

Dénombrements et probabilités

François Capaces¹, Christophe Antonini², Olivier Teytaud³, Pierre Borgnat⁴, Annie Chateau⁵, and Edouard Lebeau⁶

¹, ,

²Enseignant en CPGE, Institut Stanislas, Cannes

³Chargé de recherche INRIA, Université d'Orsay, Orsay

⁴Chargé de recherche CNRS, ENS Lyon, Lyon

⁵Maitre de conférence, Université Montpellier-2, Montpellier

⁶Enseignant en CPGE, Lycée Henri Poincaré, Nancy

2 janvier 2023



Espaces mesurables, variables aléatoires et probabilités.

1 Dénombrements et probabilités

Une première section sera dédiée aux dénombrements. On présentera tout d'abord quelques fondements des probabilités : espaces mesurés, événements, variables aléatoires (sections 1.2, 1.3, 1.4) ; on s'attardera sur l'important cas des sommes de variables aléatoires en section ?? . On présentera les lois conditionnelles en section ?? .

Le cas de modèles ayant une composante d'évolution dans le temps est fondamental : martingales (??), processus stochastiques (??).

Après un peu de zoologie des lois de probabilités (??), on consacrera trois sections aux sommes de variables aléatoires : lois des grands nombres (??), théorème central-limite (??), grandes déviations (??). On conclura par une section applicative (??), avant une brève présentation du vocabulaire statistique en section ?? .

La partie ??, consacrée aux π -systèmes, est indispensable pour s'attaquer aux probabilités.

1.1 Combinatoire et dénombrements

On consultera avec grand profit le chapitre 1 de [?], très agréable et fournissant une bonne quantité de résultats originaux, complétant utilement les résultats très classiques suivants. On pourra aussi aller consulter le paragraphe ?? pour les cardinaux d'ensembles infinis.

On se limitera ici à plus intuitif : cardinaux d'ensembles finis (1.1.1), dénombrement de fonctions (1.1.2), dénombrement d'arrangements (1.1.3), et enfin les célèbres combinaisons (1.1.4 – on parle aussi de binôme de Newton).

1.1.1 Cardinaux d'ensembles finis

Avec A un ensemble fini, on a la **Formule d'inclusion-exclusion** ; avec $F_i \in \mathcal{A} = P(A)$, on a

$$|\cup_{i \leq n} F_i| = \sum_{i \leq n} |F_i| - \sum_{1 \leq i < j \leq n} |F_i \cap F_j| + \sum_{1 \leq i < j < k \leq n} |F_i \cap F_j \cap F_k| \dots + (-1)^{n-1} |\cap_{1 \leq i \leq n} F_i|.$$

Le nombre de parties à p éléments d'un ensemble à n éléments est C_n^p ; ces coefficients binomiaux sont définis en 1.1.4.

1.1.2 Dénombrement de fonctions

On considère E et F deux ensembles finis, de cardinaux e et f .

Ensemble des applications de E dans F L'ensemble des applications de E dans F , noté F^E , a pour cardinal f^e .

Ensemble des injections de E dans F L'ensemble des injections de E dans F a pour cardinal

$$A_f^e = \frac{f!}{(f-e)!} \quad \text{si } f \geq e ; 0 \text{ sinon.}$$

La preuve se fait facilement par récurrence. Voir 1.1.3 pour les premières valeurs.

Ensemble des surjections (et bijections) de E dans F En notant S_e^f le cardinal de l'ensemble des surjections de E dans F (dans le cas $e \geq f$) divisé par $f!$ (c'est-à-dire, dans le cas de E et F totalement ordonnés, le nombre de surjections croissantes), on a les formules :

$$S_e^1 = S_e^e = 1$$

$$\forall(e, f) \quad S_{e+1}^f = S_e^{f-1} + f.S_e^f$$

On obtient ainsi les valeurs suivantes de S_e^f :

	$f = 1$	$f = 2$	$f = 3$	$f = 4$	$f = 5$
$e = 1$	1	0	0	0	0
$e = 2$	1	1	0	0	0
$e = 3$	1	3	1	0	0
$e = 4$	1	7	6	1	0
1	15	25	10	1	0

Le nombre de bijections de E vers F vaut $e!$ si $e = f$, 0 sinon.

Ensemble des applications croissantes de E vers F E et F sont maintenant munis d'un ordre total.

L'ensemble des applications croissantes de E dans F a le même cardinal que l'ensemble des applications strictement croissantes de E dans $[1, f + e - 1]$; En effet, on a :

- si u de $\{1, 2, 3, \dots, e\}$ dans $\{1, 2, 3, \dots, f\}$ est croissante alors $x \mapsto v_u(x) = u(x) + x$ de $\{1, 2, 3, \dots, e\}$ dans $\{1, 2, 3, \dots, e + f\}$ est strictement croissante
- $u \mapsto v_u$ est bijective de l'ensemble des applications croissantes de E dans F vers l'ensemble des applications strictement croissantes de E dans $[1, f + e - 1]$

Il est alors facile de montrer que cet ensemble a pour cardinal C_{f+e-1}^e .

1.1.3 Arrangements

DÉFINITION 0.1 p -arrangement de E

On appelle p -**arrangement** de E une application injective de \mathbb{N}_p dans E . On note A_n^p le cardinal de l'ensemble des p -arrangements d'un ensemble à n éléments.

Au vu des résultats précédents, on peut dire que $A_n^p = \frac{n!}{(n-p)!}$.

Les premières valeurs sont :

	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$
$n = 1$	1	0	0	0	0
$n = 2$	2	2	0	0	0
$n = 3$	3	6	6	0	0
$n = 4$	4	12	24	24	0
$n = 5$	5	20	60	120	120

1.1.4 Combinaisons

DÉFINITION 0.2 p -combinaison de E

On appelle p -**combinaison** de E tout sous-ensemble de E de cardinal p . On note C_n^p ou $\binom{n}{p}$ le cardinal de l'ensemble des p -combinaisons de E .

Intuition La notation C_n^p est d'usage en français mais on peut trouver aussi $\binom{n}{p}$ qui est la notation anglophone pour le cardinal des p -arrangements. Ces quantités sont usuellement appelées coefficients binomiaux (ou coefficients du binôme) du fait de la formule de Newton, présentée plus bas.

Démonstration On montre facilement, par récurrence, que $C_n^p = \frac{n!}{p!(n-p)!}$.

Cette formule peut aussi se déduire sans récurrence en voyant qu'il y a un nombre $p!$ de p -arrangements qui donnent une p -combinaison donnée et donc

$$C_n^p = \frac{1}{p!} A_n^p = \frac{n!}{p!(n-p)!}.$$

Elle peut aussi se déduire du fait que le groupe $\sigma(E)$ des permutations de $E = \{1, 2, \dots, e\}$ agit transitivement sur l'ensemble des p -combinaisons de E , que le stabilisateur S de $F = \{1, 2, \dots, f\} \subset$

E est le produit $A \times B$ avec A et B respectivement les groupes de permutations de $\{1, 2, \dots, f\}$ et $\{f+1, f+2, \dots, e\}$; S a donc pour cardinal $S = f!(e-f)!$, d'où

$$C_e^f = \frac{\sigma(E)}{|S|} = \frac{e!}{f!(e-f)!}.$$

Un argument similaire permet d'ailleurs de montrer la formule $A_n^p = \frac{n!}{(n-p)!}$ sans récurrence.

En outre on a les formules suivantes :

$$C_n^p = C_n^{n-p}.$$

$$C_{n+1}^p = C_n^p + C_n^{p+1}.$$

Formule de Newton, valable dans un anneau : si $a.b = b.a$ et $n > 0$, alors :

$$(a+b)^n = \sum_{k=0}^n C_n^k a^k . b^{n-k}.$$

$$\sum_{k=0}^n C_n^k = 2^n, \quad \sum_{k=0}^n (-1)^k C_n^k = 0.$$

(ces deux formules sont obtenues en spécialisant la formule de Newton)

pour $1 \leq p \leq n$, on a $p.C_n^p = n.C_{n-1}^{p-1}$.

$$\sum_{k=0}^n k.C_n^k = n.2^{n-1}, \quad \sum_{k=0}^n (C_n^k)^2 = C_{2n}^n.$$

Les premières valeurs de C_n^p sont :

	$p=0$	$p=1$	$p=2$	$p=3$	$p=4$
$n=0$	1	0	0	0	0
$n=1$	1	1	0	0	0
$n=2$	1	2	1	0	0
$n=3$	1	3	3	1	0
$n=4$	1	4	6	4	1

1.1.5 Quelques applications

On parlera ici de la dimension d'espaces de polynômes, de combinatoire, du binôme de Newton et de familles sommables infinies.

Dimension des polynômes homogènes Pour $n \geq 1$ et $d \geq 0$, notons $C_{n,d}$ la dim des polynômes de n indéterminées, homogènes de degré d (i.e. tous les termes sont de degré d). On a clairement $\forall n \geq 1, C_{n,0} = 1$ (les polynômes homogènes de degré 0 sont les constantes) et $\forall d \geq 0, C_{1,d} = 1$ (avec une seule indéterminée, l'espace des polynômes homogènes de degré d est $Vect(X^d)$). Par ailleurs, on peut montrer la relation $\forall n \geq 2, \forall d \geq 1, C_{n,d} = C_{n-1,d} + C_{n,d-1}$. Cette formule s'obtient en partitionnant les monômes homogènes unitaires en deux catégories : ceux avec X_n de puissance nulle : ce sont des polynômes homogènes de degré d et à $n-1$ indéterminées, il y en a $C_{n-1,d}$; il reste les polynômes pour lesquels X_n a une puissance ≥ 1 . En les factorisant par X_n , on trouve des polynômes unitaires de degré $d-1$, il y en a donc $C_{n,d-1}$. Notre formule est ainsi prouvée.

Par ailleurs, la quantité C_{n+d-1}^d vérifie les mêmes "conditions au bord" (valeurs pour $n=1$ ou $d=0$) et la même relation de récurrence. On peut en déduire l'égalité de ces suites doubles par récurrence, et ainsi

$$\forall n \geq 1, \forall d \geq 0, \quad C_{n,d} = C_{n+d-1,d}.$$

PROPOSITION 0.1

Soit n un entier naturel. On a alors

$$\sum_{k=0}^n C_n^k (-1)^{n-k} k^p = \begin{cases} 0 & \text{si } p < n \\ \text{si } p = n \end{cases}.$$

Une formule utile de combinatoire *Application 0.1* Ce résultat est utile pour la proposition ??.

Démonstration Comme $C_n^k = C_n^{n-k}$, la somme présentée est le coefficient c_n du produit de Cauchy suivant :

$$f_p(x) = \sum_{m \geq 0} c_m x^m = \underbrace{\left(\sum_{k=0}^n C_n^k (-1)^k x^k \right)}_{(1-x)^n} \underbrace{\left(\sum_{l \geq 0} l^p x^l \right)}_{g_p(x)},$$

où toutes les séries entières présentées ont un rayon de convergence d'au moins 1. La famille de fonctions $(g_p)_{p \geq 0}$ est également définie par récurrence comme suit :

$$g_0(x) = \frac{1}{1-x} \text{ et } g_{p+1}(x) = x g_p'(x).$$

LEMME 0.2

Pour tout entier naturel p , il existe un polynôme h_p de degré p tel que $g_p(x) = h_p(x)/(1-x)^{p+1}$.

begindivdemonstrationbeginintext endtext Récurrence évidente.

Si $p < n$, alors $f_p(x) = h_p(x)(1-x)^{n-p-1}$ est un polynôme de degré $(n-1)$, si bien que $c_n = 0$. Supposons maintenant que $p = n$. On a alors :

$$f_n(x) = \frac{h_n(x)}{1-x} = h_n(x)(1+x+x^2+\dots).$$

Comme h_n est de degré n , le coefficient c_n vaut la somme des coefficients de h_n . Autrement dit, $c_n = h_n(1)$. On déduit de la relation $g_{p+1}(x) = x g_p'(x)$ que

$$\frac{h_{p+1}(x)}{(1-x)^{p+2}} = \frac{x h_p'(x)}{(1-x)^{p+1}} + \frac{(p+1)x h_p(x)}{(1-x)^{p+2}}.$$

En multipliant par $(1-x)^{p+2}$ puis en prenant $x = 1$, il vient $h_{p+1}(1) = (p+1)h_p(1)$. Comme $h_0(1) = 1$, on obtient $h_p(1) = p!$ pour tout p . En particulier, $c_n = h_n(1) = n!$ et la proposition est démontrée.*enddivdemonstration*

Généralisation du binôme de Newton Dans un anneau commutatif, pour n non nul,

$$(x_1 + \dots + x_p)^n = \sum_{i_1+i_2+\dots+i_p=n, i_j \geq 0} \left(\frac{n!}{i_1!i_2!\dots i_p!} \right) x_1^{i_1} x_2^{i_2} \dots x_p^{i_p}$$

Le coefficient $\frac{n!}{i_1!i_2!\dots i_p!}$ pouvant se noter C_n^i avec $i = (i_1, \dots, i_p)$.

Application 0.2 Ces nombres apparaissent par exemple dans le calcul du nombre d'anagrammes d'un mot.

Un mot constitué de N lettres deux à deux différentes possède $N!$ anagrammes.

Par exemple, le mot « matheux » possède $7! = 5040$ anagrammes.

Mais plus généralement, considérons un mot de N lettres est constitué de i_1 fois une première lettre, de i_2 fois une deuxième droite, ..., de i_p fois une p -ième lettre (les p lettres étant deux-à-deux différentes) avec donc $N = \sum_{k=1}^p i_k$. Un tel mot possèdera

$$\frac{N!}{i_1!i_2!\dots i_p!} = C_N^{(i_1, i_2, \dots, i_p)}$$

anagrammes.

Par exemple, le mot « anticonstitutionnellement » possède

$$\frac{25!}{3!3!2!5!2!5!} = 7480328917501440000$$

anagrammes (environ 7,5 milliards de milliards).

Voici un programme Maple capable de calculer le nombre d'anagrammes d'un mot.

Exemple Maple

```
> Nb_Anagrammes := proc(mot) # Le mot doit être constitué de minuscules
> local occur, i, Nb; for i from "a" to "z" do    occur[i] := 0; od; for i from 1 to length(mot) do
occur[mot[i]] := occur[mot[i]] + 1; od; Nb := length(mot)!; for i from "a" to "z" do    Nb := Nb/oc-
cur[i]!; od; Nb; end;
> Nb_Anagrammes("matheux"); 5040
> Nb_Anagrammes("anticonstitutionnellement"); 7480328917501440000
```

DÉFINITION 0.3 **sommable de somme** x

Une famille $(x_i)_{i \in I}$ de nombres complexes est **sommable de somme** x si pour tout $\epsilon > 0$ il existe $J \subset I$ finie telle que, pour tout K fini, $J \subset K \subset I$ implique $|x - \sum_{i \in K} x_i| \leq \epsilon$.

LEMME 0.3

Si une famille (x_i) de nombres réels est sommable, alors la famille de ses termes positifs est sommable, et la famille de ses termes négatifs est sommable.

Familles sommables infinies

Démonstration Supposons que la famille des (x_i) soit sommable, et supposons que la famille des $(\max(x_i, 0))$ ne le soit pas. Alors la somme des $(\max(x_i, 0))$ pour $i \in J$ avec J fini peut être arbitrairement grande. Or la somme des (x_i) pour i dans $J' = \{j \in J; x_j > 0\}$ est tout aussi grande, et peut donc être arbitrairement grande elle aussi. D'où le résultat pour la famille des réels positifs. Le raisonnement pour la famille négative est le même.

PROPOSITION 0.4

Toute famille sommable de nombres réels est de support dénombrable (i.e. seule une quantité au plus dénombrable de ces réels est non nulle). Il en va de même des familles de nombres complexes.

Démonstration En vertu du lemme précédent, on se contente de démontrer ce résultat pour une famille de nombres réels positifs. Le résultat dans le cas général s'obtient par le lemme précédent.

Il existe un nombre fini de réels plus grands que $1/n$, pour tout n . En notant A_n la famille des réels $> 1/n$, on voit que la réunion des A_n est le support de la famille; une réunion dénombrable d'ensembles finis étant dénombrable, la famille est dénombrable.

1.2 Espaces mesurés

On trouvera en ?? les fondements de la topologie, et en ?? les fondements de la théorie de la mesure. On rappelle ci-dessous, pour le confort du lecteur, quelques définitions qui sont données dans les paragraphes précédemment cités.

Une **topologie** sur X est un sous-ensemble de $P(X)$ contenant \emptyset , X , et stable par réunion quelconque et par intersection finie. Les éléments d'une topologie sont appelés **ouverts**, leurs complémentaires sont appelés **fermés**.

Une **algèbre** sur X est un sous-ensemble de $P(X)$ contenant X , stable par passage au complémentaire et stable par union finie.

Une **tribu** sur X est un sous-ensemble de $P(X)$ contenant X , stable par passage au complémentaire et par union dénombrable. Une tribu est aussi appelée **σ -algèbre**.

Un **espace mesurable** est un couple (X, \mathcal{A}) avec \mathcal{A} tribu sur X .

Dans \mathbb{R} ou dans \mathbb{R}^n , ou en général dans un espace topologique, la tribu usuelle est la tribu engendrée par les ouverts. Dans le cas de \mathbb{R}^n et \mathbb{R} , cette tribu est aussi la tribu engendrée par les boules ouvertes. Une tribu engendrée par une topologie s'appelle **tribu des boréliens**; ses éléments s'appellent les **boréliens**.

Une mesure positive sur un espace mesurable (X, \mathcal{A}) est une fonction μ de \mathcal{A} dans $\overline{\mathbb{R}^+}$ telle que $\bullet \mu(\emptyset) = 0$

\bullet Si les A_i sont deux à deux disjoints et I dénombrable alors $\mu(\cup_{i \in I} A_i) = \sum_{i \in I} \mu(A_i)$

Un **espace mesuré** est un triplet (X, \mathcal{A}, μ) avec \mathcal{A} une tribu sur X , μ une mesure positive sur (X, \mathcal{A}) .

Une fonction de (X_1, \mathcal{A}_1) dans (X_2, \mathcal{A}_2) est dite **mesurable** si et seulement si l'image réciproque de tout ensemble mesurable est mesurable.

La σ -algèbre engendrée par une base d'ouverts d'une topologie est égale à la σ -algèbre engendrée par cette topologie.

Une mesure est dite **finie** si et seulement si la mesure de l'espace tout entier X est finie et alors pour tout A mesurable $\mu(A) \leq \mu(X)$.

Une mesure est une **mesure de probabilité** si la mesure de l'espace tout entier est 1.

1.3 Événements

On présente dans cette section la notion d'événements; définitions de base, exemples de mesures de probabilité.

1.3.1 Définitions de base

DÉFINITION 0.4 Définitions de base

On appelle **triplet de probabilité** ou **espace probabilisé** un triplet (Ω, \mathcal{F}, P) où \mathcal{F} est une tribu sur Ω et P une mesure de probabilité sur (Ω, \mathcal{F}) .

Ω est appelé **l'univers**.

Un élément de Ω est appelé **possible**.

On appelle **événement** une partie mesurable de Ω , c'est-à-dire un élément de \mathcal{F} , c'est-à-dire une partie \mathcal{F} -mesurable.

PROPOSITION 0.5

Si chaque $(F_n)_{n \in \mathbb{N}}$ est de mesure 1, c'est-à-dire que chaque événement F_n est réalisé presque sûrement¹, alors $\cap_{n \in \mathbb{N}} F_n$ se réalise avec une probabilité 1.

Par ailleurs, si E_n est une suite d'événements tels que $\sum_i P(E_i) < +\infty$, alors $P(\limsup_{n \rightarrow +\infty} E_n) = 0$. Ce résultat est connu sous le nom de premier lemme de **Borel-Cantelli**.

On notera une nouvelle façon de voir \limsup des E_n , avec les E_n des événements ; en l'écrivant $\cap_k \cup_{k \leq n} E_n$, on voit maintenant cette limite comme l'événement qui arrive « infiniment souvent » ; c'est l'ensemble des possibles ω qui appartiennent à une infinité de E_n .

De même on peut voir différemment \liminf des E_n , avec les E_n des événements ; en l'écrivant $\cup_k \cap_{k \leq n} E_n$, on voit $\liminf E_n$ comme l'ensemble des possibles ω tels que ω est dans tout E_n pour n assez grand ($\geq N_\omega$ avec N_ω dépendant de ω).

On peut trouver ici des corollaires du lemme de Fatou ; notamment les deux propriétés suivantes :

- $P(\liminf E_n) \leq \liminf P(E_n)$
- $P(\limsup E_n) \geq \limsup P(E_n)$

La première propriété est vraie dans le cas d'un espace mesuré quelconque ; la seconde demande que la mesure soit finie, ce qui est donc le cas dans le cadre d'un espace de probabilité.

1.3.2 Quelques mesures de probabilité

Ensemble fini Lorsque l'univers est fini, on peut prendre (et on prend usuellement) pour tribu l'ensemble de toutes les parties de l'univers. Par exemple, si l'on lance 3 fois une pièce, et que l'on peut obtenir à chaque fois pile ou face, l'univers est :

$$\{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}.$$

On peut dans ce cas prendre pour mesure l'application qui à un ensemble E associe égale à $\frac{\text{card } E}{\text{card } \Omega}$. L'univers contient des éléments correspondant à chaque manière dont peut se réaliser le phénomène aléatoire étudié (dépend de la finesse de la description). La structure de Ω lui-même est secondaire, ce sont les variables aléatoires définies dessus qui importent (pourvu que Ω soit suffisamment vaste pour les définir suffisamment fines).

Distribution sur $[0, 1]^n$ On peut utiliser comme mesure sur $[0, 1]^n$ la restriction d'une mesure sur les boréliens ou les lebesguiens de \mathbb{R}^n telle que la mesure de $[0, 1]^n$ soit 1 ; c'est en particulier le cas de la mesure de Lebesgue. On généralise facilement la méthode à une partie mesurable (de mesure > 0) de \mathbb{R}^n , en divisant par la mesure de la partie.

1.4 Variables aléatoires

Après quelques définitions, on parlera d'indépendance, d'espérance et autres moments, des différentes formes de convergence, et de l'importante notion d'espérance.

1.4.1 Définitions : variable aléatoire, loi, fonction de répartition

DÉFINITION 0.5 Variable aléatoire

Une **variable aléatoire** est une fonction mesurable d'un univers vers \mathbb{R} (muni de sa tribu borélienne pour la topologie usuelle).

On note que l'on parle ici de boréliens plutôt que de lebesguiens. Cela est conforme à la tradition, et n'est pas toujours sans conséquence ; la composée $f \circ v$ d'une variable aléatoire v avec une fonction mesurable f de domaine \mathbb{R} (mesurable au sens des Lebesguiens) n'est pas, en général, une variable aléatoire car f n'est pas forcément mesurable au sens des boréliens ! Voir notamment la définition 0.13 des mesures images pour voir les prudences qu'il faut parfois avoir.

Notons qu'en toute généralité, on peut définir des variables aléatoires comme étant des fonctions mesurables d'un univers (muni d'une σ -algèbre) vers un domaine muni d'une σ -algèbre.

On peut voir alors de nombreux outils qui seront des variables aléatoires ; par définition il suffit que $f^{-1}(B)$ soit mesurable pour tout B borélien pour que f soit une variable aléatoire (il faut bien entendu que le domaine soit un univers, donc que la mesure de l'espace soit 1). Toutes les opérations sur des fonctions à variables réelles qui conservent la mesurabilité sont alors possibles pour construire des variables aléatoires ; la somme, le produit, la valeur absolue... Ce qui est cohérent par rapport à la notion intuitive de variable aléatoire ; si le résultat d'un tirage aléatoire est un réel et si l'on répète dix fois ce tirage aléatoire, alors la somme des résultats de ces dix tirages est une variable aléatoire, le produit aussi ; ils ont eux aussi leurs *distributions de probabilité (notion définie plus tard, définition 0.14)*.

DÉFINITION 0.6 loi de probabilité \mathcal{L}_X

Supposons que X soit une variable aléatoire sur un triplet de probabilité (Ω, \mathcal{F}, P) . X est alors une application de Ω dans \mathbb{R} , et P est une application de \mathcal{F} dans $[0, 1]$. Alors on définit la **loi de probabilité** \mathcal{L}_X de X par $\mathcal{L}_X = P \circ (X^{-1})$ (X^{-1} n'est pas l'application réciproque – non nécessairement bien définie – mais l'application qui à une partie associe son image réciproque) ; \mathcal{L}_X est ainsi définie sur l'ensemble des boréliens de \mathbb{R} .

$\mathcal{L}_X(A)$ est ainsi la probabilité pour que X soit dans A .

Des définitions découle la proposition suivante :

PROPOSITION 0.6

\mathcal{L}_X est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$.

DÉFINITION 0.7 variables aléatoires indépendantes

Soit $F_X(t) = \mathcal{L}_X([-\infty, t])$; alors F_X est appelée **fonction de répartition** de X .

PROPOSITION 0.7

La donnée de F_X détermine \mathcal{L}_X de manière unique.

Démonstration $\{[-\infty, t]; t \in \mathbb{R}\}$ est un π -système qui engendre l'ensemble des boréliens. Donc par le lemme ??, \mathcal{L}_X est entièrement défini par la fonction $F_X(t) = L_X([-\infty, t]) = P(\{\omega \in \Omega; X(\omega) \leq t\})$.

PROPOSITION 0.8 Caractérisation des fonctions de répartition

$F : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction de répartition d'une certaine variable aléatoire si et seulement si les quatre propriétés suivantes sont vérifiées :

- F est croissante de \mathbb{R} dans $[0, 1]$
- $F(x) \rightarrow 1$ quand $x \rightarrow +\infty$
- $F(x) \rightarrow 0$ quand $x \rightarrow -\infty$
- F est continue à droite en tout point.

Démonstration Le sens « seulement si » ne pose pas de problème : les trois premiers points sont clairs, le quatrième utilise le fait que l'ensemble des ω inférieurs à $x + \frac{1}{n}$ « tend » vers l'ensemble des ω inférieurs à x , en décroissant (en effet la mesure d'une suite décroissante d'ensembles mesurables tend vers la mesure de l'intersection, par le théorème de convergence dominée ??).

Le sens « si » est plus délicat ; la loi X correspondant à F est définie par $X(\omega) = \inf \{z; F(z) \geq \omega\}$. Il convient alors de vérifier que X ainsi défini est mesurable.

1.4.2 Variables aléatoires indépendantes

Intuitivement des variables aléatoires sont indépendantes lorsqu'aucune d'elles ne dépend, d'aucune façon, des autres. Par exemple dans un sondage en vue d'un référendum, il est souhaitable que la variable « être sondé » soit indépendante de la variable « être partisan du oui » (chose très difficile à réaliser en pratique), sans quoi le sondage risque d'être biaisé. La notion d'indépendance est formalisée ci-dessous (et on formalisera la notion de sondage « biaisé » plus loin, en section « statistiques »).

DÉFINITION 0.8 σ -algèbres indépendantes

Soit S une σ -algèbre, et $(S_i)_{i \in I}$ une famille de sous- σ -algèbres de S ; alors les S_i sont dites σ -algèbres indépendantes si pour toute famille $(s_i)_{i \in J} \in \pi_{i \in J} S_i$ avec J partie finie de I , on a $P(\cap_{i \in J} s_i) = \pi_{i \in J} P(s_i)$.

Soit $(X_i)_{i \in I}$ une famille de variables aléatoires ; alors les X_i sont dites **variables aléatoires indépendantes** si les S_i sont des σ -algèbres indépendantes avec $S_i = X_i^{-1}(\mathcal{B})$ (\mathcal{B} la σ -algèbre borélienne de \mathbb{R}).

Des événements E_i sont dits **événements indépendants** si les σ -algèbres $\{E_i, \Omega \setminus E_i, \emptyset, \Omega\}$ sont indépendantes ; ce qui équivaut au fait que les fonctions caractéristiques des E_i , en tant que variables aléatoires, sont indépendantes.

On appelle suite de **variables aléatoires identiquement distribuées** (en abrégé i.i.d.) une suite de variables aléatoires indépendantes et ayant la même fonction de répartition.

On n'ira pas ici jusqu'à cette construction, mais il est possible de prouver (théorème d'extension de Kolmogorov) qu'étant donnée une suite de distributions de probabilités $d_1, d_2, \dots, d_n, \dots$, on peut construire une suite de variables aléatoires indépendantes x_1, x_2, \dots telles que x_i ait distribution d_i et que les x_i soient indépendantes. Mais attention ! cela est valable pour une suite, mais pas pour une famille quelconque. En effet, l'étude des champs aléatoires fournit des contre-exemples ; il s'agit d'un cadre où l'on cherche à définir une « variable aléatoire » mais à valeurs non pas dans \mathbb{R} ou \mathbb{R}^n mais dans $\mathbb{R}^{\mathbb{R}^n}$ ou \mathbb{R}^D avec $D \subset \mathbb{R}^n$; c'est-à-dire que chaque réalisation de la « variable aléatoire » est une fonction e.g. de \mathbb{R}^n dans \mathbb{R} . On est tenté d'appliquer le théorème d'extension de Kolmogorov pour définir par exemple, des variables de Bernoulli identiquement distribuées et indépendantes, non constantes, indexées par \mathbb{R} . Et bien c'est impossible ! Il faut à coup sûr certaines dépendances. Évidemment, si on ne requiert pas l'indépendance, on peut par exemple prendre des variables aléatoires de Bernoulli presque sûrement identiques. C'est donc bien « l'excès » d'indépendance qui pose souci. On pourra consulter [?] pour des compléments sur ce passionnant sujet.

PROPOSITION 0.9

Pour que des événements E_i pour $i \in I$ soient indépendants il suffit de vérifier que $P(\cap_{i \in J} E_i) = \prod_{i \in J} P(E_i)$ pour tout J fini dans I .

Démonstration Il suffit de considérer les propriétés d'additivité de P , pour voir que cette formule permet de déduire les cas où des E_i sont remplacés par leurs complémentaires. Le fait que seules des intersections finies sont à considérer est directement la définition de l'indépendance.

DÉFINITION 0.9 indépendants

Deux π -systèmes P_1 et P_2 sur un même ensemble sont dits **indépendants** si pour tout $p_1 \in P_1$ et tout $p_2 \in P_2$ on a

$$P(p_1 \cap p_2) = P(p_1) \cdot P(p_2)$$

LEMME 0.10

Supposons S_1 et S_2 deux σ -algèbres sur X engendrées respectivement par P_1 et P_2 des π -systèmes. Alors S_1 et S_2 sont indépendantes si et seulement si P_1 et P_2 sont des π -systèmes indépendants.

Démonstration Le sens « seulement si » étant immédiat on se préoccupe de l'autre sens.

Étape 1. Fixons p_1 dans P_1 , et considérons les mesures m_1 et m_2 définies sur S_2 par

$$m_1(E) = P(E \cap p_1) \text{ et } m_2(E) = P(E).P(p_1).$$

Ces deux mesures coïncident sur P_2 et donnent une mesure finie à X ; donc par le lemme ?? elles sont égales sur S_2 .

Étape 2. Fixons maintenant p_2 dans S_2 . On définit maintenant les deux mesures m_3 et m_4 sur S_1 par

$$m_3(E) = P(E \cap p_2) \text{ et } m_4(E) = P(E).P(p_2).$$

Elles coïncident sur P_1 (grâce au résultat de l'étape 1) et donnent une mesure finie à X ; donc par le lemme ?? elles sont égales.

Étape finale. Résumons. On a donc montré l'équation 0.9 pour p_1 et p_2 dans les cas suivants :

- $p_1 \in P_1$ et $p_2 \in P_2$, clair par hypothèse.
- $p_1 \in P_1$ et $p_2 \in S_2$, dans la première étape ci-dessus (grâce au lemme ??).
- p_1 quelconque dans S_1 et p_2 quelconque dans S_2 , en fixant p_2 , dans la deuxième étape (en utilisant le lemme ?? là-aussi).

Le résultat est donc prouvé.

LEMME 0.11 Second lemme de Borel-Cantelli

Soit $(E_n)_{n \in \mathbb{N}}$ une suite d'événements indépendants.

$$\text{Si } \sum_n P(E_n) = +\infty, \text{ alors } P(\limsup E_n) = 1.$$

Intuitivement, cela signifie que si la somme des probabilités pour qu'un événement arrive à l'instant n pour $n \in \mathbb{N}$ tend vers $+\infty$, alors l'événement a une probabilité 1 d'avoir lieu une infinité de fois.

Démonstration Notons que

$$\begin{aligned} (\limsup E_n)^c &= (\cap_m \cup_{n \geq m} E_n)^c \\ &= \cup_m ((\cup_{n \geq m} E_n)^c) \\ &= \cup_m \cap_{n \geq m} (E_n^c) = \liminf (E_n^c) \end{aligned}$$

(où F^c est le complémentaire de F)

Avec $p_n = P(E_n)$, pour tout $p \geq m$, par définition de l'indépendance, on a

$$P(\cap_{p \geq n \geq m} E_n^c) = \pi_{p \geq n \geq m}(1 - p_n).$$

Donc en passant à la limite, par monotonie de l'intersection des E_n^c , on a

$$P(\cap_{n \geq m} E_n^c) = \pi_{n \geq m}(1 - p_n).$$

$1 - x \leq \exp(-x)$, donc $\pi_{n \geq m}(1 - p_n) \leq \pi_{n \geq m} \exp(-p_n) \leq \exp(-\sum_{n \geq m} p_n) \leq 0$, d'où le résultat.

Le premier lemme de Borel-Cantelli, évoqué en proposition 0.5, fournit un complément (une forme de réciproque, quoique les hypothèses ne soient pas exactement les mêmes car le premier lemme ne requiert pas l'indépendance) à ce lemme.

Exemple 0.3 Application des deux lemmes de Borel-Cantelli

On définit les $(E_n)_{n \geq 1}$ comme des événements aléatoires, indépendants, par

$$E_n = X_n \in [\alpha \cdot \log(n), +\infty[$$

avec X_n variables aléatoires indépendantes définies par leur fonction de répartition $F_X(x) = \max(1 - \exp(-x), 0)$.

$P(E_n) = n^{-\alpha}$, donc $\sum_{n \geq 1} P(E_n) = +\infty$ si et seulement si $\alpha \leq 1$, et donc par les deux lemmes de Borel-Cantelli E_n a lieu infiniment souvent (c'est-à-dire que $P(\limsup E_n) = 1$) si et seulement si $\alpha \leq 1$.

Intuitivement, pour chaque instant $n \geq 1$, on tire au sort une réalisation indépendante x_n de la loi X , et on dit que E_n a lieu si $x_n \geq \alpha \log(n)$. Alors on a deux conclusions :

- si et seulement si $\alpha \leq 1$, presque sûrement, E_n a lieu pour une infinité de valeurs de n ;
- en outre, si et seulement si $\alpha > 1$, presque sûrement, il y a un nombre fini de valeurs de n telles que E_n a lieu.

Il est important de voir que ces deux lignes ne sont pas deux formulations de la même conclusion. On aurait pu imaginer que pour certaines valeurs de α , on a une chance sur deux d'avoir un nombre fini de valeurs de n telles que E_n a lieu. On vient en fait de montrer que ce n'est pas possible. Pour toute valeur de α , soit presque sûrement E_n a lieu une infinité de fois, soit presque sûrement E_n n'a lieu qu'un nombre fini de fois. Ceci est une illustration simple et amusante d'un phénomène plus général, la loi 0-1 de Kolmogorov, que nous allons voir ci-dessous.

DÉFINITION 0.10 σ -algèbre asymptotique

Étant donnée une suite de variables aléatoires X_1, \dots, X_n, \dots , on appelle **σ -algèbre asymptotique de la suite $(X_n)_{n \in \mathbb{N}}$** la σ -algèbre $\bigcap_n \tau_n$, avec τ_n la σ -algèbre engendrée par $(X_{n+1}, X_{n+2}, \dots)$.

Pour bien comprendre cette définition il faut voir que :

- τ_n est la σ -algèbre qui rend toutes les variables aléatoires X_i mesurables pour $i > n$.
- τ est l'intersection des τ_n .

Intuitivement la σ -algèbre asymptotique contient les événements qui ne dépendent que du comportement à la limite.

PROPOSITION 0.12

Les événements suivants sont par exemples mesurables pour la σ -algèbre asymptotique des X_i (on les appelle **événements asymptotiques**) :

- $\{\omega; \lim_{n \rightarrow +\infty} X_i(\omega) \text{ existe}\}$
- $\{\omega; \sum_{n \rightarrow +\infty} X_i(\omega) \text{ existe}\}$
- $\{\omega; \lim_{n \rightarrow +\infty} \sum_{i \in [1, n]} X_i(\omega) / n \text{ existe}\}$

Les variables aléatoires suivantes sont τ -mesurables (on les appelle **variables asymptotiques**) :

- $\limsup_{n \rightarrow +\infty} \sum_{i \in [[1, n]]} (X_i(\omega)/n)$
- $\liminf_{n \rightarrow +\infty} \sum_{i \in [[1, n]]} (X_i(\omega)/n)$

Pour contre-exemples on peut citer par exemple X_{10} (variable aléatoire non-asymptotique dans le cas général), ou la somme des X_i pour $0 \leq i \leq 10$.

Démonstration Pour les trois premiers points il faut et il suffit donc de montrer que l'ensemble E donné est inclus dans chaque τ_n .

• Pour $E = \{\omega; \lim_{n \rightarrow +\infty} X_i(\omega) \text{ existe}\}$, les X_i pour $i > n$ sont τ_n -mesurables, donc $\liminf X_i$ et $\limsup X_i$ sont τ_n -mesurables, donc

$$(\liminf X_i - \limsup X_i)^{-1}(\{0\})$$

est une partie τ_n -mesurable. Donc $E \in \tau_1$; de la même manière $E \in \tau_n$, pour tout n , donc $E \in \tau$.

• Pour $E = \{\omega; \sum_{n \rightarrow +\infty} X_i(\omega) \text{ existe}\}$, les X_i pour $i > n$ sont τ_n -mesurables, donc une somme finie de X_i est τ_n -mesurable, donc

$$\limsup_{m \rightarrow +\infty} \sum_{i=n+1}^{n+m} X_i$$

est τ_n -mesurable, pareil avec \liminf , d'où le résultat en considérant

$$(\limsup_{m \rightarrow +\infty} \sum_{i=n+1}^{n+m} X_i - \limsup_{m \rightarrow +\infty} \sum_{i=n+1}^{n+m} X_i)^{-1}(\{0\}).$$

• Pour $E = \{\omega; \lim_{n \rightarrow +\infty} \sum_{i \in [[1, n]]} X_i(\omega)/n \text{ existe}\}$, par une méthode similaire aux deux cas précédents on montre que E appartient à τ_1 , il suffit de voir ensuite que $\lim_{n \rightarrow +\infty} \sum_{i \in [[1, n]]} X_i(\omega)/n$ converge si et seulement si $\lim_{n \rightarrow +\infty} \sum_{i \in [0, 1]} X_i(\omega)/(n - K + 1)$ converge pour conclure que E appartient aussi à τ_K .

Pour les variables aléatoires qui suivent les façons de raisonner sont les mêmes.

THÉORÈME 0.13 Loi 0 – 1 de Kolmogorov

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes définies sur le même espace de probabilité, et soit τ la σ -algèbre asymptotique des X_n ; alors :

- tout événement asymptotique a une probabilité 0 ou 1.
- pour toute variable asymptotique Y , il existe un unique $z \in [-\infty, +\infty]$ tel que $P(Y = z) = 1$.

Démonstration On procède par étapes :

• On montre tout d'abord que les σ -algèbres suivantes sont indépendantes pour tout n :

- la σ -algèbre engendrée par X_1, \dots, X_n , notée par la suite Y_n .
- la σ -algèbre engendrée par X_{n+1}, X_{n+2}, \dots , notée comme d'habitude τ_n .

En effet :

- la première de ces deux σ -algèbres est engendrée par le π -système des ensembles de la forme $\{\omega; \forall i \in [[1, n]] X_i(\omega) \leq x_i\}$ avec $x_i \in]-\infty, +\infty]$.
- la seconde de ces deux σ -algèbres est engendrée par le π -système des ensembles de la forme $\{\omega; \forall j \in [[n+1, n+1+K]] X_j(\omega) \leq x_j\}$ avec $x_j \in]-\infty, +\infty]$.

Par le lemme 0.10, nos deux σ -algèbres sont donc indépendantes.

• Y_n et τ sont indépendantes; en effet, $\tau \subset \tau_n$, donc l'indépendance de τ_n et de Y_n entraîne l'indépendance de τ et Y_n .

• τ_1 et τ sont indépendantes.

L'union des Y_n est un π -système (facile), qui engendre τ_1 (facile aussi). D'après l'étape précédente, l'union des Y_n et τ sont indépendantes au sens des π -systèmes; donc les σ -algèbres engendrées sont indépendantes.

• τ étant inclus dans τ_1 , τ est indépendante de τ (ce n'est pas une faute de frappe!), et donc pour tout $E \in \tau$, on a

$$P(E) = P(E \cap E) = P(E).P(E).$$

Le premier des deux points est alors prouvé. Pour trouver z du second point il suffit de considérer le supremum des $z \in \mathbb{R}$ tels que $P(Y \leq z) = 0$.

Application 0.4 On trouvera une belle illustration avec l'exemple 0.3; ce n'est cependant pas à strictement parler une application, car on peut là-bas se passer du théorème général de la loi 0-1 de Kolmogorov.

1.4.3 Convergence de variables aléatoires

On consultera la partie ?? et la partie ?? pour avoir les définitions nécessaires à la bonne compréhension des notions de convergences de variables aléatoires; les mêmes notions de convergence sont utilisées ici. Il faut bien voir les variables aléatoires comme des fonctions dont le domaine est un espace mesuré (et même un espace probabilisé) pour bien comprendre pourquoi les définitions sont les mêmes que celles utilisées pour des mesures. On rappelle ici simplement la convergence en loi, que l'on utilisera intensivement dans ce chapitre.

DÉFINITION 0.11 Convergence en loi

La convergence **en loi** ou **en distribution** de X_n (suite de variables aléatoires) vers X (variable aléatoire) est la convergence, pour toute fonction f continue bornée, de $Ef(X_n)$ vers $Ef(X)$.

La convergence en loi a la force de ne pas nécessiter que l'on ait égalité des espaces de probabilité (pour les variables aléatoires X_n et X). En outre, elle est plus faible que la plupart des autres convergences (voir [?] pour une exploitation des convergences en loi).

On peut discuter un peu plus cette histoire d'égalité ou pas égalité des espaces de probabilité. Deux variables aléatoires ne peuvent être égales presque sûrement que si elles sont définies sur le même espace probabilisé. Par contre, elles peuvent être égales en distribution (avoir la même distribution) sans être définies sur le même espace probabilisé.

1.4.4 Espérance, variances et autres moments

L'espérance d'une variable aléatoire est sa « valeur moyenne ». Définissons cela formellement.

DÉFINITION 0.12 Espérance d'une variable aléatoire dans L^1

Étant donnée X une variable aléatoire de $L^1(X, \mathbb{R})$, on définit son espérance par

$$E(X) = \int_{\Omega} X.dP.$$

Cette définition peut éventuellement être étendue aux fonctions intégrables positives, sans contrainte de mesurabilité.

On définit en outre $E(X; F)$, avec X une variable aléatoire \mathcal{L}^1 ou bien une variable aléatoire intégrable positive, et F une partie mesurable, par

$$E(X; F) = \int_F X.dP = E(X.\chi_F).$$

avec χ_F la fonction caractéristique de F .

Définitions On a bien deux cas distincts, même si cela est un peu pénible : \mathcal{L}^1 , ou intégrable positive. Voir la discussion qui suit la définition ?? de l'intégrale de Lebesgue.

THÉORÈME 0.14 Théorèmes de passage à la limite en probabilités

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires et X une variable aléatoire telles que

$$P(X_n \rightarrow X) = 1 \text{ c'est-à-dire } P(\{\omega; X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

Alors les résultats de convergence monotone, de Fatou, de convergence dominée et de Scheffé, que l'on peut trouver dans la partie ??, se reformulent comme suit :

• Convergence monotone :

Si les X_n sont ≥ 0 et $X_n(\omega)$ croît vers $X(\omega)$ pour $n \rightarrow +\infty$, alors $E(X_n) \rightarrow E(X)$.

• Lemme de Fatou :

Si $X_n \geq 0$ alors $E(X) \leq \liminf E(X_n)$

• Théorème de convergence dominée de Lebesgue :

Si pour tout n et tout ω on a $|X_n(\omega)| \leq |Y(\omega)|$, avec Y une variable aléatoire telle que $E(Y) \leq +\infty$, alors $E(|X_n - X|) \rightarrow 0$, et en particulier $E(X_n) \rightarrow E(X)$.

• Lemme de Scheffé :

Si $E(|X_n|) \rightarrow E(|X|)$, alors $E(|X_n - X|) \rightarrow 0$.

Théorèmes et inégalités

Démonstration Voir le chapitre ?? pour les preuves correspondantes, qui s'appliquent directement.

THÉORÈME 0.15 Inégalité de Markov

Supposons X variable aléatoire, et f mesurable de \mathbb{R} muni des boréliens dans $[0, +\infty]$ muni des boréliens, avec f croissante. Alors

$$E(f \circ X) \geq E(f \circ X; X \geq c) \geq f(c) \cdot \int \chi_{\{\omega; X(\omega) \geq c\}}$$

qu'on peut aussi noter $E(f \circ X) \geq E(f \circ X; X \geq c) \geq f(c) \cdot P(X \geq c)$.

Démonstration Il suffit de bien voir que f est positive et que

$$Ef \circ X = E(f \circ X; X > c) + E(f \circ X; X \leq c).$$

COROLLAIRE 0.16

Avec X une variable aléatoire positive, $E(X) \geq c \cdot P(X \geq c)$

Démonstration C'est l'inégalité de Markov avec f la fonction identité.

COROLLAIRE 0.17

Pour X variable aléatoire positive et $z > 0$, $P(X \geq z) \leq E(X)/z$.

Démonstration Il s'agit simplement de l'inégalité ci-dessus, reformulée.

THÉORÈME 0.18 Inégalité de Jensen

On se donne f une application de U dans \mathbb{R} , avec U intervalle ouvert de \mathbb{R} , et X une variable aléatoire, avec les hypothèses suivantes :

f convexe;
 $P(X \in U) = 1$;
 $E(|X|) < +\infty$ (c'est-à-dire que X est intégrable); $E(|f(X)|) < +\infty$ (c'est-à-dire que $f \circ X$ est intégrable).

Alors : $E(f(X)) \geq f(E(X))$

Application 0.5 Voir par exemple les propriétés des fonctions caractéristiques en probabilités, proposition ??.

Démonstration

• Les dérivées à gauche et à droite de f , notée d^- et d^+ , existent et sont croissantes par convexité de f ; on a en outre $d^-(x) \leq d^+(x)$.

• Considérons maintenant $z \in U$, et $x \in U$.

Soit $x < u < z$, alors la pente entre x et u est inférieure à la pente entre u et z ; en faisant tendre u vers z on constate alors que la pente entre x et z est inférieure à $d^-(z)$. Donc :

$$f(x) \geq f(z) + d^-(z)(x - z).$$

De même pour $x > z$ on montre $f(x) \geq f(z) + d^+(z)(x - z)$

• Comme $d^-(z) \leq d^+(z)$, on peut résumer ces assertions en

$$f(x) \geq f(z) + d^-(z)(x - z)$$

valable pour tout x .

• Il est facile de voir que $E(X) \in U$

• On peut donc spécialiser l'affirmation de l'avant-dernier point en

$$f(x) \geq f(E(X)) + d^-(E(X))(x - E(X)).$$

• En intégrant l'inégalité ci-dessus il vient

$$E(f(X)) \geq f(E(X)).$$

La preuve est ainsi terminée.

Espaces L^p On pourra bien sûr réviser le chapitre ?? sur les espaces L^p en général.

⚠ Attention 0.6 Dans le contexte des probabilités, L^p désignera, étant donné un univers Ω , $L^p(\Omega)$, Ω étant muni d'une mesure de probabilité (L^p est en fait dépendant de l'univers Ω , de la tribu définie sur Ω , et de la mesure définie sur cette tribu). Ne pas généraliser les résultats qui suivent au cas général de $L^p(X)$ pour X espace mesuré quelconque ! Ainsi la proposition 0.19 ci-dessous n'est pas valable en toute généralité.

PROPOSITION 0.19

Pour $p \in [1, +\infty]$, alors $L^{p'} \subset L^p$ pour tout $p' \geq p$ (éventuellement p' infini). En outre pour tout X dans $L^{p'}$, on a $N_{p'}(X) \leq N_p(X)$ avec $N_q(x) = \sqrt[q]{E|x|^q}$.

Démonstration Pour l'inclusion, il suffit de voir la proposition ??.

Pour l'inégalité, on peut clairement se ramener au problème des variables aléatoires positives.

Étant donné X à valeurs positives dans $L^{p'}(X)$, on définit $X_n(\omega) = \min(X(\omega), n)$.

Alors X_n est bornée, et donc $X_n^{p'}$ et X_n^p aussi, donc $X_n^{p'}$ et X_n^p sont dans L^1 (on utilise le fait que la mesure est finie). On peut donc appliquer l'inégalité de Jensen (voir théorème 0.18) avec la variable aléatoire X_n^p et la fonction convexe $x \mapsto x^{p'/p}$, et écrire

$$E(X_n^p)^{p'/p} \leq E(X_n^{pp'/p}) = E(X_n^{p'}) \leq E(X^{p'}).$$

On applique alors le théorème de convergence monotone à X_n^p et

$$(E(X^p))^{p'/p} \leq E(X^{p'}).$$

En élevant à la puissance $1/p'$ on a alors $N_p(X) \leq N_{p'}(X)$.

La preuve est ainsi terminée.

Les résultats usuels dans L^p sont valables, notamment l'inégalité de Schwartz, de Hölder, de Minkowski, pour lesquels on consultera la partie ??.

Pour rappeler l'essentiel :

• Si X et Y sont des variables aléatoires de L^2 , alors le produit $X.Y$ appartient à L^1 , et (inégalité de Schwartz)

$$|E(X.Y)| \leq E(|X.Y|) \leq N_2(X).N_2(Y).$$

• Si X et Y sont des variables aléatoires de L^2 , alors la somme $X+Y$ appartient à L^2 , et (inégalité triangulaire)

$$N_2(X+Y) \leq N_2(X) + N_2(Y).$$

Une proposition est nécessaire pour bien comprendre ce qu'il se passe :

PROPOSITION 0.20

Soit X une variable aléatoire, et soit f une fonction mesurable de \mathbb{R} dans \mathbb{R} , alors $f \circ X$ est une variable aléatoire de L^1 (au sens donné ici, c'est-à-dire $L^1(\Omega)$, avec Ω muni d'une mesure de probabilité) si et seulement si f est dans $L^1(\mathbb{R}, L_X)$ avec L_X la loi de X .

$$\text{On a alors } E(f \circ X) = \int f(x).dL_X.$$

Voir la proposition 0.6 et la définition qui la précède pour bien cerner ce qu'est une loi de probabilité.

Démonstration La démonstration est comme suit :

• Si f est une fonction caractéristique d'un borélien, il s'agit simplement de la définition de la loi de probabilité.

• Si f est simple, alors par linéarité la propriété est aussi vraie.

• Si f est positive, alors f est limite croissante de fonctions simples, donc on peut appliquer le théorème de convergence monotone.

• Enfin dans le cas général, f s'écrit comme différence de deux fonctions mesurables positives.

DÉFINITION 0.13 Mesure image

Étant donnée f une application mesurable d'un espace Ω doté d'une mesure μ dans \mathbb{R} muni des boréliens, on note μ^f la mesure appelée **mesure image de μ par f** définie sur l'ensemble des boréliens de \mathbb{R} par

$$\mu^f(E) = \mu(f^{-1}(E)).$$

Il s'agit bien d'une mesure.

THÉORÈME 0.21 Théorème de transport

Pour toute fonction mesurable ϕ de \mathbb{R} (muni des boréliens) dans \mathbb{R} (muni des boréliens),

$$\int_{\mathbb{R}} \phi d\mu^f = \int_{\Omega} \phi \circ f d\mu.$$

On ramène ainsi les intégrales du type $\int_{\Omega} dP$ à des intégrales sur \mathbb{R} pour la mesure de Lebesgue ; on n'a pas besoin de connaître la structure de Ω , mais seulement les lois.

Démonstration Le chapitre sur l'intégration permet de comprendre clairement les notions en jeu. Il s'agit en fait simplement de vérifier la formule dans le cas d'une fonction caractéristique d'un borélien, puis d'un le cas d'une fonction simple (i.e. étagée² et mesurable) grâce à la linéarité de l'intégrale, puis pour une fonction positive par passage au sup, puis dans le cas général en exprimant une fonction comme différence de deux fonctions l'une positive et l'autre négative (utilisation du théorème de convergence monotone à la fois pour les fonctions simples tendant vers ϕ et pour les fonctions simples tendant vers $\phi \circ f$).

COROLLAIRE 0.22

On peut écrire le même théorème avec une fonction ϕ de \mathbb{R}^n dans \mathbb{R} et f de Ω dans \mathbb{R}^d .

Démonstration Même principe que ci-dessus.

DÉFINITION 0.14 densité de probabilité de X

Étant donné X une variable aléatoire, une application f_X mesurable est appelée une **densité de probabilité de X** si et seulement si pour tout borélien E de \mathbb{R} , on a $P(X^{-1}(E)) = \int_E f_X$.

Densité de probabilité Notons que $\int_{\mathbb{R}} f_X = 1$.

2. Étagée = ne prenant qu'un nombre fini de valeurs.

Variance, covariance, lois jointes, densités jointes, fonctions de répartition jointes

Nous allons ici présenter la variance et des dérivées de la variance. La variance présente notamment un intérêt fort vis à vis des inégalités de concentration, comme l'inégalité de Tchebychev ; il s'agit de dire que les choses de peu de variance sont beaucoup concentrées (autour de leur moyenne). Les inégalités de concentration ont nombre d'applications passionnantes, comme les statistiques [?] ou les processus stochastiques [?].

DÉFINITION 0.15 Covariance et variance

Étant donnée X une variable aléatoire, on définit la **déviatio**n de X par

$$\tilde{X} = X - E(X).$$

Étant données X et Y des variables aléatoires dans L^2 , on définit la **covariance** de X et Y par

$$Cov(X, Y) = E(\tilde{X} \cdot \tilde{Y}).$$

Étant donnée X une variable aléatoire dans L^2 , on définit la **variance** de X par

$$Var(X) = Cov(X, X).$$

Le **produit scalaire de deux variables aléatoires** X et Y de L^2 est l'espérance de $X \cdot Y$ (comme dans le cadre d'un espace L^2 quelconque), noté parfois $\langle X | Y \rangle$.

On appelle **corrél**ation entre deux variables aléatoires X et Y de normes N_2 non nulles le réel de $[-1, 1]$ $cor(X, Y) = \frac{\langle \tilde{X} | \tilde{Y} \rangle}{N_2(X) \cdot N_2(Y)}$. On dit que deux variables sont décorrélées ou non-corrélées lorsque leur corrélation est nulle. On appelle **angle** entre deux variables aléatoires X et Y de normes N_2 non nulles le réel θ appartenant à $[0, \pi]$ tel que $cos(\theta) = \frac{\langle X | Y \rangle}{N_2(X) \cdot N_2(Y)}$.

Plus généralement, deux variables aléatoires sont dites **non corrélées** si leur covariance est nulle (sans forcément que leur corrélation soit bien définie donc).

On appelle **matrice de covariance** d'une suite finie de variables aléatoires (X_1, \dots, X_d) la matrice symétrique M définie par $M_{i,j} = cov(X_i, X_j)$.

La définition de la covariance et de la variance se justifie par le fait que si X est dans L^2 , alors $X - E(X)$ aussi, et donc avec X et Y dans L^2 , $(X - E(X)) \cdot (Y - E(Y))$ est dans L^1 par l'inégalité de Schwartz.

La définition de la corrélation se justifie par l'inégalité de Schwartz.

La corrélation entre deux variables aléatoires est le cosinus de l'angle entre les déviations de ces variables aléatoires (voir définition 0.15).

Notons qu'on a $cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = \langle \tilde{X} | \tilde{Y} \rangle$ et $var(X) = E(X^2) - E(X)^2$.

Si X_1, \dots, X_n sont des variables aléatoires, alors

$$var\left(\sum_{i \in [1, n]} X_i\right) = \sum_{(i, j) \in [1, n]^2} cov(X_i, X_j)$$

$$var\left(\sum_{i \in [1, n]} X_i\right) = \sum_{i \in [1, n]} var(X_i) + \sum_{(i, j) \in [1, n]^2, i \neq j} cov(X_i, X_j)$$

$$\text{var} \left(\sum_{i \in [[1, n]]} X_i \right) = \sum_{i \in [[1, n]]} \text{var}(X_i) + 2 \cdot \sum_{(i, j) \in [[1, n]]^2, i < j} \text{cov}(X_i, X_j).$$

Pour plus d'informations voir la section ?? et plus spécialement ??.

COROLLAIRE 0.23 Inégalité de Tchébitchev

Pour X variable aléatoire, $P(|X - E(X)| > \epsilon) \leq \text{Var}(X)/\epsilon^2$.

Application 0.7 Voir le théorème ?? sur les polynômes de Bernstein.

Démonstration Il suffit d'appliquer le corollaire 0.17 de l'inégalité de Markov à $(X - E(X))^2$.

Le théorème suivant est particulièrement important et beaucoup moins connu que les propriétés de l'espérance d'une somme de variables aléatoires. Attention, contrairement au cas des sommes, ici l'indépendance est utile.

THÉORÈME 0.24 Produit de variables aléatoires indépendantes

Soient X et Y des variables aléatoires indépendantes appartenant à L^1 . Alors $X.Y$ appartient à L^1 et $E(X.Y) = E(X).E(Y)$. Soient X et Y des variables aléatoires indépendantes appartenant à L^2 . Alors :

$$\text{cov}(X, Y) = 0 \text{ et } \text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

Démonstration On se préoccupe tout d'abord du premier résultat.

- Si X et Y sont des fonctions caractéristiques, alors $X = \chi_E$ et $Y = \chi_F$, et $E(X.Y) = P(\chi_{E \cap F}) = P(E).P(F)$ par indépendance.

- Si X et Y sont des fonctions simples alors ce sont des combinaisons linéaires de fonctions caractéristiques, donc le résultat est aussi valable.

- Si X et Y sont positives, alors ce sont des limites croissantes de fonctions simples, donc le résultat est aussi valable par le théorème de convergence monotone.

- Dans le cas général, X et Y s'écrivent comme différences de deux fonctions positives.

La suite se déduit facilement, au vu des définitions de la covariance et de la variance.

⚠ Attention 0.8 Notez bien que X et Y sont supposées dans le premier cas appartenant à L^1 et pas nécessairement à L^2 .

Pour cerner plus précisément l'intérêt de l'indépendance des variables aléatoires, on a besoin de définitions supplémentaires utilisant les mesures produits (voir la partie ?? pour connaître les bases requises).

DÉFINITION 0.16 loi jointe

Étant données X_1, \dots, X_n des variables aléatoires, on appelle

- **loi jointe** de X_1, \dots, X_n ou simplement **loi** de X_1, \dots, X_n l'application L_{X_1, \dots, X_n} qui à un borélien E de \mathbb{R}^n associe $P(F)$ avec $F = \{\omega \in \Omega; (X_1(\omega), \dots, X_n(\omega)) \in E\}$.

• **fonction de répartition** de X_1, \dots, X_n l'application qui à (x_1, \dots, x_n) dans \mathbb{R}^n associe $L_{X_1, \dots, X_n}([-\infty, x_1], \dots, [-\infty, x_n])$.

• **densité de probabilité** ou simplement **densité** de X_1, \dots, X_n une application f (quand elle existe!) de \mathbb{R}^n dans \mathbb{R} telle que pour tout borélien E de \mathbb{R}^n on ait $L_{X_1, \dots, X_n}(E) = \int_E f$. Il s'agit donc en fait simplement de la densité de la loi par rapport à la mesure de Lebesgue. La densité est unique presque sûrement; c'est-à-dire que deux densités d'une même variable aléatoire sont presque sûrement égales.

On note que le théorème de Fubini permet d'affirmer qu'étant donnée f densité de probabilité jointe de X_1, \dots, X_n l'application

$$x \mapsto \int_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \mathbb{R}^{n-1}} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$$

est une densité de probabilité de X_i .

THÉORÈME 0.25

Soient X_1, \dots, X_n des variables aléatoires. On note L_{X_i} la loi de probabilité de X_i , F_{X_i} la fonction de répartition de X_i , L_{X_1, \dots, X_n} la loi de probabilité jointe de X_1, \dots, X_n , F_{X_1, \dots, X_n} la fonction de répartition de X_1, \dots, X_n , f_{X_i} une densité de probabilité³ de X_i , f_{X_1, \dots, X_n} une densité de probabilité de X_1, \dots, X_n .

Alors

X_1, \dots, X_n sont indépendantes

$$\iff L_{X_1, \dots, X_n} = L_{X_1} \otimes \dots \otimes L_{X_n}$$

X_1, \dots, X_n sont indépendantes

$$\iff F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n)$$

X_1, \dots, X_n sont indépendantes $\iff f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n)$ presque partout

Démonstration Admise.

PROPOSITION 0.26 Égalité de Bienaymé

Si les X_i sont deux à deux non corrélées (par exemple, mais pas nécessairement, indépendantes), alors

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i).$$

Démonstration

$$\text{Var}\left(\sum_i X_i\right) = E\left(\left(\sum_i X_i - E\left(\sum_i X_i\right)\right)^2\right) = \sum_{(i,j) \in [1,n]^2} \underbrace{E((X_i - E(X_i)) \cdot (X_j - E(X_j)))}_{=0 \text{ si } i \neq j}$$

$$= \sum_{i \in [1, n]} \text{Var}(X_i)$$

où le cas de nullité dans l'équation ?? découle du théorème ?. La preuve est terminée.

COROLLAIRE 0.27 Inégalité de Bienaymé-Tchébitchev

Si les $(X_i)_{i \in [1, n]}$ sont deux à deux indépendantes, pour $t > 0$,

$$P \left(\left| \sum_i X_i - E(X_i) \right| \geq t \right) \leq \frac{\sum_i \text{Var}(X_i)}{t^2}.$$

Démonstration Il suffit de combiner l'inégalité de Tchébitchev et l'égalité de Bienaymé.

1.5 Somme de variables aléatoires et transformée de Fourier

On a vu que la variance de variables aléatoires non corrélées s'additionne. On va voir maintenant que les convoluées de variables aléatoires indépendantes se convoluent.

DÉFINITION 0.17 Produit de convolution

On appelle **produit de convolution** de deux lois de probabilités indépendantes P^X et P^Y sur \mathbb{R} la mesure de probabilité $P^X * P^Y$ sur \mathbb{R} définie par

$$(P^X * P^Y)(E) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \chi_E(y) dP^Y(y - x) \right) dP^X(x).$$

Intuition On consultera aussi avec profit le chapitre ?? sur la convolution.

PROPOSITION 0.28 Propriétés fondamentales du produit de convolution

Soient X et Y sont deux variables aléatoires réelles indépendantes de lois P^X et P^Y .

- La loi de $X + Y$ est $P^X * P^Y$.
- $P^X * P^Y = P^Y * P^X$.
- Pour toute fonction mesurable f de \mathbb{R} dans \mathbb{R} ,

$$\int_{\mathbb{R}} f(x) . d(P^X * P^Y)(x) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x + y) dP^Y(y) \right) dP^X(x).$$

Démonstration

- La fonction de répartition de $X + Y$ est un outil commode pour démontrer le premier point.

$$\begin{aligned}
& P_{X,Y}(X+Y \leq c) \\
&= E_{X,Y} \chi_{]-\infty, c]}(Y+X) \\
&= E_X E_Y (\chi_{]-\infty, c]}(Y+X)) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \chi_{]-\infty, c]}(x+y) dP^Y(y) \right) dP^X(x) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \chi_{]-\infty, c]}(y) dP^Y(y-x) \right) dP^X(x) = P^X * P^Y(]-\infty, c])
\end{aligned}$$

et donc $P^X * P^Y$ est bien la distribution de probabilité de $X+Y$.

Le deuxième point découle de la commutativité de l'addition, et le troisième point est une application immédiate du théorème de transport 0.21.

PROPOSITION 0.29 Liste de propriétés du produit de convolution

On se donne X, Y et Z des variables aléatoires réelles et P^X, P^Y et P^Z leurs lois.

• Le produit de convolution de la loi P^X par une masse de Dirac située⁴ en 0 est la loi P^X elle-même.

• Le produit de convolution de P^X par une masse de Dirac située en x est la loi de $X+x$.

• Le produit de convolution est commutatif, associatif.

• Le produit de convolution est distributif, au sens suivant ; pour t dans $[0, 1]$, on a :

$$P^X * (t.P^Y + (1-t).P^Z) = t.P^X * P^Y + (1-t).P^X * P^Z.*$$

Démonstration Les trois premiers • sont immédiats, au vu de la proposition précédente.

Pour montrer le quatrième point, une façon “imagée” est de passer par l’intermédiaire d’une variable aléatoire indépendante U égale à 1 avec probabilité t et à 0 avec probabilité $1-t$, avec U indépendante de variables aléatoires X, Y et Z de lois respectivement P^X, P^Y et P^Z . La variable aléatoire $X+U.Y+(1-U)Z$ est alors égale à $U(X+Y)+(1-U)(X+Z)$, et leurs lois $P^X * (t.P^Y + (1-t).P^Z)$ et $t.P^X * P^Y + (1-t).P^X * P^Z$ sont donc égales.

DÉFINITION 0.18 Fonction caractéristique

Soit X une variable aléatoire à valeurs dans \mathbb{R}^d . On appelle **fonction caractéristique** de X la fonction $\phi^X : \mathbb{R}^d \rightarrow \mathbb{R}^d$

$$\text{définie par } \phi^X(t) = E(e^{i\langle t, X \rangle})$$

Cette quantité est toujours bien définie.

On reconnaît une transformée de Fourier.

Application 0.9 On verra une jolie application avec certaines formes du théorème central limite (voir la démonstration du théorème ??).

PROPOSITION 0.30

La fonction caractéristique a les propriétés suivantes :

- $\phi^X(t) = \int_{\mathbb{R}^d} \cos(\langle t, x \rangle) dP^X(x) + i \cdot \int_{\mathbb{R}^d} \sin(\langle t, x \rangle) dP^X(x)$
- $\phi^X(0) = 1$
- ϕ^X est à valeurs dans le disque unité fermé de \mathbb{C} (important !)
- $\phi^X = \phi^Y$ implique $P^X = P^Y$.

Démonstration

Point par point :

- Par définition.
- Par convexité du disque unité.
- Grâce à l'inégalité de Jensen (voir 0.18).
- Ce point, délicat, sera ici admis.

Quelques exemples de fonctions caractéristiques :

- Si P^X est un dirac en x , alors $\phi^X(t) = e^{i\langle t, x \rangle}$.
- Étant donné X une variable aléatoire à valeurs dans \mathbb{R}^d , M une matrice de type (d, d) , et C un vecteur dans \mathbb{R}^d , avec $Y = M.X + C$, on a

$$\begin{aligned} \phi^Y(t) &= E(e^{i\langle t, Y \rangle}) \\ &= E(e^{i\langle t, MX \rangle + i\langle t, C \rangle}) \\ &= e^{i\langle t, C \rangle} \cdot E(e^{i\langle t, MX \rangle}) \\ &= e^{i\langle t, C \rangle} \cdot E(e^{i\langle X, {}^tMt \rangle}) = e^{i\langle t, C \rangle} \cdot \phi^X({}^tMt) \end{aligned}$$

- On trouvera d'autres exemples dans la partie ??.

THÉORÈME 0.31 Formule d'inversion de Fourier

On suppose que ϕ^X , fonction caractéristique de la variable aléatoire X , est intégrable. Alors X admet une densité continue bornée f^X , et on a

$$f^X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \phi^X(t) \cdot dt.$$

Démonstration On se réfère à la partie consacrée aux séries de Fourier, où l'on trouvera d'ailleurs de nombreux résultats complémentaires.

DÉFINITION 0.19 moment d'ordre k

On appelle **moment d'ordre k** de la variable aléatoire X à valeurs dans \mathbb{R} l'espérance de X^k (quand elle existe). On appelle **moment centré d'ordre k** de la variable aléatoire X à valeurs dans \mathbb{R} l'espérance de $(X - E(X))^k$.

Le résultat suivant est donné sans preuve.

PROPOSITION 0.32

Deux variables aléatoires bornées ayant les mêmes moments à tous ordres ont même distribution de probabilité.

1.6 Probabilités conditionnelles

Cette partie sera indispensable pour bien comprendre la partie sur les martingales (??). Les démonstrations, souvent abstraites et difficiles, seront laissées de côté dans ce chapitre introductif.

DÉFINITION 0.20 espérance conditionnelle de X sachant S (resp. sachant Y)

Soit X une variable aléatoire réelle d'espérance finie, sur un triplet de probabilité (Ω, \mathcal{A}, P) et soit S une sous- σ -algèbre de \mathcal{A} (resp. Y une variable aléatoire sur (Ω, \mathcal{A}, P) qui engendre la σ -algèbre $S \subset \mathcal{A}$). On appelle **espérance conditionnelle de X sachant S (resp. sachant Y)** l'unique (presque partout⁵) variable aléatoire $E(X|S)$ (resp. $E(X|Y) = E(X|S)$) mesurable pour S et telle que

$$\forall s \in S, \int_s E(X|S) dP = \int_s X dP$$

(on peut aussi écrire $E(E(X|S)\chi_s) = E(X\chi_s)$).

Si X et Y sont des variables aléatoires à valeurs dans \mathbb{R}^x et \mathbb{R}^y respectivement et admettant des densités respectives f_x et f_y , alors (X, Y) a pour densité $f_{xy}(a, b) = f_x(a)f_y(b)$ et la **loi conditionnelle de X sachant Y** , notée $\mathcal{L}_{X|Y}$ ou $\mathcal{L}_{X|Y=y}$ est la loi de densité $a \mapsto f_{xy}(a, Y)/f_y(Y)$.

Les lois conditionnelles peuvent (on ne le fait pas ici, cf [?]) être définies de manière beaucoup plus générale.

PROPOSITION 0.33

Étant données X et Y deux variables aléatoires réelles définies sur un même espace de probabilité, x, y, z des réels, S une sous- σ -algèbre, on a presque sûrement :

$$\begin{aligned} E(xX + yY + z|S) &= xE(X|S) + yE(Y|S) + z; \\ E(XY|S) &= YE(X|S) \text{ si } XY \text{ est intégrable et } Y \text{ est mesurable pour } S; \\ E(X|S) &= E(X) \text{ si } X \text{ et } S \text{ sont indépendants.} \end{aligned}$$

Intuition Une compréhension intuitive des espérances conditionnelles apparaît en regardant Y à valeurs dans un domaine fini pouvant prendre n'importe quelle valeur de ce domaine avec une probabilité non nulle. L'espérance conditionnelle de X sachant Y est alors la variable aléatoire qui vaut $\frac{1}{P(Y=y)}E(X \times \chi_y(Y))$.

1.7 Martingales

Le lecteur est encouragé à approfondir cette très brève introduction aux martingales par d'autres lectures, comme [?] et [?].

DÉFINITION 0.21 espace filtré

On appelle **espace filtré** un quadruplet $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, P)$ avec (Ω, \mathcal{F}, P) triplet de probabilité, et $(\mathcal{F}_n)_{n \in \mathbb{N}}$ une **filtration**, c'est-à-dire une suite croissantes de σ -algèbres incluses dans \mathcal{F} .

On appelle **processus adapté à un espace filtré** ou plus simplement **processus** une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans \mathbb{R} telles que

$$\forall n \in \mathbb{N}, X_n \text{ est } \mathcal{F}_n\text{-mesurable.}$$

On appelle **temps d'arrêt** une application T de Ω dans \mathbb{N} telle que pour tout n $\{\omega; T(\omega) \leq n\}$ appartient à \mathcal{F}_n .

Étant donnés X un processus et T un temps d'arrêt, on note X^T le **processus X stoppé à l'instant T** défini par $X_n^T(\omega) = X_{\min(T(\omega), n)}(\omega)$.

On appelle **processus prévisible** (relativement à un espace filtré) une suite $(C_n)_{n > 0}$ de variables aléatoires à valeurs dans \mathbb{R} telles que pour tout $n > 0$ C_n est \mathcal{F}_{n-1} -mesurable.

On appelle **processus prévisible** associé à un temps d'arrêt le processus prévisible C tel que $C_n(\omega)$ est égal à 1 si $n \leq T(\omega)$ et égal à 0 sinon.

Un processus C est dit **borné** si il existe K tel que pour tout n et tout ω , $|C_n(\omega)|$ est majoré par K .

Intuition Les martingales sont très souvent illustrées sur des jeux, mais le champ d'application est beaucoup plus vaste comme le montrent les exemples plus bas. Par exemple, l'inégalité d'Hoeffding-Azuma (applicable pour des martingales) est abondamment utilisée pour estimer les probabilités de grande déviation pour des processus très loin des jeux. Intuitivement, l'espace filtré représente les connaissances disponibles à l'instant $n \in \mathbb{N}$; c'est-à-dire qu'une fonction est \mathcal{F}_n -mesurable à condition qu'elle puisse être connue à l'instant n . Ensuite le fait qu'un processus soit adapté, signifie simplement que la valeur de $X_n(\omega)$ est connue à l'instant n . Un processus prévisible est en fait un processus déterminé à l'avance, i.e. tel que le processus à l'instant n est connu dès l'instant $n - 1$. Un processus prévisible sera notamment usuellement une stratégie élaborée par un joueur, qui peut donc agir en fonction de ce qui a déjà eu lieu, la stratégie étant supposée déterministe. Un temps d'arrêt est en fait une façon de décider un instant, sachant que la décision d'un instant ne peut être faite qu'en fonction des événements antérieurs. Le processus prévisible associé à un temps d'arrêt est en fait une façon de jouer où l'on ne choisit pas la mise, mais pour laquelle on peut choisir le moment où le jeu s'arrête.

Application 0.10 On verra un temps d'arrêt sympathique et un processus stoppé sympathique en partie ??.

DÉFINITION 0.22 marche aléatoire sur \mathbb{Z}

Un processus adapté X est une **martingale** si, pour tout n , X_n appartient à L^1 et l'espérance conditionnelle vérifie $E(X_n | \mathcal{F}_{n-1}) = X_{n-1}$.

Un processus adapté X est une **surmartingale** si, pour tout n , X_n appartient à L^1 et l'espérance conditionnelle vérifie $E(X_n | \mathcal{F}_{n-1}) \leq X_{n-1}$.

Un processus adapté X est une **sous-martingale** si, pour tout n , X_n appartient à L^1 et l'espérance conditionnelle vérifie $E(X_n | \mathcal{F}_{n-1}) \geq X_{n-1}$.

Intuition On comprend bien ce que signifie le fait que X_n soit L^1 ; la condition sur l'espérance conditionnelle d'une martingale signifie, elle, simplement que la moyenne de X_n , toutes les infor-

mations étant connues jusqu'à l'étape $n - 1$, est égale à X_{n-1} . C'est-à-dire que si l'on fixe les $n - 1$ premières étapes (on regarde la distribution conditionnelle, ces $n - 1$ valeurs étant fixées), la n -ième est centrée sur sa valeur à l'étape $n - 1$.

Intuition En voyant X_n comme le gain à un jeu jusqu'à l'instant n inclus, une surmartingale est un jeu où en moyenne on perd, une sous-martingale un jeu où en moyenne on gagne.

PROPOSITION 0.34

Si X est une surmartingale, $-X$ est une sous-martingale.

X est une martingale si et seulement si X est à la fois une surmartingale et une sous-martingale.

Exemple 0.11

On aura souvent comme filtration $\mathcal{F}_n = \sigma(W_0, \dots, W_n)$ (σ -algèbre engendrée par W_0, \dots, W_n), et $X_n = f_n(W_0, \dots, W_n)$ avec f_n mesurable de \mathbb{R}^n dans \mathbb{R} comme processus adapté.

Exemple 0.12

Soit $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires indépendantes L^1 d'espérance nulle. On définit $S_n = \sum_{i \in [0, n]} X_i$. La filtration choisie est définie ainsi : \mathcal{F}_n est la σ -algèbre engendrée par (X_0, \dots, X_n) . Alors S_n est une martingale.

Avec X_i de loi $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, variable aléatoire à valeurs dans $\{-1, 1\}$ (équirépartie sur ces deux valeurs), la martingale définie ci-dessus est une **marche aléatoire sur \mathbb{Z}** .

On peut aussi prendre des variables aléatoires Y_i positives, telles que $E \ln(Y_i) = 0$, indépendantes, et définir π_n le produit des Y_i pour $i \leq n$. La filtration se définit comme dans le cas ci-dessus. En posant $X_i = \ln Y_i$ on est ramené au cas précédent avec $S_n = \ln \pi_n$ et on voit facilement que π_n est aussi une martingale.

THÉORÈME 0.35 On ne peut pas gagner si on a un porte-monnaie fini

• Si C est un processus prévisible borné et positif et si X est une surmartingale, une sous-martingale ou une martingale (respectivement), alors $(C \bullet X)$ défini par

$$(C \bullet X)_n = \sum_{i=1}^n c_i(X_i - X_{i-1})$$

est une surmartingale, une sous-martingale ou une martingale.

• Si C est un processus prévisible borné et X une martingale, alors $C \bullet X$ est une martingale.

Démonstration En utilisant les propriétés de l'espérance conditionnelle,

$$\begin{aligned} & E((C \bullet X)_n - (C \bullet X)_{n-1} | \mathcal{F}_{n-1}) \\ &= C_n E(X_n - X_{n-1} | \mathcal{F}_{n-1}) \\ &= C_n (E(X_n | \mathcal{F}_{n-1}) - E(X_{n-1} | \mathcal{F}_{n-1})) = C_n (E(X_n | \mathcal{F}_{n-1}) - X_{n-1}), \end{aligned}$$

d'où les résultats en appliquant les définitions des martingales, des surmartingales, des sousmartingales.

COROLLAIRE 0.36

Si X est une surmartingale et T un temps d'arrêt, alors le processus stoppé X^T est une surmartingale et $E(X_T) \leq E(X_0)$. Si X est une martingale et T un temps d'arrêt, alors le processus stoppé X^T est une martingale et $E(X_T) = E(X_0)$.

Le résultat suivant provient de [?] :

THÉORÈME 0.37 Théorème d'arrêt éventuel de Doob

Soit T un temps d'arrêt et X une surmartingale, alors si l'une des conditions suivantes est vérifiée :

- $\exists N; \forall \omega T(\omega) < N$
 - $\exists K; \forall (\omega, n) |X_n(\omega)| < K$ et pour presque tout ω T est fini.
 - $E(T) < \infty$ et $\exists K$ tel que $\forall (n, \omega) |X_n(\omega) - X_{n-1}(\omega)| \leq K$
- alors on peut conclure que $E(X_T) \leq E(X_0)$.

Démonstration C'est une application directe des résultats ci-dessus, en utilisant le théorème de convergence dominée de Lebesgue ?? dans le troisième cas.

1.8 Processus stochastique. Processus de Markov

DÉFINITION 0.23

On dit qu'une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires à valeurs dans un ensemble E au plus dénombrable muni de la σ -algèbre $P(E)$ est une **chaîne de Markov** dans E (appelé **espace des états**) si pour tout (i_0, \dots, i_n) suite finie d'éléments de E telle que $P(X_0 = i_0 \wedge X_1 = i_1 \wedge \dots \wedge X_{n-1} = i_{n-1}) > 0$, $P(X_n = i_n | X_0 = i_0 \wedge X_1 = i_1 \wedge \dots \wedge X_{n-1} = i_{n-1}) = P(X_n = i_n | X_{n-1} = i_{n-1})$.

La chaîne de Markov est dite **homogène** si pour tout n , la probabilité $P(X_{n+1} = j | X_n = i)$ est indépendante de n (sous condition que $P(X_n = i) > 0$).

On appelle **matrice stochastique** une application M de E^2 dans $[0, 1]$ telle que pour tout i ,

$$\sum_{j=0}^n M_{i,j} = 1.$$

Soit une chaîne de Markov homogène telle que pour tout i dans E il existe n tel que $P(X_n = i) > 0$ ⁶. On appelle **matrice stochastique**, dite aussi **matrice de transition**, associée à cette chaîne de Markov, la matrice M définie par $M_{i,j} = P(X_{n+1} = j | X_n = i)$.

Intuition Cela signifie simplement que l'état à l'instant n (i.e. X_n) ne dépend que de l'état à l'instant $n - 1$ (i.e. de X_{n-1}) et non des états aux instants antérieurs. La chaîne est homogène si les changements d'états ne dépendent que de l'état, et pas de la date. Dans beaucoup de modélisations, la chaîne est homogène.

Les marches aléatoires, définies dans la partie ??, sont des exemples de chaînes de Markov en plus d'être des martingales.

Remarquons l'égalité de Chapman-Kolmogorov :

$$P(X_{m+n} = j | X_0 = i) = \sum_{k \in E} P(X_{m+n} = j | X_n = k) P(X_n = k | X_0 = i).$$

Notons que les produits de matrices stochastiques, définis comme généralisation du produit usuel de matrice par $MN = P$ avec $P_{i,j} = \sum_{k \in E} M_{i,k} N_{k,j}$, sont bien définis et sont encore des matrices stochastiques. On remarque aussi que :

PROPOSITION 0.38

Si X est un processus de Markov de matrice de transition M^T , alors

- $P(X_0 = i_0 \wedge X_1 = i_1 \wedge X_n = i_n) = P(X_0 = i_0) M_{i_0, i_1} M_{i_1, i_2} \dots M_{i_{n-1}, i_n}$.
- $P(X_n = i) = \sum_{j \in E} (P^n)_{j,i} P(X_0 = j)$.

1.9 Zoologie des lois de probabilité

On présente ci-dessous un échantillon des lois de probabilités les plus usitées. La loi gaussienne, mono- ou multidimensionnelle, a bien sûr une importance particulière de par son statut de limite du théorème de la limite centrale. La loi de Bernoulli, en tant que loi non-triviale la plus simple qu'on puisse imaginer, a une grande importance. La binomiale généralise la loi de Bernoulli : c'est la somme d'un nombre fini de variables de Bernoulli indépendantes et identiquement distribuées. La multinomiale est une généralisation de la loi binomiale ; au lieu d'avoir une somme de variables de Bernoulli i.i.d., on a une somme de variables aléatoires i.i.d. sur $\{(1, 0, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)\}$. Intuitivement, là où la binomiale compte par exemple le nombre de victoires dans un jeu aléatoire simple, la multinomiale compte le nombre de fois où on a obtenu le résultat x pour chaque résultat x possible.

1.9.1 Lois normales ou gaussiennes

Cas unidimensionnel • Paramètres : $\mathcal{N}(m, \sigma^2)$ a pour paramètres m (sa moyenne) et σ^2 (sa variance). Le cas particulier $m = 0$ et $\sigma^2 = 1$ est appelé **loi gaussienne centrée réduite**.

- À valeurs dans \mathbb{R} .
- Densité :

$$x \mapsto \frac{\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$

- Fonction caractéristique : $\phi(t) = \exp(itm - \frac{1}{2}\sigma^2 t^2)$.
- Propriétés : voir (bien sûr!) le théorème central limite. Par ailleurs, si X_n a loi $B(n, p)$, alors $(X_n - np)/(\sqrt{np(1-p)})$ tend vers $\mathcal{N}(0, 1)$.

Cas multidimensionnel • Paramètres :

$N(m, \Gamma)$ où $m \in \mathbb{R}^d$ et Γ est une matrice symétrique semi-définie positive réelle de type (d, d) .

- À valeurs dans \mathbb{R}^d .
- Densité : $x \mapsto \frac{1}{(2\pi)^{d/2} \sqrt{\det \Gamma}} \exp(-\frac{1}{2} {}^t(x-m)\Gamma^{-1}(x-m))$.
- Espérance : m
- Covariance : Γ
- Fonction caractéristique : $\phi(t) = \exp(i {}^t m t - \frac{1}{2} {}^t t \Gamma t)$
- Propriétés : la somme de deux variables aléatoires gaussiennes indépendantes de lois $N(m_a, \Gamma_a)$ et $N(m_b, \Gamma_b)$ est une gaussienne de loi $N(m_a + m_b, \Gamma_a + \Gamma_b)$. Démonstration avec X et Y deux

variables aléatoires, supposées centrées sans perte de généralité et avec la notation x_i pour la i -ième coordonnée de x :

$$\begin{aligned}
 & E((X+Y)_i - E(X+Y)_i)((X+Y)_j - E(X+Y)_j) \\
 = & E(X_i + Y_i - EX_i - EY_i)(X_j + Y_j - EX_j - EY_j) \\
 = & E(X_i + Y_i)(X_j + Y_j) \\
 = & EX_iX_j + EY_iY_j \text{ car } EX_iY_j = EY_iX_j = 0 \text{ par indépendance de } X \text{ et } Y \\
 = & Cov(X)_{i,j} + Cov(Y)_{i,j}
 \end{aligned}$$

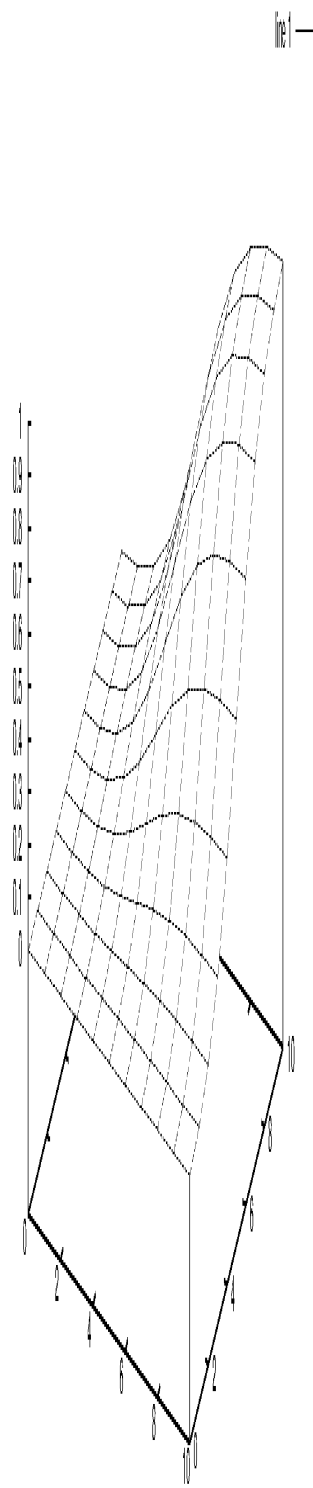
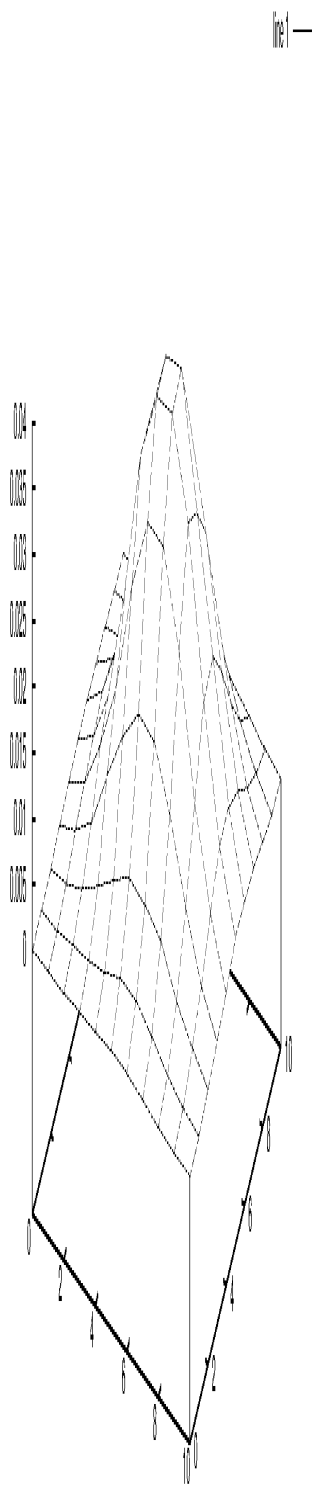
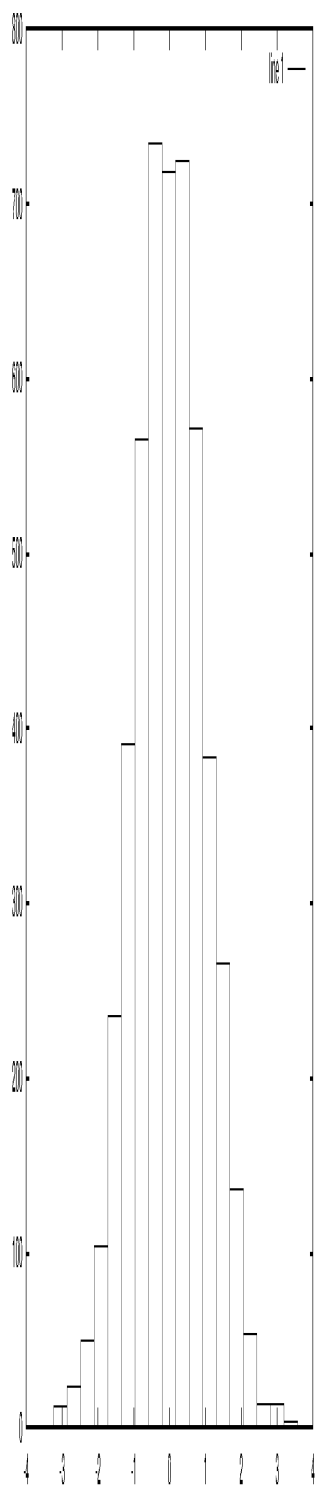
La commande Matlab « hist » permet directement d'afficher des histogrammes en dimension 1. En dimension 2 on peut utiliser par exemple les lignes fournies ci-dessous. Le dernier paragraphe de commandes fournit une estimation empirique de la fonction de répartition.

Exemple Matlab

```

X = ceil(randn(2, 50000)) + 6;
for i = 1 : 11, for j = 1 : 11;
M(i, j) = mean(prod(X(:, :) == [i; j] * ones(1, 50000)));
endfor; endfor;
mesh(M);

```



1.9.2 Loi de Bernoulli

- Paramètre : $B(p)$ a pour paramètre $p \in [0, 1]$
- À valeurs dans $\{0, 1\}$
- Loi : $P(X = 1) = p$ et $P(X = 0) = 1 - p$
- Espérance : p
- Variance : $p(1 - p)$
- Fonction caractéristique : $\phi(t) = 1 - p + pe^{it}$
- Intuition : pile ou face si $p = \frac{1}{2}$, pile ou face « biaisé » sinon.

1.9.3 Lois binomiales et multinomiales

Loi binomiale • Paramètres : $B(n, p)$ a pour paramètres n dans \mathbb{N} et $p \in [0, 1]$

- À valeurs dans $\{0, 1, 2, \dots, n\}$
- Loi : $P(X = k) = C_n^k p^k (1 - p)^{n-k}$ si $k \in [0, n]$ et $P(X = k) = 0$ sinon
- Espérance : $n.p$
- Variance : $n.p.(1 - p)$
- Fonction caractéristique : $\phi(t) = (1 - p + p.e^{it})^n$
- Intuition : somme de n variables aléatoires de Bernoulli indépendantes de même paramètre p .
- Signe particulier : la somme de deux variables aléatoires de lois binomiales $B(n_1, p)$ et $B(n_2, p)$ est une variable aléatoire de loi $B(n_1 + n_2, p)$ (les deux variables aléatoires binomiales en question étant supposées indépendantes). On peut de la même manière sommer un nombre quelconque de lois binomiales (conformément à l'intuition ci-dessus d'ailleurs).
- Cas particulier : $B(1, p) = B(p)$, loi de Bernoulli.
- Cas limite : Si $\lim_{n \rightarrow \infty} n.p_n = \lambda$, alors $B(n, p_n)$ converge en loi vers une variable aléatoire de loi de Poisson $P(\lambda)$. Noter que seul le produit $n.p_n$ compte pour ce passage à la limite ; d'où l'additivité des lois de Poisson quels que soient leurs paramètres.

Loi géométrique • Paramètre : $G(p)$ a un paramètre $p \in]0, 1]$

- À valeurs dans \mathbb{N}
- Loi : $P(X = k) = p.(1 - p)^k$
- Espérance : $\frac{1-p}{p}$
- Variance : $\frac{1-p}{p^2}$
- Fonction caractéristique : $\phi(t) = (\frac{pe^{it}}{1-(1-p)e^{it}})$
- Intuition : on tire au sort jusqu'à ce que l'on gagne, sachant qu'à chaque étape on a une probabilité p de gagner. Le nombre d'échecs avant la première victoire suit une loi géométrique $G(p)$.

loi binomiale négative • Paramètres : $B^-(n, p)$ a deux paramètres $n \in \mathbb{N}^*$ et $p \in]0, 1]$

- À valeurs dans \mathbb{N}
- Loi : $P(X = k) = C_{n+k-1}^{n-1} p^n . (1 - p)^k$ pour tout $k \in \mathbb{N}$
- Espérance : $n. \frac{1-p}{p}$
- Variance : $n. \frac{1-p}{p^2}$
- Fonction caractéristique : $\phi(t) = (\frac{p}{1-(1-p)e^{it}})^n$
- Intuition : à chaque instant dans \mathbb{N} , on joue avec probabilité p de gagner ; on compte le nombre d'échecs avant d'avoir gagné n fois.

- Cas limite : $B^-(1, p) = G(p)$.

Loi multinomiale • Paramètre : $\mathcal{M}(n, p_1, p_2, \dots, p_d)$ a pour paramètres $n \in \mathbb{N}$ et $(p_1, \dots, p_d) \in [0, 1]^d$ avec $\sum_{i=1}^d p_i = 1$

- À valeurs $(n_1, \dots, n_d) \in [[0, n]]^d$, avec $\sum_{i=1}^d n_i = n$
- Loi : $P(X = (n_1, \dots, n_d)) = \frac{n!}{n_1! n_2! \dots n_d!}$ si $\sum_{i=1}^d n_i = n$ et 0 sinon
- Espérance : $(n.p_1, n.p_2, \dots, n.p_d)$
- Matrice de covariance : $M_{i,j} = -n.p_i.p_j$ si $i \neq j$ et $M_{i,i} = n.p_i.(1 - p_i)$
- Intuition : on tire au sort n fois un nombre entier entre 1 et d avec probabilités respectives p_1, \dots, p_d , et la i -ième composante de notre variable aléatoire représente le nombre de fois que l'on a tiré l'entier i .

1.9.4 Loi de Poisson

- Paramètre : $\mathcal{P}(\lambda)$ a pour paramètre $\lambda \in]0, \infty[$
- À valeurs dans \mathbb{N}
- Loi : $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$
- Espérance : λ
- Variance : λ
- Fonction caractéristique : $\phi(t) = e^{\lambda.(e^{it} - 1)}$
- Intuition : cas limite de la loi binomiale (voir partie ??). La loi de Poisson est aussi appelée loi des événements rares. Elle sert à modéliser de nombreux phénomènes sans mémoire ; X de loi de Poisson de paramètre λ vaut k avec la probabilité pour qu'un phénomène sans mémoire survienne k fois en une seconde, sachant qu'en moyenne ce phénomène survient λ fois par seconde.
- Signe particulier : la somme de deux variables aléatoires de lois de Poisson $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$ est une variable aléatoire de loi de Poisson $\mathcal{P}(\lambda + \mu)$.
- Cas limite : Si X_λ suit une loi $\mathcal{P}(\lambda)$, alors $\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$ converge en loi vers une loi normale $\mathcal{N}(0, 1)$ quand λ tend vers $+\infty$.

1.9.5 Loi hypergéométrique

- Paramètre : $H(N, n, p)$ a pour paramètres N un entier, n un entier $\leq N$, et p de la forme q/N avec $q \in \{0, 1, \dots, N\}$.
- À valeurs dans $\{0, 1, \dots, n\}$
- Loi : $P(X = k) = \frac{C_{N-p}^k \cdot C_N^{n-k}}{C_N^n}$ si k est supérieur ou égal à 0 et à $n - N.(1 - p)$ et inférieur ou égal à n et à $N.p$.
- Espérance : $n.p$ (indépendante de N !)
- Variance : $\frac{N-n}{N-1} . n.p.(1 - p)$
- Fonction caractéristique : $\phi(t) = \frac{C_{N(1-p)}^n}{C_N^n} {}_2F_1(-n, -Np; N(1-p) - n + 1; e^{it})$ où ${}_2F_1$ est une fonction hypergéométrique.
- Intuition : une urne contient N boules, dont une proportion p de boules noires. On tire n boules ; la loi hypergéométrique $H(N, n, p)$ décrit le comportement du nombre de boules noires tirées.

- Cas limite : pour tous n, p , une suite indexée par N de variables aléatoires suivant une loi hypergéométrique $H(N, n, p)$ converge en loi quand $N \rightarrow \infty$ vers une loi binomiale $B(n, p)$ (ce qui est intuitivement logique).

1.10 Loi des grands nombres

THÉORÈME 0.39 Loi des grands nombres

Soit $(X_n)_{n \in \mathbb{N}}$ une famille de variables aléatoires à valeurs réelles, indépendantes identiquement distribuées et vérifiant $E|X_1| < \infty$. Alors avec $S_n = \sum_{i=1}^n X_i$, presque sûrement, $S_n/n \rightarrow E(X_1)$.

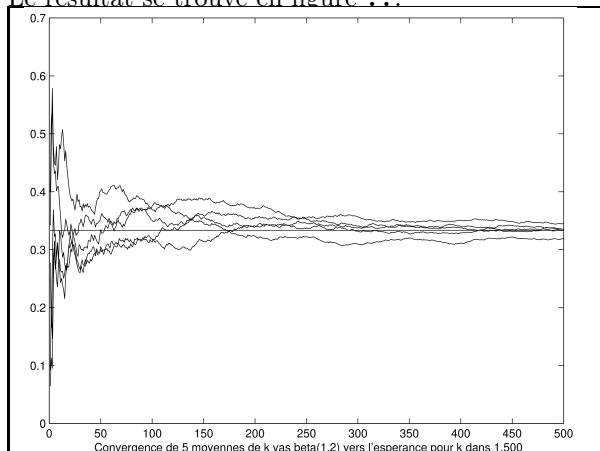
Cette version de loi des grands nombres est usuellement appelée loi forte des grands nombres par opposition à la loi faible des grands nombres, où l'on démontre (beaucoup plus facilement) la convergence en probabilité.

Notons que de nombreuses extensions, non-identiquement distribuées ou non-indépendantes existent, que l'on ne détaillera pas ici. On peut illustrer ce résultat aisément en Matlab ou Octave.

Exemple Matlab

```
x = rbeta([p, n], i, j);
m = cumsum(x')'. * (ones([1, p])' * (1./(1 : n)));
m = [m; ((i/(i + j)) * ones([1, n]))];
plot(m');
xlabel(sprintf('Convergence de %d moyennes
de k vas beta(%d,%d) vers
l\'esperance pour k dans 1,%d', p, i, j, n));
```

Le résultat se trouve en figure ??.



1.11 Théorème central limite

THÉORÈME 0.40 Théorème de la limite centrale

Soit $(X_n)_{n \in \mathbb{N}}$ une famille de variables aléatoires réelles indépendantes identiquement distribuées et de variance finie. Alors avec $S_n = \sum_{i=1}^n X_i$,

$$\frac{S_n - nE(X_1)}{\sqrt{n \operatorname{Var} X_1}}$$

converge en loi vers une variable aléatoire gaussienne centrée réduite.

Application 0.13 De nombreuses extensions, non-identiquement distribuées ou non-indépendantes existent. Une application du théorème est la construction d'intervalles de confiance (voir partie statistiques).

Démonstration On peut supposer sans perte de généralité que la variance est 1 et que la moyenne est 0. Il est ensuite suffisant de montrer que la fonction caractéristique de S_n/\sqrt{n} tend vers celle de la loi normale centrée réduite. Or, la fonction caractéristique de S_n/\sqrt{n} en t est la puissance n -ième de celle de X_i en t/\sqrt{n} (par indépendance des X_i , voir propriétés de la transformée de Fourier et de la convolution). X_i étant de carré intégrable, sa fonction caractéristique est deux fois dérivable (voir ce lemme dans [p145]BL). On peut donc écrire, avec ϕ la fonction caractéristique de X_1 , que

$$\phi(t) = 1 + \underbrace{\phi'(0)}_{=0 \text{ car } EX_1=0} t + \frac{1}{2} \underbrace{\phi''(0)}_{=-E(X_1^2)=-1} t^2 + o(t^2) = 1 - \frac{1}{2}t^2 + o(t^2)$$

grâce à la formule de Taylor-Young. On en déduit donc que, avec ψ_n la fonction caractéristique de S_n/\sqrt{n} ,

$$\psi_n(t) = (1 - t^2/2n + o(1/n))^n$$

qui tend (le voir en passant au logarithme) vers $\exp(-\frac{t^2}{2})$, qui est la fonction caractéristique de la loi centrée réduite.

On donne ici deux exemples d'illustration, l'un en matlab, l'autre en Maple.

Exemple Matlab

```
function f = tcl(i,j,n,p,nb,step,A)
N = n/step;
x = sum(rbeta([p,N,step],i,j),3);
E = step*i/(i + j);
sigma = (i*j/((i + j + 1)*(i + j)2))*step;
m = (cumsum(x')'.*(ones([1,p])*(1./(1 :N))) - E)/sqrt(sigma);
rf = pnorm(- A :2*A/nb :A,0,1);
index = cumsum(ones([1,nb])/(nb + 2));
for k = 1 :N,
    vecteur = (m( :,k))*sqrt(k);
    e = quantile(vecteur,index);
    plot(- A :2*A/nb :A,rf,e,index);
    xlabel(sprintf('Fonction de repartition de
la moyenne de %d vas beta( %d, %d)',k*step,i,j));
    if (k = N)
        sprintf('Appuyez sur une touche pour
la suite'), pause;
    end
end;
```

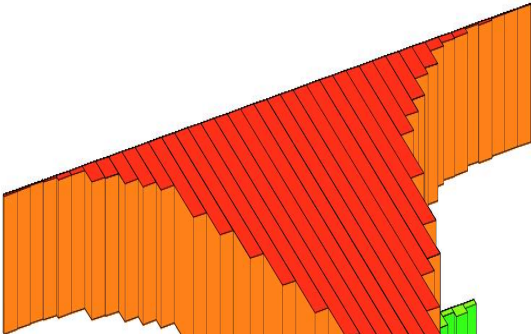
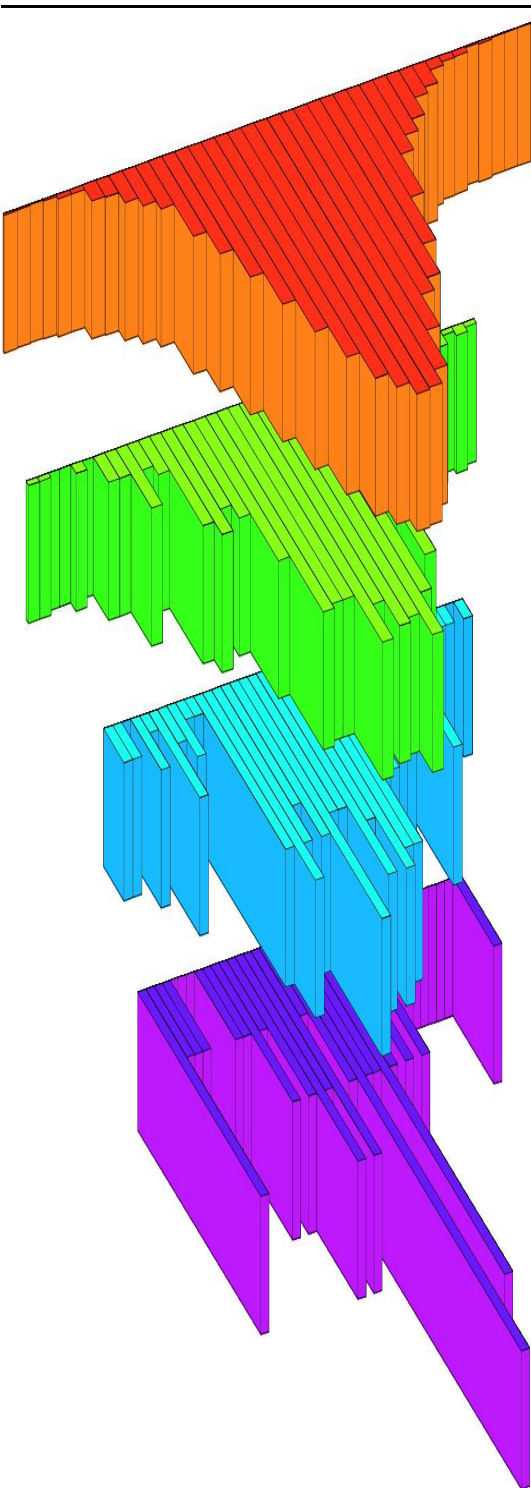
L'intérêt de cet exemple (notamment par rapport à l'illustration suivante en Maple) est le fait qu'ici on montre la convergence simple de la fonction de répartition, alors que la convergence illustrée en examinant des histogrammes est moins directement liée au théorème central limite. On ne représente pas la figure obtenue car il s'agit d'une suite, qu'on ne peut rendre sur papier sans occuper une place abusive.

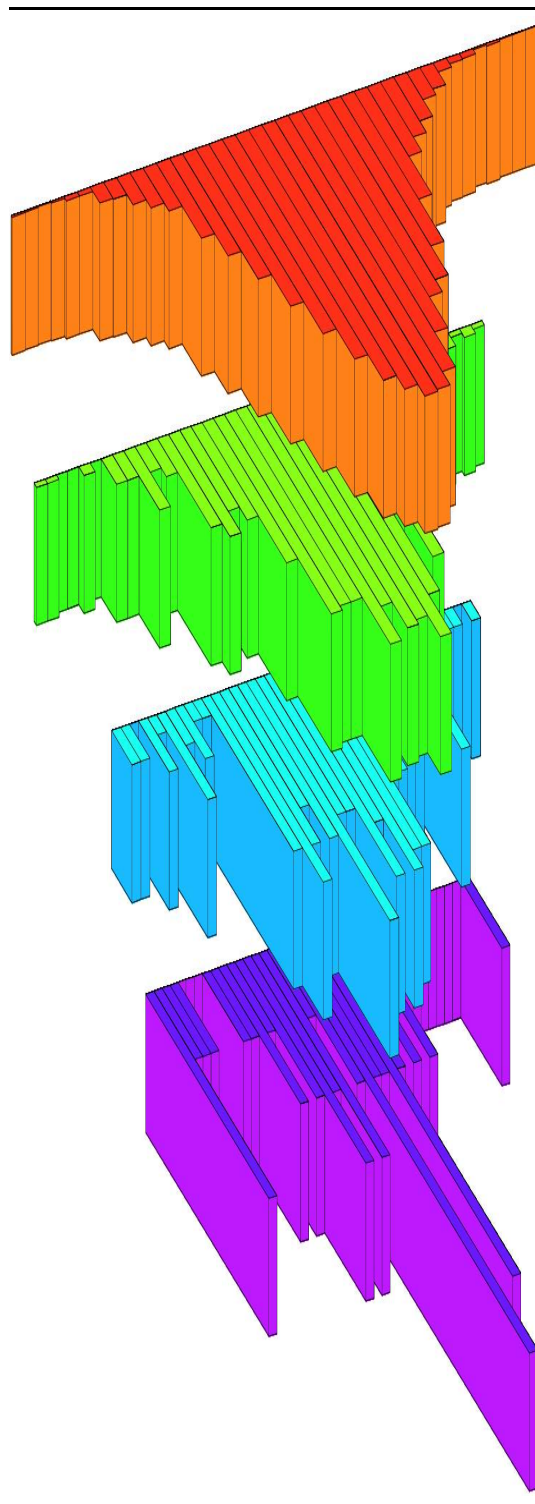
Exemple Maple

```
[anova, describe, fit, importdata, random, statevalf, statplots, transform]
unif := i -> random[uniform[0,1]](i);
unif:= random_uniform_0, 1
normale := i -> random[normald[0,1]](i);
normale := random_normald_0, 1
fit[leastsquare[[x,y,z],z=d*x*y+e*x2+f*y2+a*x+b*y+c, {a,b,c,d,e,f}]]([1,1,1,2,2,2,3,3,3],[1,2,3,1,2,3,1,2,3], [2,4,6,9,12,16,12,13,14]);

$$z = -\frac{1}{2}xy - \frac{23}{6}x^2 + \frac{1}{6}y^2 + \frac{125}{6}x + \frac{5}{2}y - \frac{160}{9}$$

histogram([normale(50)],[normale(200)],[normale(800)],[normale(6400)],numbars=40,area=1);
[boxplot, histogram, scatterplot, xscale, xshift, xyexchange, xzexchange, yscale, yshift,yzexchange,
zscale, zshift]
```





THÉORÈME 0.41 Théorème multivarié de la limite centrale

Soit $(X_n)_{n \in \mathbb{N}}$ une famille de variables aléatoires indépendantes à valeurs dans \mathbb{R}^n identiquement distribuées et de variance finie. Alors avec $S_n = \sum_{i=1}^n X_i$,

$$\frac{S_n - nE(X_1)}{\sqrt{n}}$$

converge en loi vers une variable aléatoire gaussienne centrée de matrice de variance covariance la matrice de variance covariance de X_1 .

Des extensions du théorème pour des variables aléatoires non-identiquement distribuées ou non-indépendantes existent.

1.12 Inégalité de Cramer, grandes déviations

On pourra consulter le livre [?], lecture 16, avec profit pour plus d'informations. Les grandes déviations sont indispensables pour étudier de nombreux phénomènes aléatoires. On pourra par exemple se pencher sur les applications présentées dans le livre de Lugosi et Cesa-Bianchi, « Prediction, learning and games ». On donne ici sans démonstration le résultat suivant :

THÉORÈME 0.42 Probabilités de grandes déviations

Soit $(X_n)_{n \in \mathbb{N}}$ une famille de variables aléatoires indépendantes identiquement distribuées et telles que $P(X_n > t)$ soit nul pour t assez grand. Alors avec $S_n = \sum_{i=1}^n X_i$, la probabilité $P(|S_n|/n > c)$ avec $c > 0$ décroît exponentiellement en n (i.e. est inférieur à $A \exp(-Kn)$ pour certains $A, K > 0$).

Ce résultat est initialement dû à Cramer. Le programme suivant illustre les prédictions du théorème ci-dessus.

Exemple Maple

```

function f = gd(i,j,n,p,c,astuce)
x = mean(rbeta([p,n,astuce],i,j),3);
M = (abs(cumsum(x)')*(ones([1,p])*(1./(1:n))) - i/(i+j));
m(1,:) = mean(M > c);
m(2,:) = mean(M > 2*c);
m(3,:) = mean(M > 3*c);
m(4,:) = mean(M > 4*c);
m = (log(m))/astuce;
plot(m');
title(sprintf('1/k log de la proportion des %d moyennes
de k vas beta(%d,%d) a distance > %g x 1 :4 de
l'esperance pour k dans 1,%d',p,i,j,c,n*astuce));

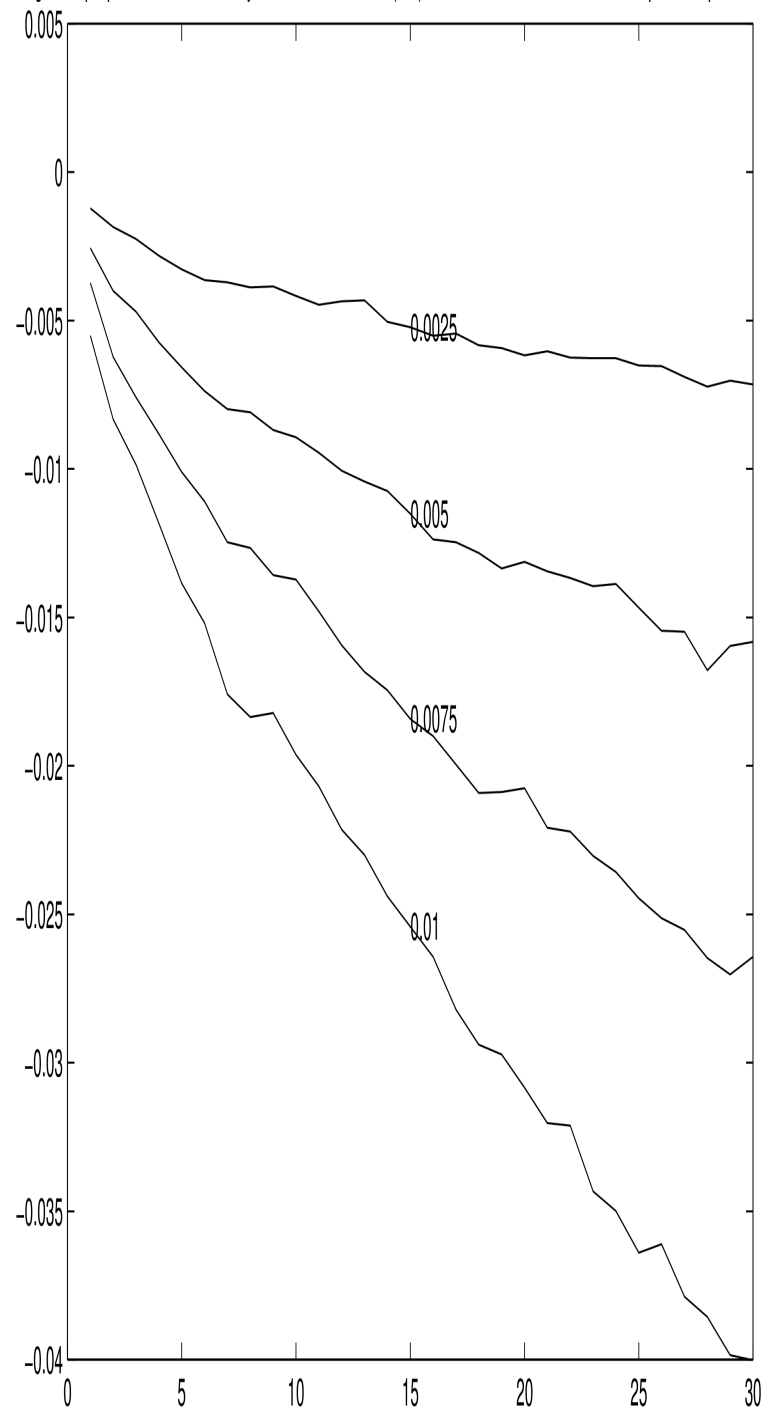
```

Exemple Matlab

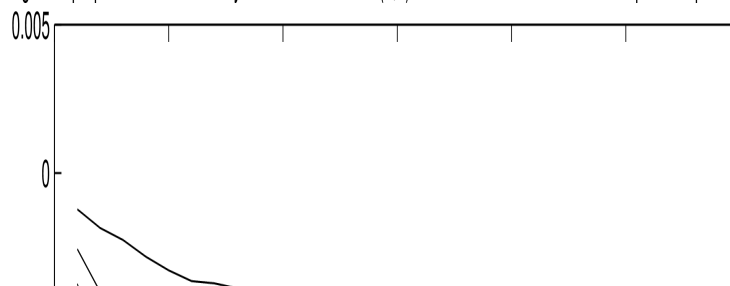
```
text( $n/2$ ,  $m(1, \text{floor}(n/2))$ , sprintf(' %g',  $c$ ));  
text( $n/2$ ,  $m(2, \text{floor}(n/2))$ , sprintf(' %g',  $2 * c$ ));  
text( $n/2$ ,  $m(3, \text{floor}(n/2))$ , sprintf(' %g',  $3 * c$ ));  
text( $n/2$ ,  $m(4, \text{floor}(n/2))$ , sprintf(' %g',  $4 * c$ ));
```

Le résultat est illustré en figure ?? ; il faut noter le fait que la courbe est bien linéaire.

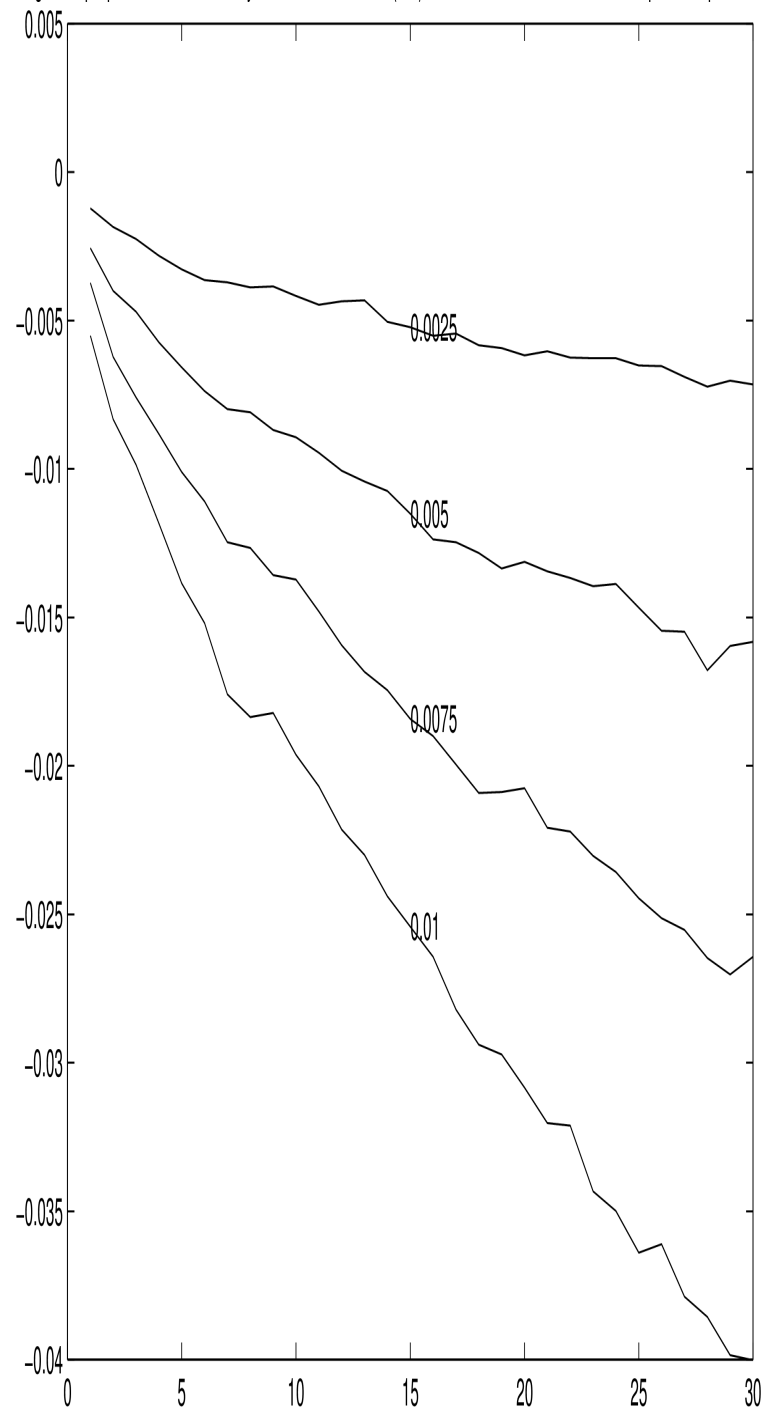
1/k log de la proportion des 2000 moyennes de k vas $\text{beta}(1,1)$ a distance $> 0.0025 \times 1:4$ de l'esperance pour k dans 1,900



1/k log de la proportion des 2000 moyennes de k vas $\text{beta}(1,1)$ a distance $> 0.0025 \times 1:4$ de l'esperance pour k dans 1,900



1/k log de la proportion des 2000 moyennes de k vas $\text{beta}(1,1)$ a distance $> 0.0025 \times 1:4$ de l'esperance pour k dans 1,900



1.13 Applications des probabilités

Les modèles probabilistes sont utilisés massivement dans l'industrie (modélisation des pannes, de la météorologie, médicale, de consommation), en biologie (processus de branchement, distances phylogénétiques) et en sciences physiques (physique statistique, physique quantique, télécommunications). Cette section propose un certain nombre d'applications ; la liste est loin d'être exhaustive. Le chapitre « statistiques » (??) est aussi un vaste champ d'applications des probabilités.

On pourra, outre les exemples ci-dessous, s'intéresser au calcul de longueur de courbe présenté dans [p33]BL.

1.13.1 Application des probabilités au calcul d'intégrales : méthode de Monte-Carlo

On se donne une fonction f intégrable de $[0, 1]$ dans \mathbb{R} . On va chercher à calculer l'intégrale I de f sur $[0, 1]$.

I est l'espérance de f , vue comme variable aléatoire. Donc par l'inégalité de Tchebitchev, on peut écrire que

$$P(|\sum_{i=1}^n f(X_i) - nI| \leq \epsilon) \leq (n \text{Var } f(X_1))/t^2$$

avec les X_i des variables aléatoires identiquement distribuées uniformes sur $[0, 1]$.

Cette méthode est en fait plutôt utilisé pour des fonctions de $[0, 1]^d$ dans \mathbb{R} , avec d grand (ou bien lorsque la fonction n'a aucune propriété de régularité), sinon des techniques d'analyse numérique sont plus adéquates (voir [?]). Diverses techniques sont utilisées pour « réduire » la variance : tirer au sort des points en quantité plus importantes là où la variance est forte, où là où l'espérance (en valeur absolue) est forte. On pourra consulter notamment le livre de P.S. Toulouse (voir [44]).

1.13.2 Calcul de surface minimale

On se donne un compact K de \mathbb{R}^2 , et ∂K son contour. On suppose donnée une fonction g définie sur ∂K . Soit E l'ensemble des applications de K dans \mathbb{R} . Chaque f appartenant à E définit une surface, l'ensemble $\{(x, y, f(x, y)); (x, y) \in K\}$.

On cherche parmi E une fonction ayant bien une surface, coïncidant avec g sur ∂K , et parmi ces fonctions la fonction définissant la surface minimale. On admet le fait que la fonction vérifiant cette propriété est une fonction ayant un laplacien, et dont le laplacien est nul ; cette fonction est unique. On va s'intéresser ici à une méthode probabiliste résolvant le problème discrétisé. On pourrait bien sûr s'attaquer à un problème plus général, mais par simplicité de notations on considèrera que $K = [0, n]^2$, et on s'intéressera seulement aux points de coordonnées entières de K . La fonction g peut-être quelconque ; on s'intéressera pour nos représentations schématiques à la fonction définie ci-dessous, `front.m` (attention, on y travaille sur les coordonnées normalisées, i.e. ramenées à $[0, 1]^2$, et non $K = [0, n]^2$) :

Exemple Matlab

```
a = abs(x - round(x));  
b = abs(y - round(y));
```

Exemple Matlab

```
if (a < b) g = 1; else g = 0; end;
```

Pour résoudre le problème, on calcule séparément les valeurs de f en les différents points de coordonnées entières de K . Considérons par exemple $(i, j) \in [0, n]^2$. On considère alors le processus de Markov $(X^{(i,j)})_n$ ayant K pour espace des états, partant de (i, j) , et effectuant une marche aléatoire simple sur K (i.e. les 4 directions sont équiprobables). On définit un temps d'arrêt T égal au nombre d'étapes avant que la marche aléatoire atteigne ∂K , i.e. une abscisse ou une ordonnée égale à 0 ou n , ce qui a une probabilité 1 d'arriver. On considère alors $f \in E$ définie par $f(i, j) = E(g((X^{(i,j)})_T))$.

Il est clair que l'application f ainsi définie vérifie bien $\Delta f = 0$. Le programme matlab correspondant est le suivant :

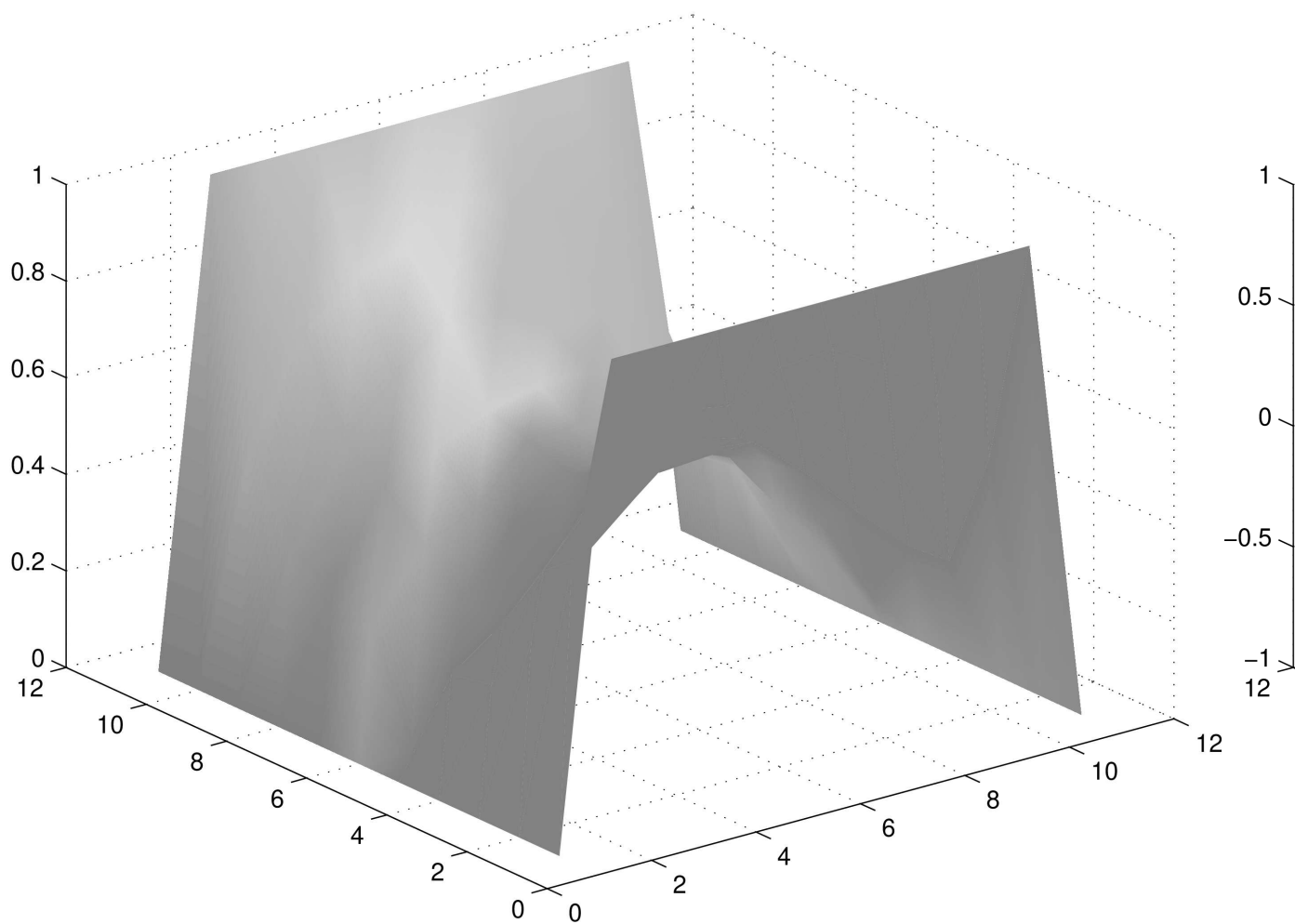
Exemple Matlab

```
function v = lapla(n,e)
u = zeros(n + 1,n + 1);
for i = 0 :n,
    for j = 0 :n,
        disp(sprintf('%g %g', (i*(n + 1) + j)*100/((n + 1)*(n + 1))))
        nb = 0;
        t = 0;
        err = [];
        while ((2*t > e) | (nb < 30))
            a = i;
            b = j;
            while((a < n)&(a > 0)&(b > 0)&(b < n))
                switch(floor(rand*4))
                    case 0
                        a = a + 1;
                    case 1
                        a = a - 1;
                    case 2
                        b = b + 1;
                    case 3
                        b = b - 1;
                end;
            end;
            a = a/n; b = b/n; nb = nb + 1; err = [err, front(a,b)];
            if (nb > 1) t = std(err)/sqrt(nb - 1); end;
        end;
        u(i + 1,j + 1) = mean(err);
    end;
end;
```

Exemple Matlab

```
surf1(u)
shadinginterp
colormapautumn
v = u;
```

On pourra regretter que les pourcentages affichés pendant le calcul ne sont pas les pourcentages du temps de calcul, mais les pourcentages du nombre de points calculés. La figure obtenue par « `lapla(10,0.05)` » est ??, à gauche. En remplaçant la fonction « `front.m` » par « `cos(4*atan((x-0.5)/(y-0.5)))` », on obtient la figure ??, à droite.



1.14 Statistique

Cette très brève introduction aux statistiques ne peut remplacer la lecture d'ouvrages de référence. Nous introduirons ici simplement un peu de terminologie utile à la vie quotidienne. On

pourra s'initier aux statistiques avec [?, ?]. Pour un cadre plus financier, on pourra se pencher sur [?]. On pourra s'orienter vers la théorie des sondages avec [?].

1.14.1 Quelques notions élémentaires

Définitions On considère ici x_i , pour $i \in [[1, n]]$, des nombres réels. Dans un grand nombre de cas, il sera intéressant de considérer le cas de n variables aléatoires, possiblement i.i.d.

On appelle **moyenne arithmétique** de n nombres réels x_1, \dots, x_n la quantité $\frac{\sum_{i=1}^n x_i}{n}$. On l'appelle aussi **moyenne** tout court lorsqu'il n'y a pas de risque de confusion, et on la note \bar{x} .

On appelle **moyenne géométrique** de n nombres réels x_1, \dots, x_n la quantité $\sqrt[n]{\prod_{i=1}^n x_i}$ lorsqu'elle est définie.

On appelle **moyenne harmonique** des x_i l'inverse de la moyenne arithmétique des inverses des x_i :

$$\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$$

On appelle **moyenne quadratique** des x_i , lorsqu'ils sont positifs, la racine carrée de la moyenne arithmétique des carrés des x_i :

$$\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

On appelle **médiane** d'une mesure finie sur un espace ordonné tout élément x tel que la mesure de $\{y; y > x\}$ est égale à la mesure de $\{y; y < x\}$.

On appelle **effectif cumulé croissant** d'une distribution sur un espace ordonné la fonction qui à x associe la mesure de $\{y; y < x\}$, et **effectif cumulé décroissant** la fonction qui à x associe la mesure de $\{y; y > x\}$. Les effectifs cumulés croissants sont aussi appelés **effectifs cumulés** tout simplement. Ces notions sont définies lorsque les mesures correspondantes sont bien finies.

On appelle **k -ième percentile** d'une distribution sur \mathbb{R} une valeur x telle que les effectifs cumulés en x représentent $k\%$ de la mesure de tout l'espace; on parle aussi de quantile $k/100$ ou de quantile à $k\%$. On définit de même des **quartiles**, des **déciles** : premier quartile = quantile à 25 %, troisième quartile = quantile à 75 %, premier décile à 10 %, etc. On appelle **interquartile** la différence entre le troisième et le premier quartile.

On appelle **mode** ou **dominante** d'une distribution toute valeur x telle que la densité de probabilité en x soit localement maximale. S'il y a plusieurs modes la distribution est dite **pluri-modale**.

On appelle **déviations** de x_i la valeur $x_i - \bar{x}$.

On appelle **écart moyen** la moyenne des $|x_i - \bar{x}|$; c'est donc $\overline{|x_i - \bar{x}|}$.

On appelle **variance** la moyenne des $(x_i - \bar{x})^2$; on la note souvent V ou σ^2 . Pour des raisons de qualité d'estimation, on utilise en fait en général

$$\frac{1}{n-1} \sum (x_i - \bar{x})^2$$

comme variance approchée et non $\frac{1}{n} \sum (x_i - \bar{x})^2$.

En effet, l'équation ?? présente l'avantage d'être, si les x_i sont des variables aléatoires i.i.d., en moyenne égale à la variance de x_1 , propriété que n'a pas l'équation ?? :

$$E_{x_1, \dots, x_n} \frac{1}{n-1} \sum (x_i - \bar{x})^2 = E_{x_1} (x_1 - E x_1)^2.$$

On dit alors que l'estimateur ?? est non-biaisé, alors que l'estimateur ?? est biaisé (il sous-estime la variance, à moins qu'elle soit nulle).

On appelle **écart type** ou **écart quadratique moyen** la racine carrée de la variance. On le note souvent σ ; $\sigma = \sqrt{V}$.

On procède à un **changement d'origine** lorsque l'on remplace les données x_i par les y_i définis par $y_i = x_i - C$, avec C une constante.

On procède à un **changement d'échelle** lorsque l'on remplace les données x_i par les y_i définis par $y_i = C.x_i$, avec C une constante.

On appelle **moment d'ordre** p des x_i par rapport à y la moyenne des $(x_i - y)^p$. Pour $p = 1$ et $y = 0$ il s'agit donc de la moyenne (arithmétique), pour $p = 2$ et $y = \bar{x}$ il s'agit de la variance.

Propriétés On note les propriétés immédiates suivantes :

- Le logarithme de la moyenne géométrique est la moyenne arithmétique des $\log(x_i)$.
- Moyenne harmonique \leq moyenne géométrique \leq moyenne arithmétique \leq moyenne quadratique.
- La moyenne arithmétique est peu sensible aux fluctuations d'échantillonnage.
- La médiane est peu sensible aux valeurs aberrantes.
- La somme des déviations est nulle.
- La variance V est aussi égale à $V = \overline{x^2} - \bar{x}^2$, avec $\overline{x^2}$ la moyenne arithmétique des x_i^2 , et \bar{x}^2 le carré de la moyenne des x_i . On le prouve facilement en développant $\sum (x_i - \bar{x})^2$.
- Multiplier les données par C multiplie la moyenne arithmétique par C , la variance par C^2 , et l'écart-type par C .
- Translater les données de C ajoute C à la moyenne arithmétique, et ne change ni la variance ni l'écart-type.

1.14.2 Applications des probabilités à l'échantillonnage

Cette partie ne se veut qu'une très brève introduction aux statistiques. Il est bien évident que dans le cadre de l'option probabilités de l'agrégation, il est indispensable de se référer à un livre plus complet. Pour une introduction concise on pourra consulter le livre « Thèmes de probabilité et statistiques » de P.S. Toulouse, Dunod 1999.

Soit X_1, \dots, X_n variables aléatoires indépendantes identiquement distribués L^1 , ou du moins telles que le théorème central limite ?? sous une forme ou une autre est vérifié. Intuitivement, les X_i sont des mesures ; par exemple, on mesure la taille de 50 français pour évaluer la taille moyenne des français. L'intérêt des probabilités va être de fournir des bornes sur l'erreur commise par une telle évaluation.

On se donne donc $m = \frac{1}{n}(X_1 + X_2 + \dots + X_m)$. On cherche $[a, b]$ tel que $M = E(X)$ soit compris dans $[a, b]$. Il faut alors noter que bien entendu, on ne peut être certain que M soit dans l'intervalle $[a, b]$, quel que soit l'intervalle que l'on donne, simplement au vu des X_i . Il est toujours possible que l'on ait été particulièrement malchanceux dans les tirages des X_i et que la moyenne soit très différente de ce que l'on suppose au vu des données. On doit donc plutôt donner α un réel (petit

de préférence) et z tel que avec probabilité $1 - \alpha$, pour toute loi de X_1 , $|m - M| \leq z$ soit vrai. a et b seront alors $m - z$ et $m + z$ respectivement.

Concrètement on procède comme suit :

- On évalue (empiriquement) l'écart type σ de X_i .
- On repère t_α tel que $P(|N| \leq t_\alpha) = 1 - \alpha$, avec N loi normale centrée réduite (espérance nulle et écart-type 1). Les valeurs de t_α sont tabulées (il s'agit simplement de la fonction de répartition de la loi normale). Le plus courant est de choisir $\alpha = 0.05$, t_α étant alors environ égal à 2.

• On détermine $a = m - t_\alpha \sigma / \sqrt{n}$ et $b = m + t_\alpha \sigma / \sqrt{n}$.

• On peut alors écrire que, au **seuil de confiance** α , M est compris entre a et b . Ceci constitue un **intervalle de confiance**. Il faut bien noter le caractère approximatif (asymptotique) de cette conclusion. On pourrait s'affranchir de cette hypothèse asymptotique, en utilisant des inégalités exactes, par exemple en utilisant l'inégalité de Hoeffding, ou de Chernoff.

Il faut bien cerner la notion de seuil de confiance. On ne se trompe, au pire cas, que dans $100 \times (1 - \alpha)\%$ des cas en utilisant ce système (à l'approximation asymptotique près).

On peut ainsi dire que la moyenne arithmétique est un estimateur de l'espérance ; que la formule ?? est un estimateur non-biaisé de la variance ; que la formule ?? est un estimateur biaisé de la variance.

On peut citer les développements suivants :

- le cas des petits échantillons ($n < 30$). Il n'est alors pas adéquat d'utiliser la loi normale comme approximation asymptotique. Il faut alors utiliser la loi de Student, sous certaines hypothèses (hypothèse de normalité des x_i , i.e. hypothèse selon laquelle les x_i sont distribués selon une distribution normale).
- le cas où l'on ne s'intéresse pas à la probabilité pour que la moyenne soit *mal* évaluée, mais à la probabilité pour que la moyenne soit *sur*-évaluée. Il suffit, pour construire un intervalle de confiance de la forme $] - \infty, b]$, de constater que

$$P(N > t) = \frac{1}{2}P(|N| > t)$$

pour toute variable aléatoire N symétrique, et en particulier donc la loi normale. On parle alors de test unilatéral (ou d'intervalle de confiance unilatéral), au lieu d'un test bilatéral.

- le cas de X_i à valeur dans $\{0, 1\}$, que l'on peut simplifier et étudier facilement sans hypothèse asymptotique ; plus généralement le cas de variables bornées peut aussi être commodément étudié sans hypothèse asymptotique (voir les inégalités de Hoeffding ou de Chernoff).
- le cas où l'on n'étudie pas la moyenne des X_i mais leur max.
- le cas de X_i non indépendants.
- le cas de X_i non identiquement distribués.
- le bootstrap, comme moyen d'évaluer des intervalles de confiance et des biais de manière très astucieuse.
- le test du χ^2 et celui de Kolmogorov-Smirnov sont deux développements indispensables des statistiques. Ils permettent de tester le fait que deux échantillons proviennent d'une même distribution, ou qu'un échantillon est bien distribué suivant une certaine distribution de probabilité.

Ces études et d'autres encore constituent la théorie des tests et font appel à des variantes parfois beaucoup plus difficiles du théorème central limite (par exemple le bootstrap utilise des extensions difficiles de ce théorème). La façon d'échantillonner, de manière plus sophistiquée, est aussi un développement important des statistiques : on peut formaliser l'intuition selon laquelle il est plus important d'avoir un grand nombre de points dans les catégories les plus variables. Outre cet aspect, consistant à biaiser l'échantillonnage pour améliorer la précision d'estimateurs, il existe aussi des méthodes dites de quasi-monte-carlo, notamment pour les espaces continus : plutôt qu'échantillonner de manière aléatoire simple⁸ et uniforme un domaine $[0, 1]^d$, pour calculer l'espérance de $f(X)$ avec X une variable aléatoire uniforme sur $[0, 1]^d$, on peut parfois choisir les $(x_i)_{i \in [[1, n]]}$ de manière « plus régulière » dans $[0, 1]^d$ qu'en les tirant au sort. Ceci est le principe de base des méthodes dites de quasi-Monte-Carlo ; on parle de suites à faible-discrépance pour ces suites de points très régulières.

Références

- [1] W. Giorgi, *Thèmes mathématiques pour l'agrégation*, Masson, 1998.
- [2] E. VanMarcke, *Random Fields : Analysis and Synthesis*, MIT Press, Cambridge MA, 1998.
- [3] A.W. van der Vaart, J.A. Wellner, *Weak Convergence and Empirical Processes, With Applications to Statistics*, Springer-Verlag, 1996.
- [4] E. Amzallag, N. Piccioli, F. Bry, *Introduction à la statistique*, Herman, 1978.
- [5] P. Barbe, M. Ledoux, *Probabilité*, Belin, 1998.
- [6] D. Williams, *Probability with martingales*, Cambridge University Press, 1991.
- [7] Y.G. Sinai *Probability theory - An introduction course*, Springer Textbook, 1992.
- [8] J.-P. Demailly, *Analyse numérique et équations différentielles*, Presses Universitaires de Grenoble, 1996.
- [9] G. Demange, J.-C. Rochet, *Méthodes mathématiques de la finance*, Economica, 2ème édition, 1997.
- [10] Y. Tillé, *Théorie des sondages. Echantillonnage et estimation en populations finies*, Dunod, 2001.

8. Échantillonnage aléatoire simple = échantillonnage i.i.d.