Probabilités, Statistiques, Combinatoire

Adrian Tanasă

Université de Bordeaux – Licence Informatique

CM 3

Rappel - définition de la classe combinatoire

Définition :

Une **classe combinatoire** est la donnée d'un ensemble C, et d'une fonction

$$t : C \to N,$$

t. q. pour tout n, l'ensemble C_n des éléments de C qui ont n pour image par t,

$$C_n = \{x \in C \mid t(x) = n\},$$

est fini.

La **suite de comptage** de la classe C est la suite $(cn)_{n \geq 0}$ définie par : pour tout n, $c_n = \# C_n$.

La fonction t est appelée fonction de **taille** pour la classe.

si on change t, on change généralement de suite de comptage (et même, on peut ne plus avoir une classe combinatoire).

Exemples

Soit un alphabet à 2 lettres, $A = \{a, b\}$, et comme ensemble, $C = A^*$ (tous les mots sur l'alphabet A).

Plusieurs choix de fonctions taille :

- fonction taille, la longueur du mot, $t(w) = |w|$: on a une classe combinatoire, et la suite de comptage est $a_n = 2^n$ (car $C_n = A^n$)
- avec comme fonction taille, le nombre d'occurrences de la lettre a ($t'(w) = |w|_a$).
  on n'a pas une classe combinatoire : on peut former une infinité de mots de "taille" 0 (mots sans la lettre a : b, bb, $bbb \ldots$).
- comme fonction taille :

  $$t''(w) = |w|_a + 2|w|_b$$

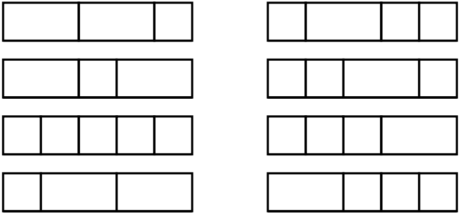
  on a bien une classe combinatoire
  On peut calculer "à la main" les premiers termes : $C''_0 = \{\varepsilon\}$, $C''_1 = \{a\}$, $C''_2 = \{aa, b\}$, $C''_3 = \{ab, ba, aaa\}$, $C''_4 = \{aaaa, aab, aba, baa, bb\}$.

Pavages : segments et dominos

- On se donne un entier n.

  Question : de combien de façons on peut recouvrir un "segment" de longueur n (n cases) avec des segments de longueur 1 (case isolée) ou 2 ("domino") ?

Exemple : $n = 5$ (les 8 segments de taille 5)



Pavages, codage et classe combinatoire

- **Codage :** une case isolée est codée par a ; un domino est codé par b.
- un mot code ainsi un segment ; la taille du segment est donc le nombre de a, plus deux fois le nombre de b.
- On retrouve la classe combinatoire $\{a, b\}^*$ avec la taille $t(w) = |w|_a + 2|w|_b$.

Permutations

Objet très souvent utilisé en combinatoire.

On s'intéresse aux permutations sur l'ensemble $[\![1, n]\!]$.

1. math. : Une *permutation* est une bijection de l'ensemble $[\![1, n]\!]$ vers lui-même.
2. info : Une *permutation* est un mot sur les lettres de l'alphabet $[\![1, n]\!]$ où chaque lettre apparaît une seule fois.
3. pour les enfants : Une *permutation* est un mélange de cartes numérotés de 1 à n.

Les permutations comme classe combinatoire

Soit S_n l'ensemble des permutations sur l'ensemble $[[1, n]]$

Soit $S = \cup_{n \in \mathbb{N}} S_n$. L'ensemble S est une classe combinatoire où on prend comme "taille" le nombre d'éléments de l'ensemble S_n.

La suite de comptage des permutations est :

Théorème

Pour tout entier $n \geq 1$, le nombre de permutations d'un ensemble de cardinal n, est $n! = n.(n-1)\dots 2.1$ ("factorielle n").

Comptage des permutations

Preuve par comptage : on compte les façons de construire le mot, lettre par lettre.

- il y a n choix pour la première lettre ;
- une fois fixée la première lettre, il y a $n-1$ choix restants pour la deuxième lettre, quelle que soit la première ;
- une fois fixées les deux premières lettres, quelles qu'elles soient, il y a $n-2$ choix pour la troisième ;
- le schéma continue : une fois fixées les k premières lettres, quelles qu'elles soient, il y a $n-k$ choix pour la $(k+1)$-ème lettre ;
- on est dans un schéma multiplicatif, le nombre de mots est donc $n(n-1)\dots 2 \cdot 1 = n!$ (CQFD).

Corollaire : en arrêtant le calcul aux mots de seulement k lettres, toutes différentes (pour $k \leq n$), on obtient le nombre de séquences de longueur k d'entiers **tous différents** entre 1 et n : $n(n-1)\dots(n-k+1)$, soit $\frac{n!}{(n-k)!}$.

Formule pour les coefficients binomiaux

- On vient de voir que le nombre de séquences de k entiers entre 1 et n, **tous différents**, est $n!/(n-k)!$.
- Mais une autre façon de décrire cette séquence, c'est la suivante :
  - on choisit l'ensemble des k entiers qui vont apparaître dans la séquence (il y a $\binom{n}{k}$ choix possibles) ;
  - une fois l'ensemble choisi, on choisit dans quel ordre les k entiers vont apparaître dans la séquence
    (il y a $k!$ choix, quelque soit l'ensemble choisi).
- On a

$$\frac{n!}{(n-k)!} = \binom{n}{k} \cdot k!$$

ce qui implique une formule explicite pour $\binom{n}{k}$:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Toujours sur les coefficients binomiaux

- La formule $\frac{n!}{k!(n-k)!}$ rend claire la symétrie entre k et $n-k$ (changer k pour $n-k$ laisse le dénominateur inchangé)

- On peut simplifier la formule en supprimant, au numérateur et au dénominateur, les facteurs de 1 à $n-k$:

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!}$$

- En particulier, pour de petites valeurs de k :

$$\binom{n}{0} = 1; \; \binom{n}{1} = n; \; \binom{n}{2} = \frac{n(n-1)}{2}; \; \binom{n}{3} = \frac{n(n-1)(n-2)}{6}$$

Le triangle de Pascal - relation de récurrence

- On peut également prouver une **relation de récurrence** ("triangle de Pascal") :

  pour $0 < k < n$,

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

Le triangle de Pascal

$n \backslash k$	0	1	2	3	4	5	6	7	8
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

Généralisation : les coefficients multinomiaux

- Le coefficient $\binom{n}{k}$ donne le nombre de façons de "partager" (partitionner) $[[1, n]]$ en une partie A de cardinal k et une partie B de cardinal $n - k$.
- Une extension naturelle, c'est de compter les *partitions* de $[[1, n]]$ en i ensembles $A_1, \ldots, A_i$ (au lieu de juste $i = 2$)
  - chaque $a \in [[1, n]]$ doit apparaître dans **un et un seul** des A_j.
- Si on impose les cardinaux des A_j ($\#A_j = a_j$, avec $a_1 + \cdots + a + i = n$), on obtient les *coefficients multinomiaux*,

$$\binom{n}{a_1, a_2, \ldots, a_i} = \frac{n!}{a_1!a_2!\ldots a_i!}$$

- Par exemple, cela permet de compter les **anagrammes** d'un mot avec des lettres répétées comme ABRACADABRA : il y a 5 occurrences de A, 2 de B, 1 de C, 1 de D, 2 de R :

$$\binom{11}{5, 2, 1, 1, 2} = 83160$$

Comptage de mots

Souvent, on fixe un alphabet A (de petite taille, disons d), et on veut compter, selon leur longueur, les mots de A^* qui satisfont une certaine **contrainte**.

- Notion naturelle de taille : la **longueur** des mots.
- **Exemples** de contraintes possibles :
  - certaines lettres ne peuvent pas être consécutives ;
  - interdire plus de k répétitions consécutives d'une même lettre ;
  - fixer le nombre total d'occurrences d'une certaine lettre ;
  - contrainte plus difficile : mots w tels que, pour chaque k, il y ait parmi les k premières lettres de w au moins autant de a que de b.
- On a forcément une classe combinatoire : le nombre de mots acceptables de longueur n est forcément inférieur à d^n (nombre de mots de longueur n, sans contrainte).
- On appelle **langage**, n'importe quel ensemble $L \subset A^*$ de mots.
- Fréquemment, on définit le langage choisi à partir d'une règle de codage pour des objets qu'on veut compter (Ex. 2.1, feuille TD2 : codages de chemins)

Un exemple de contrainte locale

- On part d'un alphabet à 3 lettres : $A = \{a, b, c\}$
- On interdit certaines lettres consécutives, par exemple : **pas de b immédiatement après un a, pas de c immédiatement après un c.**
- On veut déterminer le nombre ℓ_n de mots de longueur n dans un tel langage L
- Chaque a peut être suivi de 2 lettres possibles, a ou c ; chaque b, de 3 lettres possibles ; chaque c, de 2 lettres possibles.
- Donc chaque mot de L, quelque soit sa dernière lettre, peut être prolongé de 2 ou 3 façons par ajout d'une lettre.
- Ça nous assure une inégalité : $2\ell_n \leq \ell_{n+1} \leq 3\ell_n$
- Avec la condition initiale $\ell_1 = 3$, ça permet de décrire un encadrement très grossier : $3.2^{n-1} \leq \ell_n \leq 3^n$.

Contrainte locale (suite)

- Pour trouver un encadrement plus fin, ou un comptage exact, il faut raffiner le comptage, en distinguant selon la dernière lettre des mots : on pose a_n pour le nombre de mots de L se terminant par a ; b_n pour le nombre de mots de L se terminant par b ; c_n pour le nombre de mots se terminant par c.
- On obtient des relations en examinant quelle lettre peut précéder la dernière :

$$\begin{aligned}
a_{n+1} &= a_n + b_n + c_n \\
b_{n+1} &= b_n + c_n \\
c_{n+1} &= a_n + b_n
\end{aligned}$$

- Ça permet au moins de calculer rapidement a_n, b_n, c_n rapidement jusqu'à un n fixé ; et $\ell_n = a_n + b_n + c_n$.
- Pour être plus précis, ça demande plus de travail.
- Dans cet exemple, on peut obtenir

$$\ell_{n+1} = 2\ell_n + \ell_{n-1} - \ell_{n-2}$$