . On a donc

$$\begin{aligned} \text{Moy}(G_m) &= \frac{1}{2}(m^2 - 1); \\ \text{Var}(G_m) &= \frac{1}{6}m^2(m^2 - 1). \end{aligned}$$

*Encore une victoire de la trigonométrie.  
Peut-on interpréter cela par des pièces de monnaies qu'on lance dans les angles du polygone ?*

Remarquez que cela entraîne les identités

$$\frac{m^2 - 1}{2} = \sum_{k=1}^{(m-1)/2} \frac{1}{p_k} = \sum_{k=1}^{(m-1)/2} \left(1 / \sin \frac{(2k-1)\pi}{2m}\right)^2;$$

$$\frac{m^2(m^2 - 1)}{6} = \sum_{k=1}^{(m-1)/2} \left(\cot \frac{(2k-1)\pi}{2m} / \sin \frac{(2k-1)\pi}{2m}\right)^2.$$

Le troisième cumulant de cette distribution est égal à  $\frac{1}{30}m^2(m^2 - 1)(4m^2 - 1)$  ; cependant, les suivants ne se factorisent pas aussi bien. Il existe une façon beaucoup plus simple de trouver la moyenne : on a  $G_m + A_1 + \dots + A_l = z(A_1 + \dots + A_l) + 1$  ; donc, si  $z = 1$ ,  $G'_m = A_1 + \dots + A_l$ . Comme  $G_m = 1$  lorsque  $z = 1$ , on montre aisément par induction que  $A_k = 4k$ .

**8.57** On a  $A:A \geq 2^{l-1}$ ,  $B:B < 2^{l-1} + 2^{l-3}$  et  $B:A \geq 2^{l-2}$  ; l'inégalité  $B:B - B:A \geq A:A - A:B$  n'est donc possible que si  $A:B > 2^{l-3}$ . Cela signifie que  $\bar{\tau}_2 = \tau_3$ ,  $\tau_1 = \tau_4$ ,  $\tau_2 = \tau_5$ , ...,  $\tau_{l-3} = \tau_l$ . Mais alors  $A:A \approx 2^{l-1} + 2^{l-4} + \dots$ ,  $A:B \approx 2^{l-3} + 2^{l-6} + \dots$ ,  $B:A \approx 2^{l-2} + 2^{l-5} + \dots$  et  $B:B \approx 2^{l-1} + 2^{l-4} + \dots$  ; donc, finalement,  $B:B - B:A$  est inférieur à  $A:A - A:B$ . (Guibas et Odlyzko [168] trouvent des résultats plus forts ; ils montrent que les chances de Bill sont toujours majorées par  $P\tau_1 \dots \tau_{l-1}$  ou par  $F\tau_1 \dots \tau_{l-1}$  ; en fait, la stratégie qui donne à Bill les meilleures chances de gagner est unique ; voyez l'exercice suivant).

**8.58** (Solution de J. Csirik). Si la suite  $A$  est égale à  $P^l$  ou à  $F^l$ , l'une des deux autres suites ne peut pas être utilisée. Sinon, soient  $\hat{A} = \tau_1 \dots \tau_{l-1}$ ,  $P = P\hat{A}$  et  $F = F\hat{A}$ . Il n'est pas difficile de vérifier que  $P:A = F:A = \hat{A}:\hat{A}$ ,  $P:P + F:F = 2^{l-1} + 2(\hat{A}:\hat{A}) + 1$  et  $A:P + A:F = 1 + 2(A:A) - 2^l$ . Par conséquent, l'équation

$$\frac{P:P - P:A}{A:A - A:P} = \frac{F:F - F:A}{A:A - A:F}$$

implique que les deux fractions sont égales à

$$\frac{P:P - P:A + F:F - F:A}{A:A - A:P + A:A - A:F} = \frac{2^{l-1} + 1}{2^l - 1}.$$

On peut alors réorganiser les fractions d'origine pour montrer que

$$\frac{P:P - P:A}{F:F - F:A} = \frac{A:A - A:P}{A:A - A:F} = \frac{p}{q},$$

avec  $p \perp q$ . De plus  $(p+1) \setminus \text{pgcd}(2^{l-1} + 1, 2^l - 1) = \text{pgcd}(3, 2^l - 1)$  ; nous pouvons donc supposer que  $l$  est pair et que  $p = 1$ ,  $q = 2$ . Il s'ensuit que  $A:A - A:P = (2^l - 1)/3$  et  $A:A - A:F = (2^{l+1} - 2)/3$ , donc  $A:P - A:F =$

$(2^l - 1)/3 \geq 2^{l-2}$ . On a  $A:P \geq 2^{l-2}$  si et seulement si  $A = (FP)^{l/2}$ . Alors  $P:P - P:A = A:A - A:P$ , donc  $2^{l-1} + 1 = 2^l - 1$  et  $l = 2$ .

Csirik [69] continue en montrant que, si  $l \geq 4$ , ce qu'Alice peut faire de mieux est de jouer  $PF^{l-3}P^2$ . Cependant, même avec cette stratégie, Bill gagnera avec une probabilité à peu près égale à  $\frac{2}{3}$ .

**8.59** Selon (8.82), il faut que  $B:B - B:A > A:A - A:B$ . Voici une solution :  $A = FFPP$ ,  $B = PPP$ .

**8.60** (a) Il y a deux cas, selon que  $h_k \neq h_n$  ou  $h_k = h_n$  :

$$\begin{aligned} G(w, z) &= \frac{m-1}{m} \left( \frac{m-2+w+z}{m} \right)^{k-1} w \left( \frac{m-1+z}{m} \right)^{n-k-1} z \\ &\quad + \frac{1}{m} \left( \frac{m-1+wz}{m} \right)^{k-1} wz \left( \frac{m-1+z}{m} \right)^{n-k-1} z. \end{aligned}$$

(b) On peut raisonner algébriquement, en calculant les dérivées partielles de  $G(w, z)$  par rapport à  $w$  et  $z$  et en posant  $w = z = 1$ . On peut aussi voir les choses de façon combinatoire : quelles que soient les valeurs de  $h_1, \dots, h_{n-1}$ , la valeur moyenne de  $P(h_1, \dots, h_{n-1}, h_n; n)$  ne change pas (la moyenne est prise sur les  $h_n$ ), car la suite  $(h_1, \dots, h_{n-1})$  détermine une suite de longueurs de listes  $(n_1, n_2, \dots, n_m)$  telle que la valeur moyenne considérée est égale à  $((n_1+1)+(n_2+1)+\dots+(n_m+1))/m = (n-1+m)/m$ . Par conséquent, la variable aléatoire  $E P(h_1, \dots, h_n; n)$  est indépendante de  $(h_1, \dots, h_{n-1})$ , donc indépendante de  $P(h_1, \dots, h_n; k)$ .

**8.61** Si  $1 \leq k < l \leq n$ , l'exercice précédent montre que le coefficient de  $s_k s_l$  dans la variance de la moyenne est nul. Par conséquent, il nous suffit de considérer le coefficient de  $s_k^2$ , qui est égal à

$$\sum_{1 \leq h_1, \dots, h_n \leq m} \frac{P(h_1, \dots, h_n; k)^2}{m^n} - \left( \sum_{1 \leq h_1, \dots, h_n \leq m} \frac{P(h_1, \dots, h_n; k)}{m^n} \right)^2,$$

la variance de  $((m-1+z)/m)^{k-1}z$ . Elle est égale à  $(k-1)(m-1)/m^2$ , comme nous l'avons vu dans l'exercice 30.

**8.62** La fgp  $D_n(z)$  satisfait la récurrence

$$\begin{aligned} D_0(z) &= z; \\ D_n(z) &= z^2 D_{n-1}(z) + 2(1-z^3)D'_{n-1}(z)/(n+1), \quad \text{pour } n > 0. \end{aligned}$$

Nous pouvons en déduire la récurrence

$$D''_n(1) = (n-11)D''_{n-1}(1)/(n+1) + (8n-2)/7,$$

qui admet la solution  $\frac{2}{637}(n+2)(26n+15)$  pour tout  $n \geq 11$  (quelles que soient les conditions initiales). Par conséquent, la variance est égale à  $\frac{108}{637}(n+2)$  pour tout  $n \geq 11$ .

**8.63** (Un autre problème consiste à se demander si une suite donnée de candidats cumulants provient bien d'une distribution de probabilité ; par exemple,  $\kappa_2$  doit être positif ou nul et  $\kappa_4 + 3\kappa_2^2 = E((X - \mu)^4)$  doit être supérieur ou égal à  $(E((X - \mu)^2))^2 = \kappa_2^2$ , etc ; Hamburger [6], [175] a donné une condition nécessaire et suffisante pour ce problème-ci).

**9.1** C'est vrai si toutes les fonctions sont positives. Sinon, on peut avoir, par exemple,  $f_1(n) = n^3 + n^2$ ,  $f_2(n) = -n^3$ ,  $g_1(n) = n^4 + n$ ,  $g_2(n) = -n^4$ .

**9.2** (a) On a  $n^{\ln n} \prec c^n \prec (\ln n)^n$ , car  $(\ln n)^2 \prec n \ln c \prec n \ln \ln n$ .  
(b)  $n^{\ln \ln \ln n} \prec (\ln n)! \prec n^{\ln \ln n}$ . (c) Appliquez le logarithme pour montrer que c'est  $(n!)!$  qui gagne. (d)  $F_{[H_n]}^2 \asymp \phi^{2 \ln n} = n^{2 \ln \phi}$  ; c'est  $H_{F_n} \sim n \ln \phi$  qui gagne parce que  $\phi^2 = \phi + 1 < e$ .

**9.3** Lorsqu'on remplace  $kn$  par  $O(n)$ , il faut considérer un  $C$  différent pour chaque  $k$ , mais à chaque  $O$  correspond un seul  $C$ . En fait, dans le contexte de l'exercice, le  $O$  doit représenter un ensemble de fonctions de deux variables  $k$  et  $n$ . Voici ce qu'on a le droit d'écrire :  $\sum_{k=1}^n kn = \sum_{k=1}^n O(n^2) = O(n^3)$ .

**9.4** Par exemple,  $\lim_{n \rightarrow \infty} O(1/n) = 0$ . Dans le membre gauche,  $O(1/n)$  est l'ensemble des fonctions  $f(n)$  pour lesquelles il existe des constantes  $C$  et  $n_0$  telles que  $|f(n)| \leq C/n$  pour tout  $n \geq n_0$ . Comme la limite de toutes les fonctions de cet ensemble vaut zéro, le membre gauche est égal au singleton  $\{0\}$ . Dans le membre droit, il n'y a pas de variable ;  $0$  représente  $\{0\}$ , l'ensemble (singleton) des "fonctions de zéro variable, dont la valeur est zéro". (Si vous ne voyez pas la logique de tout cela, attendez quelques mois et relisez de nouveau cet exercice).

**9.5** Soit  $f(n) = n^2$  et  $g(n) = 1$  ; alors  $n$  appartient à l'ensemble de gauche mais pas à celui de droite. La proposition est donc fausse.

**9.6**  $n \ln n + \gamma n + O(\sqrt{n} \ln n)$ .

**9.7**  $(1 - e^{-1/n})^{-1} = nB_0 - B_1 + B_2 n^{-1}/2! + \dots = n + \frac{1}{2} + O(n^{-1})$ .

**9.8** Soient  $f(n) = \lfloor n/2 \rfloor^2 + n$  et  $g(n) = (\lceil n/2 \rceil - 1)! \lceil n/2 \rceil! + n$ . Notons en passant que ces fonctions satisfont  $f(n) = O(ng(n))$  et  $g(n) = O(nf(n))$  ; on peut bien sûr trouver des exemples tout à fait différents.

**9.9** (Pour que le propos soit plus général, nous ajoutons la condition  $n \rightarrow \infty$  à l'équation ; ceci implique l'existence de deux constantes pour chaque  $O$ ). Toute fonction du membre gauche est de la forme  $a(n) + b(n)$ , et il existe des constantes  $m_0$ ,  $B$ ,  $n_0$  et  $C$  telles que  $|a(n)| \leq B|f(n)|$  pour  $n \geq m_0$  et  $|b(n)| \leq C|g(n)|$  pour  $n \geq n_0$ . Par conséquent, cette fonction vaut au plus  $\max(B, C)(|f(n)| + |g(n)|)$ , pour  $n \geq \max(m_0, n_0)$ . Elle appartient donc au membre droit.

**9.10** Soit une fonction  $g(x)$  qui appartient au membre gauche. Il existe alors  $y$  tel que  $g(x) = \cos y$ , et  $C$  tel que  $|y| \leq C|x|$ . Par conséquent,  $0 \leq 1 - g(x) = 2 \sin^2(y/2) \leq \frac{1}{2}y^2 \leq \frac{1}{2}C^2x^2$ . L'ensemble de gauche est donc contenu dans celui de droite, et la formule est vraie.

**9.11** La proposition est vraie. En effet, si  $|x| \leq |y|$ , alors  $(x+y)^2 \leq 4y^2$ . Ainsi,  $(x+y)^2 = O(x^2) + O(y^2)$ . Par conséquent,  $O(x+y)^2 = O((x+y)^2) = O(O(x^2) + O(y^2)) = O(O(x^2)) + O(O(y^2)) = O(x^2) + O(y^2)$ .

**9.12**  $1+2/n+O(n^{-2}) = (1+2/n)(1+O(n^{-2})/(1+2/n))$  d'après (9.26), et  $1/(1+2/n) = O(1)$ ; utilisez alors (9.26).

**9.13**  $n^n(1+2n^{-1}+O(n^{-2}))^n = n^n \exp(n(2n^{-1}+O(n^{-2}))) = e^{2n} + O(n^{n-1})$ .

**9.14** C'est égal à  $n^{n+\beta} \exp((n+\beta)(\alpha/n - \frac{1}{2}\alpha^2/n^2 + O(n^{-3})))$ .

**9.15**  $\ln \binom{3n}{n,n,n} = 3n \ln 3 - \ln n + \frac{1}{2} \ln 3 - \ln 2\pi + (\frac{1}{36} - \frac{1}{4})n^{-1} + O(n^{-3})$ , et voici donc la réponse :

$$\frac{3^{3n+1/2}}{2\pi n} \left(1 - \frac{2}{9}n^{-1} + \frac{2}{81}n^{-2} + O(n^{-3})\right).$$

**9.16** Soit  $l$  un entier tel que  $a \leq l < b$ . Alors

$$\begin{aligned} \int_0^1 B(x)f(l+x) dx &= \int_{1/2}^1 B(x)f(l+x) dx - \int_0^{1/2} B(1-x)f(l+x) dx \\ &= \int_{1/2}^1 B(x)(f(l+x) - f(l+1-x)) dx. \end{aligned}$$

Comme  $l+x \geq l+1-x$  lorsque  $x \geq \frac{1}{2}$ , cette intégrale est strictement positive lorsque  $f(x)$  est croissante.

**9.17**  $\sum_{m \geq 0} B_m (\frac{1}{2}) z^m / m! = ze^{z/2}/(e^z - 1) = z/(e^{z/2} - 1) - z/(e^z - 1)$ .

**9.18** Le calcul effectué dans le chapitre pour le cas  $\alpha = 1$  peut être généralisé :

$$\begin{aligned} b_k(n) &= \frac{2^{(2n+1/2)\alpha}}{(2\pi n)^{\alpha/2}} e^{-k^2\alpha/n}, \\ c_k(n) &= 2^{2n\alpha} n^{-(1+\alpha)/2+3\epsilon} e^{-k^2\alpha/n}; \end{aligned}$$

la réponse est donc  $2^{2n\alpha} (\pi n)^{(1-\alpha)/2} \alpha^{-1/2} (1 + O(n^{-1/2+3\epsilon}))$ .

**9.19**  $H_{10} = 2,928968254 \approx 2,928968256$ ;  $10! = 3628800 \approx 3628712,4$ ;  $B_{10} = 0,075757576 \approx 0,075757494$ ;  $\pi(10) = 4 \approx 10,0017845$ ;  $e^{0,1} = 1,10517092 \approx 1,10517083$ ;  $\ln 1,1 = 0,0953102 \approx 0,0953083$ ;  $1,111111 \approx 1,1111000$ ;  $1,1^{0,1} = 1,00957658 \approx 1,00957643$ . (L'approximation de  $\pi(n)$  donne plus de chiffres significatifs lorsque  $n$  est plus grand : par exemple,  $\pi(10^9) = 50847534 \approx 50840742$ ).

*(Il est intéressant de comparer cette formule avec le résultat correspondant pour le coefficient binomial, dans l'exercice 9.60).*

**9.20** (a) Oui ; le membre gauche est en  $o(n)$  et le membre droit est équivalent à  $O(n)$ . (b) Oui ; le membre gauche est égal à  $e \cdot e^{O(1/n)}$ . (c) Non ; le membre gauche est à peu près égal à  $\sqrt{n}$  fois la borne de droite.

**9.21** On a  $P_n = p = n(\ln p - 1 - 1/\ln p + O(1/\log n)^2)$ , avec

$$\begin{aligned}\ln p &= \ln n + \ln \ln p - 1/\ln n + \ln \ln n / (\ln n)^2 + O(1/\log n)^2; \\ \ln \ln p &= \ln \ln n + \frac{\ln \ln n}{\ln n} - \frac{(\ln \ln n)^2}{2(\ln n)^2} + \frac{\ln \ln n}{(\ln n)^2} + O(1/\log n)^2.\end{aligned}$$

Il s'ensuit que

$$\begin{aligned}P_n &= n \left( \ln n + \ln \ln n - 1 \right. \\ &\quad \left. + \frac{\ln \ln n - 2}{\ln n} - \frac{(\ln \ln n)^2/2 - 3 \ln \ln n}{(\ln n)^2} + O(1/\log n)^2 \right).\end{aligned}$$

(Il est possible d'obtenir une approximation un peu meilleure, dans laquelle le  $O(1/\log n)^2$  est remplacé par  $-5.5/(\ln n)^2 + O(\log \log n/\log n)^3$  ; on trouve alors  $P_{1000000} \approx 15480992.8$ ).

*Que dit un mathématicien quand il se noie ?*

*log log log log ...*

**9.22** Remplacez  $O(n^{-2k})$  par  $-\frac{1}{12}n^{-2k} + O(n^{-4k})$  dans le développement de  $H_{nk}$ . Cela a pour effet de transformer le  $O(\Sigma_3(n^2))$  de (9.53) en  $-\frac{1}{12}\Sigma_3(n^2) + O(\Sigma_3(n^4))$ . On a

$$\Sigma_3(n) = \frac{3}{4}n^{-1} + \frac{5}{36}n^{-2} + O(n^{-3}),$$

donc le terme  $O(n^{-2})$  de (9.54) peut être remplacé par  $-\frac{19}{144}n^{-2} + O(n^{-3})$ .

**9.23**  $nh_n = \sum_{0 \leq k < n} h_k/(n-k) + 2cH_n/(n+1)(n+2)$ . Prenez  $c = e^{\pi^2/6} = \sum_{k \geq 0} g_k$ , de sorte que  $\sum_{k \geq 0} h_k = 0$  et  $h_n = O(\log n)/n^3$ . Si on développe  $\sum_{0 \leq k < n} h_k/(n-k)$  comme en (9.60), on obtient  $nh_n = 2cH_n/(n+1)(n+2) + O(n^{-2})$ , donc

$$g_n = e^{\pi^2/6} \left( \frac{n + 2 \ln n + O(1)}{n^3} \right).$$

**9.24** (a) Si  $\sum_{k \geq 0} |f(k)| < \infty$  et si  $f(n-k) = O(f(n))$  lorsque  $0 \leq k \leq n/2$ , alors

$$\sum_{k=0}^n a_k b_{n-k} = \sum_{k=0}^{n/2} O(f(k)) O(f(n)) + \sum_{k=n/2}^n O(f(n)) O(f(n-k)),$$

ce qui donne  $2O(f(n) \sum_{k \geq 0} |f(k)|)$ , donc ce cas est prouvé. (b) Par contre, si  $a_n = b_n = \alpha^{-n}$ , la convolution  $(n+1)\alpha^{-n}$  n'est pas en  $O(\alpha^{-n})$ .

**9.25**  $S_n / \binom{3n}{n} = \sum_{k=0}^n n^k / (2n+1)^k$ . On peut restreindre l'intervalle de sommation à  $0 \leq k \leq (\log n)^2$ . Dans cet intervalle,  $n^k = n^k (1 - \frac{k}{2})/n + O(k^4/n^2)$  et  $(2n+1)^k = (2n)^k (1 + \frac{k+1}{2})/2n + O(k^4/n^2)$ , donc le terme général est égal à

$$\frac{1}{2^k} \left( 1 - \frac{3k^2 - k}{4n} + O\left(\frac{k^4}{n^2}\right) \right).$$

Par conséquent, la somme sur  $k$  donne  $2 - 4/n + O(1/n^2)$ . On peut alors appliquer la formule de Stirling à  $\binom{3n}{n} = (3n)!/(2n)!n!$  pour prouver (9.2).

**9.26** Le minimum apparaît en un terme  $B_{2m}/(2m)(2m-1)n^{2m-1}$ , où  $2m \approx 2\pi n + \frac{3}{2}$ , et ce terme est approximativement égal à  $1/(\pi e^{2\pi n} \sqrt{n})$ . Par conséquent, si  $n$  est plus grand que  $e^{2\pi+1}$  à peu près, l'erreur absolue de l'estimation de  $\ln n!$  est trop importante pour que l'on puisse déterminer exactement  $n!$  par un arrondi.

**9.27** Nous pouvons supposer que  $\alpha \neq -1$ . Soit  $f(x) = x^\alpha$ ; la réponse est

$$\begin{aligned} \sum_{k=1}^n k^\alpha &= C_\alpha + \frac{n^{\alpha+1}}{\alpha+1} + \frac{n^\alpha}{2} \\ &\quad + \sum_{k=1}^m \frac{B_{2k}}{2k} \binom{\alpha}{2k-1} n^{\alpha-2k+1} + O(n^{\alpha-2m-1}). \end{aligned}$$

(La constante  $C_\alpha$  est égale à  $\zeta(-\alpha)$ , et c'est en fait ainsi qu'elle est définie lorsque  $\alpha > -1$ ).

**9.28** Plus généralement, posons  $f(x) = x^\alpha \ln x$  dans la formule de sommation d'Euler, lorsque  $\alpha \neq -1$ . En procédant comme dans l'exercice précédent, on trouve

$$\begin{aligned} \sum_{k=1}^n k^\alpha \ln k &= C'_\alpha + \frac{n^{\alpha+1} \ln n}{\alpha+1} - \frac{n^{\alpha+1}}{(\alpha+1)^2} + \frac{n^\alpha \ln n}{2} \\ &\quad + \sum_{k=1}^m \frac{B_{2k}}{2k} \binom{\alpha}{2k-1} n^{\alpha-2k+1} (\ln n + H_\alpha - H_{\alpha-2k+1}) \\ &\quad + O(n^{\alpha-2m-1} \log n); \end{aligned}$$

on peut montrer [74, §3.7] que la constante  $C'_\alpha$  est égale à  $-\zeta'(-\alpha)$ . (Le facteur  $\log n$  du terme en  $O$  peut être supprimé lorsque  $\alpha$  est un entier strictement positif  $\leq 2m$ ; dans ce cas, on peut aussi remplacer le  $k$ ème terme de la somme du membre droit par  $B_{2k} \alpha! (2k-2-\alpha)! (-1)^\alpha n^{\alpha-2k+1} / (2k)!$  lorsque  $\alpha < 2k-1$ ). Pour résoudre notre problème particulier, posons  $\alpha = 1$  et  $m = 1$ , et prenons l'exponentielle des deux membres pour obtenir

$$Q_n = A \cdot n^{n^2/2+n/2+1/12} e^{-n^2/4} (1 + O(n^{-2})) ,$$

En particulier,  
 $\zeta(0) = -1/2$   
et  $\zeta(-n) = -B_{n+1}/(n+1)$   
pour tout entier  
 $n > 0$ .

où  $A = e^{1/12 - \zeta'(-1)} \approx 1,2824271291$  est la “constante de Glaisher”.

**9.29** Soit  $f(x) = x^{-1} \ln x$ . En modifiant légèrement les calculs de l'exercice précédent, on trouve

$$\begin{aligned} \sum_{k=1}^n \frac{\ln k}{k} &= \frac{(\ln n)^2}{2} + \gamma_1 + \frac{\ln n}{2n} \\ &\quad - \sum_{k=1}^m \frac{B_{2k}}{2k} n^{-2k} (\ln n - H_{2k-1}) + O(n^{-2m-1} \log n), \end{aligned}$$

où  $\gamma_1 \approx -0,07281584548367672486$  est une “constante de Stieltjes” (voir la réponse à l'exercice 9.57). En appliquant l'exponentielle, on obtient

$$e^{\gamma_1} \sqrt{n^{\ln n}} \left( 1 + \frac{\ln n}{2n} + O\left(\frac{\log n}{n}\right)^2 \right).$$

**9.30** Soit  $g(x) = x^l e^{-x^2}$  et  $f(x) = g(x/\sqrt{n})$ . Alors  $n^{-l/2} \sum_{k \geq 0} k^l e^{-k^2/n}$  est égal à

$$\begin{aligned} \int_0^\infty f(x) dx &= \sum_{k=1}^m \frac{B_k}{k!} f^{(k-1)}(0) - (-1)^m \int_0^\infty \frac{B_m(x)}{m!} f^{(m)}(x) dx \\ &= n^{l/2} \int_0^\infty g(x) dx - \sum_{k=1}^m \frac{B_k}{k!} n^{(k-1)/2} g^{(k-1)}(0) + O(n^{-m/2}). \end{aligned}$$

Comme  $g(x) = x^l - x^{2+l}/1! + x^{4+l}/2! - x^{6+l}/3! + \dots$ , les dérivées  $g^{(m)}(x)$  s'expriment simplement et la réponse est

$$\frac{1}{2} n^{(l+1)/2} \Gamma\left(\frac{l+1}{2}\right) - \frac{B_{l+1}}{(l+1)! 0!} + \frac{B_{l+3} n^{-1}}{(l+3)! 1!} - \frac{B_{l+5} n^{-2}}{(l+5)! 2!} + O(n^{-3}).$$

**9.31** En raison de la surprenante identité  $1/(c^{m-k} + c^m) + 1/(c^{m+k} + c^m) = 1/c^m$ , la somme des termes sur  $0 \leq k \leq 2m$  est égale à  $(m + \frac{1}{2})/c^m$ . Les termes restant sont

$$\begin{aligned} \sum_{k \geq 1} \frac{1}{c^{2m+k} + c^m} &= \sum_{k \geq 1} \left( \frac{1}{c^{2m+k}} - \frac{1}{c^{3m+2k}} + \frac{1}{c^{4m+3k}} - \dots \right) \\ &= \frac{1}{c^{2m+1} - c^{2m}} - \frac{1}{c^{3m+2} - c^{3m}} + \dots, \end{aligned}$$

et cette série peut être tronquée en n'importe quel point sans que l'erreur soit supérieure au premier terme supprimé.

**9.32** D'après la formule de sommation d'Euler,  $H_n^{(2)} = \pi^2/6 - 1/n + O(n^{-2})$ , et  $H_n$  est donné par (9.89). Voici donc la réponse :

$$ne^{\gamma+\pi^2/6} \left( 1 - \frac{1}{2} n^{-1} + O(n^{-2}) \right).$$

*Les trois constantes les plus célèbres,  $e$ ,  $\pi$  et  $\gamma$ , apparaissent dans cette formule.*

**9.33** On a  $n^k/n^k = 1 - k(k-1)n^{-1} + \frac{1}{2}k^2(k-1)^2n^{-2} + O(k^6n^{-3})$ ; en divisant par  $k!$  et en sommant sur  $k \geq 0$ , on obtient  $e - en^{-1} + \frac{7}{2}en^{-2} + O(n^{-3})$ .

**9.34**  $A = e^\gamma$ ;  $B = 0$ ;  $C = -\frac{1}{2}e^\gamma$ ;  $D = \frac{1}{2}e^\gamma(1-\gamma)$ ;  $E = \frac{1}{8}e^\gamma$ ;  $F = \frac{1}{12}e^\gamma(3\gamma+1)$ .

**9.35** Comme  $1/k(\ln k + O(1)) = 1/k \ln k + O(1/k(\log k)^2)$ , la somme donnée est égale à  $\sum_{k=2}^n 1/k \ln k + O(1)$ . D'après la formule de sommation d'Euler, la somme qui reste vaut  $\ln \ln n + O(1)$ .

**9.36** Soit  $S_n$  notre somme. La formule de sommation d'Euler marche à la perfection dans ce cas :

$$\begin{aligned} S_n &= \sum_{0 \leq k < n} \frac{1}{n^2 + k^2} + \frac{1}{n^2 + x^2} \Big|_0^n \\ &= \int_0^n \frac{dx}{n^2 + x^2} + \frac{1}{2} \frac{1}{n^2 + x^2} \Big|_0^n + \frac{B_2}{2!} \frac{-2x}{(n^2 + x^2)^2} \Big|_0^n + O(n^{-5}). \end{aligned}$$

Par conséquent,  $S_n = \frac{1}{4}\pi n^{-1} - \frac{1}{4}n^{-2} - \frac{1}{24}n^{-3} + O(n^{-5})$ .

**9.37** Cela donne

$$\begin{aligned} \sum_{k,q \geq 1} (n-qk)[n/(q+1) < k \leq n/q] \\ &= n^2 - \sum_{q \geq 1} q \left( \binom{\lfloor n/q \rfloor + 1}{2} - \binom{\lfloor n/(q+1) \rfloor + 1}{2} \right) \\ &= n^2 - \sum_{q \geq 1} \binom{\lfloor n/q \rfloor + 1}{2}. \end{aligned}$$

La somme qui reste est identique à (9.55), sauf que le facteur  $\mu(q)$  manque ici. En appliquant la même méthode que pour (9.55), on trouve  $\zeta(2)$  à la place de  $1/\zeta(2)$ , et la réponse est donc  $(1 - \frac{\pi^2}{12})n^2 + O(n \log n)$ .

**9.38** Remplacez  $k$  par  $n-k$  et posez  $a_k(n) = (n-k)^{n-k} \binom{n}{k}$ . Alors  $\ln a_k(n) = n \ln n - \ln k! - k + O(kn^{-1})$ , et on peut appliquer la méthode de changement de queue sur cette somme, avec  $b_k(n) = n^n e^{-k}/k!$ ,  $c_k(n) = kb_k(n)/n$  et  $D_n = \{k \mid k \leq \ln n\}$ , pour obtenir  $\sum_{k=0}^n a_k(n) = n^n e^{1/e} (1 + O(n^{-1}))$ .

**9.39** On utilise la méthode de changement de queue, avec  $b_k(n) = (\ln n - k/n - \frac{1}{2}k^2/n^2)(\ln n)^k/k!$ ,  $c_k(n) = n^{-3}(\ln n)^{k+3}/k!$  et  $D_n = \{k \mid 0 \leq k \leq 10 \ln n\}$ . Lorsque  $k \approx 10 \ln n$ , on a  $k! \asymp \sqrt{k}(10/e)^k(\ln n)^k$ , donc le  $k$ ième terme est en  $O(n^{-10 \ln(10/e)} \log n)$ . On trouve donc  $n \ln n - \ln n - \frac{1}{2}(\ln n)(1 + \ln n)/n + O(n^{-2}(\log n)^3)$ .

**9.40** En combinant les termes deux à deux, on trouve que  $H_{2k}^m - (H_{2k} - \frac{1}{2k})^m = \frac{m}{2k} H_{2k}^{m-1}$  plus des termes dont la somme sur tout  $k \geq 1$  est en  $O(1)$ . Supposons que  $n$  soit pair. La formule de sommation d'Euler implique que

$$\begin{aligned} \sum_{k=1}^{n/2} \frac{H_{2k}^{m-1}}{k} &= \sum_{k=1}^{n/2} \frac{(\ln 2e^\gamma k)^{m-1} + O(k^{-1}(\log k)^{m-2})}{k} \\ &= \frac{(\ln e^\gamma n)^m}{m} + O(1); \end{aligned}$$

donc la somme est égale à  $\frac{1}{2} H_n^m + O(1)$ . Dans le cas général, la réponse est  $\frac{1}{2}(-1)^n H_n^m + O(1)$ .

**9.41** Soit  $\alpha = \bar{\phi}/\phi = -\phi^{-2}$ . On a

$$\begin{aligned} \sum_{k=1}^n \ln F_k &= \sum_{k=1}^n (\ln \phi^k - \ln \sqrt{5} + \ln(1-\alpha^k)) \\ &= \frac{n(n+1)}{2} \ln \phi - \frac{n}{2} \ln 5 + \sum_{k \geq 1} \ln(1-\alpha^k) - \sum_{k > n} \ln(1-\alpha^k). \end{aligned}$$

Cette dernière somme vaut  $\sum_{k>n} O(\alpha^k) = O(\alpha^n)$ . La réponse est donc

$$\phi^{n(n+1)/2} 5^{-n/2} C + O(\phi^{n(n-3)/2} 5^{-n/2}),$$

où  $C = (1-\alpha)(1-\alpha^2)(1-\alpha^3)\dots \approx 1,226742$ .

**9.42** La suggestion se déduit du fait que

$$\binom{n}{k-1} / \binom{n}{k} = \frac{k}{n-k+1} \leq \frac{\alpha n}{n-\alpha n+1} < \frac{\alpha}{1-\alpha}.$$

Soit  $m = \lfloor \alpha n \rfloor = \alpha n - \epsilon$ . Alors

$$\begin{aligned} \binom{n}{m} &< \sum_{k \leq m} \binom{n}{k} \\ &< \binom{n}{m} \left( 1 + \frac{\alpha}{1-\alpha} + \left( \frac{\alpha}{1-\alpha} \right)^2 + \dots \right) = \binom{n}{m} \frac{1-\alpha}{1-2\alpha}. \end{aligned}$$

Donc  $\sum_{k \leq \alpha n} \binom{n}{k} = \binom{n}{m} O(1)$ , et il reste à approximer  $\binom{n}{m}$ . D'après la formule de Stirling,  $\ln \binom{n}{m} = -\frac{1}{2} \ln n - (\alpha n - \epsilon) \ln(\alpha - \epsilon/n) - ((1-\alpha)n + \epsilon) \ln(1 - \alpha + \epsilon/n) + O(1) = -\frac{1}{2} \ln n - \alpha n \ln \alpha - (1-\alpha)n \ln(1-\alpha) + O(1)$ .

**9.43** Le dénominateur contient des facteurs de la forme  $z - \omega$ , où  $\omega$  est une racine complexe de l'unité. Seul le facteur  $z - 1$  apparaît avec multiplicité 5. Par conséquent, d'après (7.31), une seule des racines a un coefficient en  $\Omega(n^4)$ , et ce coefficient est  $c = 5/(5! \cdot 1 \cdot 5 \cdot 10 \cdot 25 \cdot 50) = 1/1500000$ .

**9.44** D'après la formule de Stirling,  $\ln(x^{-\alpha}x!/(x-\alpha)!)$  admet un développement asymptotique

$$\begin{aligned} -\alpha - (x + \frac{1}{2} - \alpha) \ln(1 - \alpha/x) & - \frac{B_2}{2 \cdot 1} (x^{-1} - (x-\alpha)^{-1}) \\ & - \frac{B_4}{4 \cdot 3} (x^{-3} - (x-\alpha)^{-3}) - \dots \end{aligned}$$

dans lequel chaque coefficient de  $x^{-k}$  est un polynôme en  $\alpha$ . Par conséquent,  $x^{-\alpha}x!/(x-\alpha)! = c_0(\alpha) + c_1(\alpha)x^{-1} + \dots + c_n(\alpha)x^{-n} + O(x^{-n-1})$  lorsque  $x \rightarrow \infty$ , où  $c_n(\alpha)$  est un polynôme en  $\alpha$ . Nous savons que  $c_n(\alpha) = [\alpha]_{\alpha-n} (-1)^n$  pour tout  $\alpha$  entier, et que  $[\alpha]_{\alpha-n}$  est un polynôme en  $\alpha$  de degré  $2n$ ; donc  $c_n(\alpha) = [\alpha]_{\alpha-n} (-1)^n$  pour tout  $\alpha$  réel. En d'autres termes, les formules asymptotiques

(Voir [220] pour d'autres détails).

$$\begin{aligned} x^{\underline{\alpha}} &= \sum_{k=0}^n \left[ \begin{matrix} \alpha \\ \alpha-k \end{matrix} \right] (-1)^k x^{\alpha-k} + O(x^{\alpha-n-1}), \\ x^{\overline{\alpha}} &= \sum_{k=0}^n \left[ \begin{matrix} \alpha \\ \alpha-k \end{matrix} \right] x^{\alpha-k} + O(x^{\alpha-n-1}) \end{aligned}$$

sont des généralisations des équations (6.28) et (6.11) qui s'appliquent au cas entier.

**9.45** Soient  $\langle a_1, a_2, \dots \rangle$  les quotients partiels de  $\alpha$ , et soit  $\alpha_m$  la fraction continue  $1/(a_m + \alpha_{m+1})$  pour  $m \geq 1$ . Alors  $D(\alpha, n) = D(\alpha_1, n) < D(\alpha_2, [\alpha_1 n]) + a_1 + 3 < D(\alpha_3, [\alpha_2 [\alpha_1 n]]) + a_1 + a_2 + 6 < \dots < D(\alpha_{m+1}, [\alpha_m \dots [\alpha_1 n] \dots]) + a_1 + \dots + a_m + 3m < \alpha_1 \dots \alpha_m n + a_1 + \dots + a_m + 3m$ , pour tout  $m$ . Divisez par  $n$  et faites tendre  $n$  vers l'infini ; les limites sont majorées par  $\alpha_1 \dots \alpha_m$  pour tout  $m$ . Finalement, on trouve

$$\alpha_1 \dots \alpha_m = \frac{1}{K(a_1, \dots, a_{m-1}, a_m + \alpha_m)} < \frac{1}{F_{m+1}}.$$

**9.46** Pour simplifier, nous écrirons simplement  $m$  à la place de  $m(n)$ . D'après la formule de Stirling, l'expression  $k^n/k!$  est maximale lorsque  $k \approx m \approx n/\ln n$  ; remplaçons donc  $k$  par  $m+k$  pour trouver que

$$\begin{aligned} \ln \frac{(m+k)^n}{(m+k)!} &= n \ln m - m \ln m + m - \frac{\ln 2\pi m}{2} \\ &\quad - \frac{(m+n)k^2}{2m^2} + O(k^3 m^{-2} \log n). \end{aligned}$$

En fait, c'est par  $\lfloor m \rfloor + k$  que nous voulons remplacer  $k$  ; cela ajoute un terme  $O(km^{-1} \log n)$ . La méthode de changement de queue, avec  $|k| \leq m^{1/2+\epsilon}$ ,

nous permet de sommer sur  $k$  et d'obtenir une approximation assez précise en fonction de la somme  $\Theta$  de (9.93) : *C'est une très Bell somme.*

$$\begin{aligned}\varpi_n &= \frac{e^{m-1} m^{n-m}}{\sqrt{2\pi m}} (\Theta_{2m^2/(m+n)} + O(1)) \\ &= e^{m-n-1/2} m^n \sqrt{\frac{m}{m+n}} \left(1 + O\left(\frac{\log n}{n^{1/2}}\right)\right).\end{aligned}$$

Il n'est pas difficile d'en déduire la formule voulue, avec une erreur relative en  $O(\log \log n / \log n)$ .

**9.47** Soit  $\log_m n = l + \theta$ , avec  $0 \leq \theta < 1$ . La somme de parties entières inférieures est égale à  $l(n+1) + 1 - (m^{l+1} - 1)/(m-1)$  ; la somme de parties entières supérieures est égale à  $(l+1)n - (m^{l+1} - 1)/(m-1)$  ; la somme exacte est égale à  $(l+\theta)n - n/\ln m + O(\log n)$ . Si on ignore les termes en  $o(n)$ , la différence entre la somme de parties entières supérieures et la somme exacte est  $(1-f(\theta))n$ , est la différence entre la somme exacte et la somme de parties entières inférieures est  $f(\theta)n$ , avec

$$f(\theta) = \frac{m^{1-\theta}}{m-1} + \theta - \frac{1}{\ln m}.$$

Le maximum de cette fonction est égal à  $f(0) = f(1) = m/(m-1) - 1/\ln m$ , tandis que son minimum vaut  $\ln \ln m / \ln m + 1 - (\ln(m-1)) / \ln m$ . C'est la somme de parties entières supérieures qui est la plus proche lorsque  $n$  est proche d'une puissance de  $m$ , mais c'est la somme de parties entières inférieures qui gagne lorsque  $\theta$  est compris entre 0 et 1.

**9.48** Soit  $d_k = a_k + b_k$ , où  $a_k$  compte les chiffres qui sont à gauche de la virgule. Alors  $a_k = 1 + \lfloor \log H_k \rfloor = \log \log k + O(1)$ , où "log" désigne le logarithme en base 10. Pour approximer  $b_k$ , examinons le nombre de chiffres nécessaires pour distinguer  $y$  de ses voisins immédiats  $y - \epsilon$  et  $y + \epsilon'$  : soit  $\delta = 10^{-b}$  la longueur de l'intervalle qui contient tous les nombres dont l'arrondi est égal à  $\hat{y}$ . On a  $|y - \hat{y}| \leq \frac{1}{2}\delta$ , ainsi que  $y - \epsilon < \hat{y} - \frac{1}{2}\delta$  et  $y + \epsilon' > \hat{y} + \frac{1}{2}\delta$ . Donc  $\epsilon + \epsilon' > \delta$ . De plus, si  $\delta < \min(\epsilon, \epsilon')$ , l'arrondi permet bien de distinguer  $\hat{y}$  de  $y - \epsilon$  et de  $y + \epsilon'$ . Par conséquent,  $10^{-b_k} < 1/(k-1) + 1/k$  et  $10^{1-b_k} \geq 1/k$  ; on a  $b_k = \log k + O(1)$ . Donc, finalement,  $\sum_{k=1}^n d_k = \sum_{k=1}^n (\log k + \log \log k + O(1))$ , ce qui est égal à  $n \log n + n \log \log n + O(n)$  d'après la formule de sommation d'Euler.

**9.49** On a  $H_n > \ln n + \gamma + \frac{1}{2}n^{-1} - \frac{1}{12}n^{-2} = f(n)$ , où  $f(x)$  est croissante pour tout  $x > 0$  ; donc, si  $n \geq e^{\alpha-\gamma}$ , alors  $H_n \geq f(e^{\alpha-\gamma}) > \alpha$ . On a aussi  $H_{n-1} < \ln n + \gamma - \frac{1}{2}n^{-1} = g(n)$ , où  $g(x)$  est croissante pour tout  $x > 0$  ; donc, si  $n \leq e^{\alpha-\gamma}$ , alors  $H_{n-1} \leq g(e^{\alpha-\gamma}) < \alpha$ . Par conséquent,  $H_{n-1} \leq \alpha \leq H_n$  implique que  $e^{\alpha-\gamma} + 1 > n > e^{\alpha+\gamma} - 1$ . (Des résultats plus précis ont été obtenus par Boas et Wrench [33]).

**9.50** (a) La moyenne de la somme gagnée est  $\sum_{1 \leq k \leq N} k/(k^2 H_N^{(2)}) = H_N/H_N^{(2)}$ , et nous voulons sa valeur asymptotique à  $O(N^{-1})$  près :

$$\frac{\ln N + \gamma + O(N^{-1})}{\pi^2/6 - N^{-1} + O(N^{-2})} = \frac{6 \ln 10}{\pi^2} n + \frac{6\gamma}{\pi^2} + \frac{36 \ln 10}{\pi^4} \frac{n}{10^n} + O(10^{-n}).$$

Le coefficient  $(6 \ln 10)/\pi^2 \approx 1,3998$  indique qu'on peut espérer à peu près 40% de bénéfices.

(b) La probabilité de faire un bénéfice est  $\sum_{n < k \leq N} 1/(k^2 H_N^{(2)}) = 1 - H_n^{(2)}/H_N^{(2)}$ . Comme  $H_n^{(2)} = \frac{\pi^2}{6} - n^{-1} + \frac{1}{2}n^{-2} + O(n^{-3})$ , cette probabilité vaut

$$\frac{n^{-1} - \frac{1}{2}n^{-2} + O(n^{-3})}{\pi^2/6 + O(N^{-1})} = \frac{6}{\pi^2} n^{-1} - \frac{3}{\pi^2} n^{-2} + O(n^{-3}).$$

Remarquez qu'elle décroît si  $n$  augmente. (La valeur moyenne de la question (a) est élevée car elle tient compte de sommes si importantes que toute l'économie mondiale serait affectée si on devait réellement les payer).

**9.51** Strictement parlant, c'est faux car la fonction représentée par  $O(x^{-2})$  peut très bien ne pas être intégrable (par exemple  $[x \in S]/x^2$ , où  $S$  est un ensemble non mesurable). Toutefois, si on suppose que  $f(x)$  est une fonction intégrable telle que  $f(x) = O(x^{-2})$  lorsque  $x \rightarrow \infty$ , alors  $|\int_n^\infty f(x) dx| \leq \int_n^\infty |f(x)| dx \leq \int_n^\infty Cx^{-2} dx = Cn^{-1}$ .

(Par opposition  
à une fonction  
exécrable).

**9.52** En fait, l'empilement de  $n$  peut être remplacé par n'importe quelle fonction  $f(n)$  qui tend vers l'infini. Soit  $\langle m_0, m_1, m_2, \dots \rangle$  une suite telle que  $m_0 = 0$  et que  $m_k$  soit le plus petit entier  $> m_{k-1}$  tel que

$$\left(\frac{k+1}{k}\right)^{m_k} \geq f(k+1)^2.$$

Soit maintenant  $A(z) = \sum_{k \geq 1} (z/k)^{m_k}$ . Cette série converge pour tout  $z$  car les termes pour lesquels  $k > |z|$  sont majorés par une série géométrique. On a aussi  $A(n+1) \geq ((n+1)/n)^{m_n} \geq f(n+1)^2$ , donc  $\lim_{n \rightarrow \infty} f(n)/A(n) = 0$ .

**9.53** On peut montrer par induction que le terme en  $O$  est égal à  $(m-1)!^{-1} \int_0^x t^{m-1} f^{(m)}(x-t) dt$ . Comme  $f^{(m+1)}$  et  $f^{(m)}$  sont de signes opposés, la valeur absolue de cette intégrale est majorée par  $|f^{(m)}(0)| \int_0^x t^{m-1} dt$ ; l'erreur est donc majorée par la valeur absolue du premier terme supprimé.

**9.54** Soit  $g(x) = f(x)/x^\alpha$ . Alors  $g'(x) \sim -\alpha g(x)/x$  lorsque  $x \rightarrow \infty$ . D'après le théorème de la moyenne, il existe un  $y$  compris entre  $x - \frac{1}{2}$  et  $x + \frac{1}{2}$  tel que  $g(x - \frac{1}{2}) - g(x + \frac{1}{2}) = -g'(y) \sim \alpha g(y)/y$ . Or,  $g(y) = g(x)(1 + O(1/x))$ , donc  $g(x - \frac{1}{2}) - g(x + \frac{1}{2}) \sim \alpha g(x)/x = \alpha f(x)/x^{1+\alpha}$ . Par

Ce théorème ne me plaît pas.

conséquent,

$$\sum_{k \geq n} \frac{f(k)}{k^{1+\alpha}} = O\left(\sum_{k \geq n} (g(k - \frac{1}{2}) - g(k + \frac{1}{2}))\right) = O(g(n - \frac{1}{2})).$$

**9.55** On peut préciser l'approximation de  $(n+k+\frac{1}{2})\ln(1+k/n) + (n-k+\frac{1}{2})\ln(1-k/n)$  en  $k^2/n + k^4/6n^3 + O(n^{-3/2+5\epsilon})$ . En apparence, il nous faut donc un facteur supplémentaire  $e^{-k^4/6n^3}$  dans  $b_k(n)$ , et  $c_k(n) = 2^{2n}n^{-2+5\epsilon}e^{-k^2/n}$ . Il vaut mieux toutefois ne pas toucher à  $b_k(n)$  et poser

$$c_k(n) = 2^{2n}n^{-2+5\epsilon}e^{-k^2/n} + 2^{2n}n^{-5+5\epsilon}k^4e^{-k^2/n},$$

remplaçant ainsi  $e^{-k^4/6n^3}$  par  $1+O(k^4/n^3)$ . La somme  $\sum_k k^4 e^{-k^2/n}$  est en  $O(n^{5/2})$ , comme on le montre dans l'exercice 30.

**9.56** Si  $k \leq n^{1/2+\epsilon}$ , alors  $\ln(n^k/n^k) = -\frac{1}{2}k^2/n + \frac{1}{2}k/n - \frac{1}{6}k^3/n^2 + O(n^{-1+4\epsilon})$  d'après la formule de Stirling. Par conséquent,

$$n^k/n^k = e^{-k^2/2n}(1 + k/2n - \frac{2}{3}k^3/(2n)^2 + O(n^{-1+4\epsilon})).$$

En sommant avec l'identité de l'exercice 30, et en ne tenant pas compte du terme pour  $k=0$ , on trouve  $-1 + \Theta_{2n} + \Theta_{2n}^{(1)} - \frac{2}{3}\Theta_{2n}^{(3)} + O(n^{-1/2+4\epsilon}) = \sqrt{\pi n/2} - \frac{1}{3} + O(n^{-1/2+4\epsilon})$ .

**9.57** La somme devient  $\int_0^\infty ue^{-u}\zeta(1+u/\ln n)$  si on suit le conseil donné. La fonction zêta peut être définie par la série

$$\zeta(1+z) = z^{-1} + \sum_{m \geq 0} (-1)^m \gamma_m z^m/m!,$$

où  $\gamma_0 = \gamma$ , et où  $\gamma_m$  est la constante de Stieltjes [341, 201]

$$\lim_{n \rightarrow \infty} \left( \sum_{k=1}^n \frac{(\ln k)^m}{k} - \frac{(\ln n)^{m+1}}{m+1} \right).$$

Par conséquent, notre somme est égale à

$$\ln n + \gamma - 2\gamma_1(\ln n)^{-1} + 3\gamma_2(\ln n)^{-2} - \dots.$$

**9.58** Soient  $0 \leq \theta \leq 1$  et  $f(z) = e^{2\pi iz\theta}/(e^{2\pi iz} - 1)$ . On a

$$|f(z)| = \frac{e^{-2\pi y\theta}}{1 + e^{-2\pi y}} \leq 1, \quad \text{lorsque } x \bmod 1 = \frac{1}{2};$$

$$|f(z)| \leq \frac{e^{-2\pi y\theta}}{|e^{-2\pi y} - 1|} \leq \frac{1}{1 - e^{-2\pi\epsilon}}, \quad \text{lorsque } |y| \geq \epsilon.$$

Donc  $|f(z)|$  est majorée sur le contour et l'intégrale est en  $O(M^{1-m})$ . Le résidu de  $2\pi i f(z)/z^m$  en  $z = k \neq 0$  est  $e^{2\pi i k \theta}/k^m$ ; le résidu en  $z = 0$  est le coefficient de  $z^{-1}$  dans

$$\frac{e^{2\pi iz\theta}}{z^{m+1}} \left( B_0 + B_1 \frac{2\pi iz}{1!} + \dots \right) = \frac{1}{z^{m+1}} \left( B_0(\theta) + B_1(\theta) \frac{2\pi iz}{1!} + \dots \right),$$

soit  $(2\pi i)^m B_m(\theta)/m!$ . Par conséquent, la somme des résidus à l'intérieur du contour vaut

$$\frac{(2\pi i)^m}{m!} B_m(\theta) + 2 \sum_{k=1}^M e^{\pi i m/2} \frac{\cos(2\pi k\theta - \pi m/2)}{k^m}.$$

Elle est, comme l'intégrale du contour, en  $O(M^{1-m})$ , donc elle tend vers zéro lorsque  $M \rightarrow \infty$ .

**9.59** Pour toute fonction  $F(x)$  qui se comporte "suffisamment bien", on a l'identité

$$\sum_k F(k+t) = \sum_n G(2\pi n) e^{2\pi i nt},$$

où  $G(y) = \int_{-\infty}^{+\infty} e^{-iyx} F(x) dx$ . (Il s'agit de la "formule de sommation de Poisson" qu'on trouve dans la littérature standard, par exemple dans Henrici [182, Théorème 10.6e]).

**9.60** D'après l'exercice 5.22, la formule considérée est équivalente à

$$n^{1/2} = n^{1/2} \left( 1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} - \frac{21}{32768n^4} + O(n^{-5}) \right)$$

Les exercices 6.64 et 9.44 permettent d'en déduire le résultat.

**9.61** L'idée consiste à prendre un  $\alpha$  "presque" rationnel. Soit  $a_k = 2^{2^{2^k}}$  le  $k$ ième quotient partiel de  $\alpha$ , et soit  $n = \frac{1}{2} a_{m+1} q_m$ , où  $q_m = K(a_1, \dots, a_m)$  et  $m$  est pair. Alors  $0 < \{q_m \alpha\} < 1/K(a_1, \dots, a_{m+1}) < 1/(2n)$ , et si on prend  $v = a_{m+1}/(4n)$  on obtient une discrépance  $\geq \frac{1}{4} a_{m+1}$ . Si c'était plus petit que  $n^{1-\epsilon}$ , on aurait  $a_{m+1}^\epsilon = O(q_m^{1-\epsilon})$ ; or,  $a_{m+1} > q_m^{2^m}$ .

**9.62** Voir Canfield [48]; voir aussi David et Barton [71, chapitre 16] pour des équivalents asymptotiques des nombres de Stirling des deux espèces.

**9.63** Soit  $c = \phi^{2-\Phi}$ . Une première approximation,  $cn^{\Phi-1} + o(n^{\Phi-1})$ , a été donnée par Fine [150]. Ilan Vardi fait remarquer que l'estimation plus précise donnée dans l'énoncé peut se déduire du fait que le terme d'erreur  $e(n) = f(n) - cn^{\Phi-1}$  satisfait la récurrence approximée  $c^\Phi n^{2-\Phi} e(n) \approx -\sum_k e(k) [1 \leq k < cn^{\Phi-1}]$ . La fonction

$$\frac{n^{\Phi-1} u(\ln \ln n / \ln \phi)}{\ln n}$$

satisfait asymptotiquement cette récurrence si  $u(x+1) = -u(x)$ . Vardi conjecture qu'il existe une fonction  $u$  telle que

$$f(n) = n^{\phi-1} \left( c + u\left(\frac{\ln \ln n}{\ln \phi}\right) (\ln n)^{-1} + O((\log n)^{-2}) \right).$$

L'expérience montre que  $f(n)$  est égal à l'entier le plus proche de  $cn^{\phi-1}$  pour tout  $1 \leq n \leq 400$  sauf pour une valeur :  $f(273) = 39 > c \cdot 273^{\phi-1} \approx 38,4997$ . Cependant, lorsque  $n$  croît, les erreurs deviennent très grandes, notamment en raison des propriétés démontrées dans l'exercice 2.36. Par exemple,  $e(201636503) \approx 35,73$  ;  $e(919986484788) \approx -1959,07$ .

**9.64** (A partir de cette identité, qui concerne  $B_2(x)$ , on peut facilement démontrer celle de l'exercice 58 par induction sur  $m$ ). Si  $0 < x < 1$ , l'intégrale  $\int_x^{1/2} \sin N\pi t / \sin \pi t$  peut s'exprimer comme une somme de  $N$  intégrales qui sont toutes en  $O(N^{-2})$ , donc elle est en  $O(N^{-1})$  ; la constante liée à ce  $O$  peut dépendre de  $x$ . En intégrant l'identité  $\sum_{n=1}^N \cos 2n\pi t = \Re(e^{2\pi i t}(e^{2N\pi i t} - 1)/(e^{2\pi i t} - 1)) = -\frac{1}{2} + \frac{1}{2} \sin(2N+1)\pi t / \sin \pi t$  et en posant  $N \rightarrow \infty$ , on trouve  $\sum_{n \geq 1} (\sin 2n\pi x)/n = \frac{\pi}{2} - \pi x$ , une relation qu'Euler connaissait ([107] et [110, partie 2, §92]). En intégrant encore une fois, on obtient la formule désirée. (Cette solution a été suggérée par E. M. E. Werthum [367] ; le calcul original d'Euler ne satisfaisait pas les standards de rigueur actuels).

**9.65** Comme  $a_0 + a_1 n^{-1} + a_2 n^{-2} + \dots = 1 + (n-1)^{-1}(a_0 + a_1(n-1)^{-1} + a_2(n-1)^{-2} + \dots)$ , on obtient la récurrence  $a_{m+1} = \sum_k \binom{m}{k} a_k$ , qui coïncide avec celle des nombres de Bell. Par conséquent,  $a_m = \omega_m$ .

Il existe une preuve un peu plus longue, mais aussi plus instructive, basée sur le fait que  $1/(n-1) \dots (n-m) = \sum_k \binom{k}{m} / n^k$ , d'après (7.48).

**9.66** Si  $f$  est une application aléatoire de  $\{1, 2, \dots, n\}$  dans lui-même, le nombre moyen d'éléments distincts dans la suite  $1, f(1), f(f(1)), \dots$  est la fonction  $Q(n)$  de l'exercice 56, qui est égale à  $\frac{1}{2}\sqrt{2\pi n} + O(1)$  ; ceci pourrait peut-être expliquer le facteur  $\sqrt{2\pi n}$ .

**9.67** On sait que  $\ln \chi_n \sim \frac{3}{2} n^2 \ln \frac{4}{3}$  ; des expérimentations montrent que la constante  $e^{-\pi/6}$  est juste au moins sur les huit premiers chiffres significatifs.

**9.68** Cela serait faux si, par exemple, il existait un entier  $m$  et un  $0 < \epsilon < \frac{1}{8}$  tels que  $e^{n-\gamma} = m + \frac{1}{2} + \epsilon/m$ , mais on ne connaît pas de contre-exemple.

*"The paradox is now fully established that the utmost abstractions are the true weapons with which to control our thought of concrete fact."*

— A. N. Whitehead [372]