

Les Mathématiques pour l'Agrégation

C. Antonini
J.-F. Quint
P. Bognat
J. Bérard
E. Lebeau
E. Souche
A. Chateau
O. Teytaud

21 mai 2002

Table des matières

1	Combinatoire et dénombrements	2
1.1	Cardinaux d'ensembles finis	2
1.2	Dénombrement de fonctions	2
1.2.1	Ensemble des applications de E dans F	2
1.2.2	Ensemble des injections de E dans F	3
1.2.3	Ensemble des surjections (et bijections) de E dans F	3
1.2.4	Ensemble des applications croissantes de E vers F	3
1.3	Arrangements	4
1.4	Combinaisons	4
1.5	Quelques applications	5
1.5.1	Une formule utile	6
1.5.2	Généralisation du binôme de Newton	7
1.5.3	Familles sommables infinies	7
2	Probabilités	8
2.1	Espaces mesurés	8
2.2	Evènements	9
2.2.1	Définitions de base	9
2.2.2	Quelques mesures de probabilité	10
2.3	Variables aléatoires	10
2.3.1	Définitions : variable aléatoire, loi, fonction de répartition	10
2.3.2	Variables aléatoires indépendantes	12
2.3.3	Espérance	16
2.4	Somme de variables aléatoires et transformée de Fourier	24
2.5	Probabilités conditionnelles	27
2.6	Martingales	27
2.7	Processus stochastique. Processus de Markov	30
2.8	Zoologie des lois de probabilité	30
2.8.1	Lois normales	30
2.8.2	Loi de Bernoulli	31
2.8.3	Loi binomiale et multinomiale	31
2.8.4	Loi de Poisson	33
2.8.5	Loi hypergéométrique	33
2.9	Loi des grands nombres	34
2.10	Théorème central limite	35
2.11	Inégalité de Cramer-Chernoff, grandes déviations	36
2.12	Applications des probabilités	38
2.12.1	Liste d'exemples simples	38

2.12.2	Application des probabilités au calcul d'intégrales	38
2.12.3	ABRACADABRA : application des martingales	39
2.12.4	Application au calcul de la longueur d'une courbe	39
2.12.5	Application à l'évaluation de la perte de précision dans un algorithme	39
2.12.6	Application des probabilités à la géométrie euclidienne	40
2.12.7	Probabilité pour que le rapport de #piles/(#piles+#faces) tende vers $\frac{1}{2}$	40
2.12.8	Proportion de diviseurs de n dans $[1, i]$	40
2.12.9	Processus de branchement	42
2.12.10	Calcul de surface minimale	43
3	Statistique	48
3.1	Quelques notions élémentaires	48
3.1.1	Définitions	48
3.1.2	Propriétés	50
3.2	Applications des probabilités à l'échantillonnage	50

Chapitre 1

Combinatoire et dénombrements

On consultera avec grand profit le chapitre 1 de [10], très bien fait 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.

1.1 Cardinaux d'ensembles finis

Proposition 1 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 ; voir à 1.4.

1.2 Dénombrement de fonctions

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

1.2.1 Ensemble des applications de E dans F

Proposition 2 L'ensemble des applications de E dans F , noté F^E , a pour cardinal f^e .

1.2.2 Ensemble des injections de E dans F

Proposition 3 L'ensemble des injections de E dans F a pour cardinal $A_f^e = \frac{f!}{(f-e)!}$ si $f \geq e$; 0 sinon.

La preuve se fait facilement par récurrence. Voir 1.3 pour les premières valeurs.

1.2.3 Ensemble des surjections (et bijections) de E dans F

Proposition 4 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
$e = 5$	1	15	25	10	1

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

1.2.4 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 même cardinal que l'ensemble des applications strictement croissantes de E dans $[1, f + e - 1]$; une fois que l'on s'est convaincu de cela (en voyant que 1] 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 2] $u \mapsto v_u$ 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 facile de montrer que cet ensemble a pour cardinal C_{f+e-1}^e (les coefficients binomiaux C_n^p sont présentés plus bas).

1.3 Arrangements

Définition 5 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

1.4 Combinaisons

Définition 6 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 .

On

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

Cette formule peut aussi se déduire sans récurrence en voyant que $p!$ p -arrangements donnent la même p -combinaison 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 :

Proposition 7

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

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

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

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

$$\sum_{k=0}^n C_n^k = 2^n$$

$$\sum_{k=0}^n (-1)^k C_n^k = 0$$

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

$$1 \leq p \leq n \rightarrow p.C_n^p = n.C_{n-1}^{p-1}$$

$$\sum_{k=0}^n k.C_n^k = n.2^{n-1}$$

$$\sum_{k=0}^n (-1)^k C_n^k = 1$$

$$\sum_{k=0}^n (C_n^k)^2 = C_{2n}^n$$

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

1.5 Quelques applications

FLEMMARD dim espaces polynomes homogenes, anagrammes, tirages du loto, dates de naissance

1.5.1 Une formule utile

Proposition 8 Soit n un entier naturel. On a alors

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



Ce résultat sera 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 9 Pour tout entier naturel p , il existe un polynôme h_p de degré p tel que $g_p(x) = h_p(x)/(1-x)^{p+1}$.

Démonstration : Récurrence évidente. \square

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. \square

1.5.2 Généralisation du binôme de Newton

Proposition 10 (Généralisation de la formule de Newton) Dans un anneau commutatif, pour n non nul,

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

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

1.5.3 Familles sommables infinies

Définition 11 Une famille $(x_i)_{i \in I}$ de nombres complexes est **sommable de somme** x si pour tout ϵ il existe $J \subset I$ fini telle que, pour tout K fini, $J \subset K \subset I$ implique $|x - \sum_{i \in K} x_i| \leq \epsilon$.

Lemme 12 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 qui sont négatifs est sommable.

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 13 Toute famille sommable de nombres réels est de support dénombrable (ie 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, je me contente de démontrer ce résultat pour une famille de nombres réels positifs. Le résultat dans le cas général se démontre de manière similaire.

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. \square

Chapitre 2

Probabilités

2.1 Espaces mesurés

On trouvera en ?? les fondements de la topologie, et en ?? les fondements de la théorie de la mesure. Je ne rappelle ci-dessous que quelques définitions qui sont données dans les paragraphes que je viens de citer.

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 \mathbb{R}^+ telle que

- $\mu(\emptyset) = 0$
- 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.

La partie ?? est indispensable pour s'attaquer aux probabilités.

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

Définition 14 Une mesure est une **mesure de probabilité** si la mesure de l'espace tout entier est 1.

2.2 Evènements

2.2.1 Définitions de base

Définition 15 (Définitions de base) On appelle **triplet de probabilité** un triplet (Ω, \mathcal{F}, P) , avec P une mesure de probabilité sur (Ω, \mathcal{F}) .

Ω est appelé **l'univers**.

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

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

Proposition 16 Si chaque $(F_n)_{n \in \mathbb{N}}$ est de mesure 1, c'est à dire que chaque évènement F_n est réalisé *presque sûrement*^a, alors $\bigcap_{n \in \mathbb{N}} F_n$ se réalise avec une probabilité 1. Si E_n est une suite d'évènements tels que $\sum_i P(E_i) < +\infty$, alors $P(\limsup E_n) = 0$. Ce résultat est connu sous le nom de **premier lemme de Borel-Cantelli**.

^aNotez que **presque sûrement** est l'analogie, pour une mesure de probabilité, de **presque partout**, en théorie de la mesure.

On notera une nouvelle façon de voir \limsup des E_n , avec les E_n des évènements ; en l'écrivant $\bigcap_n \bigcup_{k \leq n} E_k$, on voit maintenant cette limite comme l'évènement qui arrive "infiniment souvent" ; c'est l'ensemble de ω qui appartiennent à une infinité de E_n . De même on peut voir différemment \liminf des E_n , avec les E_n des évènements ; en l'écrivant $\bigcup_n \bigcap_{k \leq n} E_k$, on voit $\liminf E_n$ comme l'ensemble des 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.

2.2.2 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 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 à $[0, 1]^n$ telle que la mesure de l'espace 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. Dans le cas d'une distribution uniforme, on choisit bien sûr la mesure de Lebesgue.

2.3 Variables aléatoires

2.3.1 Définitions : variable aléatoire, loi, fonction de répartition

Définition 17 (Variable aléatoire) Une **variable aléatoire** est simplement une fonction mesurable d'un univers vers \mathbb{R} (muni de sa tribu borélienne pour la topologie usuelle).

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 l'espace de départ 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 un tirage aléatoire donne 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 à définir plus tard).

Définition 18 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 la probabilité pour que X soit dans A .

Proposition 19 \mathcal{L}_X est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B})$.

Démonstration : Pas dur !□

Définition 20 Soit $F_X(t) = \mathcal{L}_X(]-\infty, t])$; alors F_X est appelée **fonction de répartition** de X .

Proposition 21 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 $F_X(t) = \mathcal{L}_X(]-\infty, t]) = P(\{\omega \in \Omega / X(\omega) \leq t\})$.□

Proposition 22 (Quelques propriétés des distributions) Soit F 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 trop 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 est égale à la mesure de l'intersection).

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 FLEMMARD □

2.3.2 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 sondée" 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.

Définition 23 (σ -algèbres indépendantes) Soit S un 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 $(s_i)_{i \in J} \in \prod_{i \in J} S_i$ avec J partie finie de I , on a $P(\cap_{i \in J} s_i) = \prod_{i \in J} P(s_i)$.
 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** une suite de variables aléatoires indépendantes et ayant la même fonction de répartition.

Rappel de vocabulaire :

- Un triplet de probabilité est un triplet (Ω, \mathcal{F}, P) , avec \mathcal{F} une tribu sur X , et P une mesure de probabilité sur (Ω, \mathcal{F}) .
- Ω est appelé univers.
- Un événement est une partie mesurable de Ω , c'est à dire un élément de \mathcal{F} .
- Une variable aléatoire sur Ω est une application mesurable de Ω dans \mathbb{R} (muni des boréliens usuels).
- La loi de probabilité d'une variable aléatoire X est l'application \mathcal{L}_X qui à un borélien B inclus dans \mathbb{R} associe $P(X^{-1}(B))$.
- La fonction de répartition d'une variable aléatoire X est l'application qui à un réel t associe $P(\{\omega \in \Omega / X(\omega) \leq t\})$.

Proposition 24 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 (cas de E_i remplacé par l'ensemble vide ou Ω trivial). \square

Définition 25 Deux Π -système 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 26 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 trivial on se préoccupe de l'autre sens.

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

$$m_1(E) = P(E \cap p_1)$$

et

$$m_2(E) = P(E) \cdot P(p_1)$$

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

Fixons maintenant p_2 dans P_2 .

On définit maintenant les deux mesures m_3 et m_4 sur S_1 par

$$m_3(E) = P(E \cap p_2)$$

et

$$m_4(E) = P(E) \cdot P(p_2)$$

Elles coïncident sur P_1 et donnent une mesure finie à X ; donc par le lemme ?? elles sont égales.

On a donc montré le résultat souhaité 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 quelconque, par le lemme ??.
- p_1 quelconque et p_2 quelconque, en fixant p_2 et en utilisant le lemme ??.

Le résultat est donc prouvé. \square

Lemme 27 (Second lemme de Borel-Cantelli) Si $(E_n)_{n \in \mathbb{N}}$ est une suite d'évènements indépendants, alors

$$\sum_n P(E_n) = +\infty \rightarrow P(\limsup E_n) = 1$$

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

Démonstration : Notons que

$$\begin{aligned} (\limsup E_n)^c &= (\bigcap_m \bigcup_{n \geq m} E_n)^c \\ &= \bigcup_m ((\bigcup_{n \geq m} E_n)^c) \\ &= \bigcup_m \bigcap_{n \geq m} (E_n^c) \\ &= \liminf (E_n^c) \end{aligned}$$

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

$$P(\bigcap_{p \geq n \geq m} E_n^c) = \prod_{p \geq n \geq m} (1 - p_n)$$

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

$$P(\bigcap_{n \geq m} E_n^c) = \prod_{n \geq m} (1 - p_n)$$

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

Le premier lemme de Borel-Cantelli, évoqué en partie ??, fournit un complément à ce lemme.

Exemple 28 (Application des deux lemmes de Borel-Cantelli) On définit les $(E_n)_{n \geq 1}$ comme des évènements aléatoires, indépendants, par

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

avec X variable aléatoire définie par sa fonction de répartition $F_X = e^{-x}$.

$P(E_n) = n^{-\alpha}$, 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 $\alpha > 1$.

Définition 29 (σ -algèbre asymptotique) Etant 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 évènements qui ne dépendent que du comportement à la limite.

Proposition 30 Les évènements suivants sont par exemples mesurables pour la σ -algèbre asymptotique des X_i (on les appelle **évènements asymptotiques**) :

- $\{\omega / \lim_{n \rightarrow +\infty} X_i(\omega) \text{ converge}\}$
- $\{\omega / \sum_{n \rightarrow +\infty} X_i(\omega) \text{ converge}\}$
- $\{\omega / \lim_{n \rightarrow +\infty} \text{sum}_{i \in [0,1]} X_i(\omega) / n \text{ converge}\}$

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

- $\limsup_{n \rightarrow +\infty} \text{sum}_{i \in [0,1]} X_i(\omega) / n$
- $\liminf_{n \rightarrow +\infty} \text{sum}_{i \in [0,1]} X_i(\omega) / n$

Pour contre-exemples on peut citer par exemple $X_1 0$ (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 donc montrer que l'ensemble E donné est inclus dans chaque τ_n .

- Les X_i sont mesurables, donc $\liminf X_i$ et $\limsup X_i$ sont mesurables, donc $(\liminf X_i - \limsup X_i)^{-1}(\{0\})$ est une partie mesurable. Donc $E \in \tau_1$; de la même manière $E \in \tau_n$, pour tout n , donc $E \in \tau$.
- Les X_i sont mesurables, donc une somme finie de X_i est mesurables, donc

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

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

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

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

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

Théorème 31 (Loi 0 – 1 de Kolmogorov) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes, et soit τ la σ -algèbre asymptotique des X_n ; alors :

- Tout évènement 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 26, nos deux σ -algèbres sont donc indépendantes.

- Y_n et τ sont indépendantes

Il suffit de voir que $\tau \subset \tau_n$.

- 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 -algèbre engendrées sont indépendantes.

- τ étant inclus dans τ_1 , τ est indépendante de τ , 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 plus grand z tel que $P(Y \leq z) = 0$. \square

2.3.3 Espérance

▣ Définitions

Définition 32 (Espérance d'une variable aléatoire dans L^1) Etant 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 variables aléatoires intégrables positives.

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

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

avec χ_F la fonction caractéristique de F .

□ **Théorèmes et inégalités**

Théorème 33 (Théorèmes de passage à la limite en probabilités) Soit X_n une suite de variables aléatoires et X une variable aléatoire telles que

$$P(X_n \rightarrow X) = 1$$

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 $X_n(\omega)$ croit 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 $E(Y) \leq +\infty$, 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 $E(|X_n - X|) \rightarrow 0$.

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

Théorème 34 (Inégalité de Markov) Supposons X variable aléatoire, et f mesurable de \mathbb{R} dans $[0, +\infty]$, avec f croissante. Alors

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

que l'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 n'y a rien à prouver, il suffit de bien voir que f est positive. □

Corollaire 35 Avec X une variable aléatoire positive,

$$E(X) \geq c.P(X \geq c)$$

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

Corollaire 36 Pour X variable aléatoire positive, $P(X \geq z) \leq E(X)/z$.

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

Théorème 37 (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

$$P(X \in U) = 1$$

$$E(|X|) < +\infty \text{ (c'est à dire que } X \text{ est intégrable)}$$

$$E(|f(X)|) < +\infty \text{ (c'est à dire que } f \circ X \text{ est intégrable)}$$

Alors :

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

↗ Voir par exemple les propriétés des fonctions caractéristiques en probabilités, 55.

Démonstration : • Les dérivées à gauche et à droite de f , notée d^- et d^+ , existent et sont croissantes ; 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 montrerait

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

• 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 $E(X) \in G$

- 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 complète \square

□ Espaces L^p

   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 !

Proposition 38 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 $N_{p'}(X) \leq N_p(X)$.

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

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

Etant 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 37) 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 complète. \square

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

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

- Si X et Y sont des variables aléatoires de L^2 , alors la somme $X + Y$ appartient à L^2 , et

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

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

Proposition 39 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 .

On a alors

$$E(f \circ X) = \int f(x).dL_X$$

Voir la proposition 19 et la définition qui la précède pour bien cerner ce qu'est une loi de probabilité.

Démonstration : Si vous n'arrivez pas à le faire vous-mêmes, mieux vaut relire tout le chapitre. La méthode est la suivante :

- 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 40 (Mesure image) Etant donnée f une application mesurable d'un espace Ω mesuré par 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 41 (Théorème de transport) Pour toute fonction mesurable ϕ de \mathbb{R} dans \mathbb{R} ,

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

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. étagée¹ et mesurable) grâce à la linéarité de l'intégrale, puis pour une fonction positive

¹Etagée = ne prenant qu'un nombre fini de valeurs

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 $\phi \circ f$). \square

Corollaire 42 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. \square

▣ **Densité de probabilité**

Définition 43 Etant 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} $P(X^{-1}(E)) = \int_E f_X$.

Notons que bien sûr $\int_{\mathbb{R}} f_X = 1$

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

Définition 44 (Covariance et variance) Etant donnée X une variable aléatoire, on définit la **déviatio**n de X par $\tilde{X} = X - E(X)$.
Etant données X et Y des variables aléatoires dans L^2 , on définit la **covariance** de X et Y par

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

Etant 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 $\langle X|Y \rangle$.

On appelle **corrél**ation entre deux variables aléatoires X et Y de norme 2 non nulles le réel de $[0, 1]$ $cor(X, Y) = \frac{\langle \tilde{X}|\tilde{Y} \rangle}{N_2(\tilde{X}) \cdot N_2(\tilde{Y})}$. On appelle **angle** entre deux variables aléatoires X et Y de norme 2 non nulles le réel θ appartenant à $[0, \Pi]$ tel que $cos(\theta) = \frac{\langle X|Y \rangle}{N_2(X) \cdot N_2(Y)}$.

Deux variables aléatoires sont dites **non corrélées** si leur covariance est nulle. On appelle **matrice de covariance** d'une suite finie de variables aléatoires (X_1, \dots, X_d) la matrice M définie par $M_{i,j} = cov(X_i, X_j)$.

Corollaire 45 (Inégalité de Tchébitchev) Pour X variable aléatoire, $P(|X - E(X)| > \epsilon) \leq Var(X)/\epsilon^2$.

↗ Voir le théorème ?? sur les polynomes de Bernstein.

Démonstration : Il suffit d'appliquer le corollaire 36 de l'inégalité de Markov à $(X - E(X))^2$. □

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 $(X - E(X)).(Y - E(Y))$ 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.

On a $cov(X, Y) = E(X.Y) - E(X).E(Y) = \langle \tilde{X} | \tilde{Y} \rangle$ et $var(X) = E(X^2) - E(X)^2$.

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

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

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

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

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

Théorème 46 (Une propriété fondamentale des variables aléatoires indépendantes)

Soient X et Y des variables aléatoires indépendantes appartenant à L^1 . Alors $X.Y$ est L^1 et

$$E(X.Y) = E(X).E(Y)$$

Soient X et Y des variables aléatoires indépendantes appartenant à L^2 . Alors :

$$cov(X, Y) = 0$$

$$var(X + Y) = var(X) + var(y)$$

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 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. □

⚠ Notez bien qu'il n'y a PAS d'erreur dans l'énoncé, 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 47 Etant 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 probabilité** de X_1, \dots, X_n une application f de \mathbb{R}^n dans \mathbb{R} telle que pour tout borélien E de \mathbb{R}^n on ait $L_{X_1, \dots, X_n}(E) = \int_E f$.

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 48 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 probabilité de X_i , f_{X_1, \dots, X_n} une densité de probabilité de X_1, \dots, X_n .

Alors

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

$$X_1, \dots, X_n \text{ sont indépendantes} \iff$$

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

$$X_1, \dots, X_n \text{ sont indépendantes} \iff$$

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

Démonstration : Admise.□

Proposition 49 (Égalité de Bienaymé) Si les X_i sont deux à deux non corrélées (par exemples indépendantes), alors

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

Démonstration :

$$\begin{aligned} & \text{Var}\left(\sum_i X_i\right) \\ &= E\left(\left(\sum X_i - E\left(\sum X_i\right)\right)^2\right) \\ &= E\left(\left(\sum_i X_i - E(X_i)\right)^2\right) \\ &= \sum_{(i,j) \in [1,n]^2} E\left((X_i - E(X_i)) \cdot (X_j - E(X_j))\right) \\ &= \sum_{i \in [1,n]} \text{Var}(X_i) \end{aligned}$$

La preuve est complète...□

Corollaire 50 (Inégalité de Bienaymé-Tchébitchev) Si les $(X_i)_{i \in [1,n]}$ sont deux à deux indépendantes, pour $t > 0$,

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

↗ On peut par exemple voir 2.12.5.

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

2.4 Somme de variables aléatoires et transformée de Fourier

Définition 51 (Produit de convolution) On appelle **produit de convolution** de deux lois de probabilités indépendantes P^X et P^Y sur \mathbb{R} la loi $P^X * P^Y$ sur \mathbb{R} définie par

$$(P^X * P^Y)(E) = \int_E \left(\int_E dP^Y(x-y) \right) dP^X(x)$$

Proposition 52 (Propriétés fondamentales du produit de convolution) Si 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 :

$$P^{X+Y}(E) = P^{X \times Y}(\text{som}(X, Y) \in E)$$

avec $\text{som} : (x, y) \mapsto x + y$

$$\begin{aligned} &= \int_{\Omega} \int_{\Omega} \chi_E(Y(\omega_1) + X(\omega_2)) d\omega_2 d\omega_1 \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_E(Y(\omega_1) + X(\omega_2)) dP^X dP^Y \end{aligned}$$

Et le premier point en découle ; le deuxième point découle de la commutativité de l'addition, et le troisième point est une application immédiate du théorème de transport. \square

Proposition 53 (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 située^a 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, en un sens bien précis ; pour t dans $[0, 1]$, on a :

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

^aUne masse de Dirac en X est une mesure δ_x telle que $\delta_x(E) = 1$ si $x \in E$ et $\delta_x(E) = 0$ sinon.

Démonstration : Les trois premiers \bullet sont évidents, au vu de la proposition précédente. Le quatrième vient simplement du fait que $t.P^Y + (1-t).P^Z$ est bien une loi

de probabilité.□

Définition 54 (Fonction caractéristique) Soit X une variable aléatoire à valeurs dans \mathbb{R}^d . On appelle **fonction caractéristique de X** la fonction

$$\phi^X : \mathbb{R}^d \rightarrow \mathbb{C}$$

définie par

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

Les adeptes auront reconnu une transformée de Fourier.

Proposition 55 • $\phi^X(t) = \int_{\mathbb{R}^d} \cos(\langle t, x \rangle) dP^X(x) + i \int_{\mathbb{R}^d} \sin(\langle t, x \rangle) dP^X(x)$

- $\phi^X(0) = 1$
- ϕ^X est à valeurs dans le disque unité fermé de \mathbb{C}
- $\phi^X = \phi^Y$ implique $P^X = P^Y$.

Démonstration :

Point par point :

- Par définition !
- Clair !
- Grâce à l'inégalité de Jensen (voir 37).
- Ce point, délicat, sera ici admis.□

Quelques exemples de fonctions caractéristiques :

- Si P^X est un dirac en x , alors $\phi^X(t) = e^{i\langle t, x \rangle}$.
- Etant 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 2.8.

Théorème 56 (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 57 On appelle **moment d'ordre k** de la variable aléatoire X à valeurs dans \mathbb{R} l'espérance de X^k . 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 58 Deux variables aléatoires bornées ayant les mêmes moments à tous ordres sont égales.

2.5 Probabilités conditionnelles

Cette partie sera indispensable pour bien comprendre la partie sur les martingales (2.6). Les démonstrations, souvent abstraites et difficiles, seront laissés de côté. FLEM-MARD

2.6 Martingales

Définition 59 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** 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 **processus prévisible** (relativement à un espace filtré) une suite $(X_n)_{n > 0}$ de variables aléatoires à valeurs dans \mathbb{R} telles que pour tout $n > 0$ C_n est \mathcal{F}_{n-1} -mesurable.

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

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.

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

Etant donné un processus prévisible C et une martingale X , on note $(C \bullet X)_n = \sum_{i=1}^n C_i(X_i - X_{i-1})$ pour $n > 1$.

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 .



En fait l'espace filtré représente les connaissances disponibles à l'instant $n \in \mathbb{N}$, dans un espace à temps discret ; 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, ie 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. $(C \bullet X)_n$ représente le total des gains à l'instant n , C_n représentant la mise, et $X_n - X_{n-1}$ le gain avant multiplication par la mise. 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.

➤ On verra un temps d'arrêt sympathique et un processus stoppé sympathiques en partie 2.12.10.

Définition 60 Un processus adapté X est une **martingale** si pour tout n X_n est L^1 ET si pour tout n l'espérance conditionnelle $E(X_n | \mathcal{F}_{n-1})$ est égale à X_{n-1} .
 Un processus adapté X est une **surmartingale** si pour tout n X_n est L^1 ET si pour tout n l'espérance conditionnelle $E(X_n | \mathcal{F}_{n-1})$ est \leq à X_{n-1} .
 Un processus adapté X est une **sous-martingale** si pour tout n X_n est L^1 ET si pour tout n l'espérance conditionnelle $E(X_n | \mathcal{F}_{n-1})$ est \geq à X_{n-1} .



On comprend bien ce que signifie le fait que X_n soit L^1 ; la condition sur l'espérance conditionnelle, signifie, elle, simplement que la moyenne de X_n , toutes les informations 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, la n -ième est centrée (a sa moyenne) sur l'étape $n - 1$.



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 61 Si X est une surmartingale, $-X$ est une sous-martingale. X est une surmartingale si et seulement si X est une surmartingale et une sous-martingale.

Exemple 62 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 63 Soit $(X_n)_{n \in \mathbb{N}}$ des variables aléatoires indépendantes L^1 d'espérance nulle.

On définit $S_n = \sum_{i \in [0, n]} Y_i$. La filtration choisie est définie par : \mathcal{F}_n est la σ -algèbre engendrée par (X_0, \dots, X_n) . Alors S_n est une martingale.

Avec $X_i = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, variable aléatoire à valeurs dans $\{-1, 1\}$ (équirépartie sur ces deux valeurs), on a une **marche aléatoire sur \mathbb{Z}** .

On peut aussi prendre des variables aléatoires X_i positives, d'espérance 1, indépendantes, pour X , et définir Π_n le produit des X_i pour $i \leq n$. La filtration se définit comme dans le cas ci-dessus.

Théorème 64 (On peut pas gagner si on a un porte-monnaie fini) • Si C est un processus prévisible et borné et positif et si X est une surmartingale, une sous-martingale, une martingale (respectivement), alors $(C \bullet X)$ est une surmartingale, une sous-martingale, une martingale.
 • Si C est un processus prévisible et borné et X une martingale, alors $C \bullet X$ est une martingale.

Démonstration : En utilisant les propriétés de l'espérance conditionnelle,

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

$$C_n (E(X_n | \mathcal{F}_{n-1}) - E(X_{n-1} | \mathcal{F}_{n-1})) = C_n (E(X_n | \mathcal{F}_{n-1}) - X_{n-1}),$$

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

Corollaire 65 Si X est une surmartingale, et T un temps d'arrêt, alors le processus stoppé X^T est une surmartingale et $E(X_{T^n}) \leq E(X_0)$. Si X est une martingale, et T un temps d'arrêt, alors le processus stoppé X^T est une martingale et $E(X_{T^n}) = E(X_0)$.

Le résultat suivant provient de [21] :

Théorème 66 (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 $\exists K / \forall (n, \omega) |X_n(\omega) - X_{n-1}(\omega)| \leq K$

on peut conclure que $E(X_T) \leq E(X_0)$

Démonstration : Application facile des résultats ci-dessus, en utilisant la convergence dominée de Lebesgue ?? dans le troisième cas. \square

2.7 Processus stochastique. Processus de Markov

Définition 67 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 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 $P(X_{n+1} = j | X_n = i)$ est indépendant de n tel que $P(X_n = i) > 0$.

On appelle **matrice stochastique** une application M de E^2 dans $[0, 1]$ telle que pour tout i $\sum_{j=0}^n M_{i,j} = 1$. La **matrice stochastique**, dite aussi **matrice de transition**, associée à une chaîne de Markov homogène telle que pour tout i dans E il existe n tel que $P(X_n = i) > 0$ ^a est la matrice M définie par $M_{i,j} = P(X_{n+1} = j | X_n = i)$.

^aCas auquel on peut toujours se ramener, en restreignant E .



Cela signifie simplement que l'état à l'instant n (ie X_n) ne dépend que de l'état à l'instant x_{n-1} et pas 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 2.6, sont des exemples de chaînes de Markov.

Remarquons l'égalité de Chapman-Kolmogorov : $P(X_{m+n} = j | X_0 = i) = \sum_{k \in E} P(X_m = j | X_0 = k) P(X_n = k | X_0 = i)$

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 68 Si X est un processus de Markov de matrice de transition M

^a

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

^aCeci impliquant que X est une chaîne de Markov homogène et que pour tout i dans E il existe n tel que $P(X_n = i) > 0$.

FLEMMARD

2.8 Zoologie des lois de probabilité

2.8.1 Lois normales

FLEMMARD

2.8.2 Loi de Bernoulli

Proposition 69 (Loi de Bernoulli) • Paramètre : $B(p)$ a pour paramètre $p \in [0, 1]$

- A valeurs dans $\{0, 1\}$
- Loi : $P(X = 1) = p$ & $P(X = 0) = 1 - p$
- Espérance : p
- Variance : $p \cdot (1 - p)$
- Fonction caractéristique : $\phi(t) = 1 - p + p \cdot e^{it}$
- Intuition : pile ou face si $p = \frac{1}{2}$, pile ou face "biaisé" sinon

2.8.3 Loi binomiale et multinomiale

▣ Loi binomiale

Proposition 70 (Loi binomiale) • Paramètres : $B(n, p)$ a pour paramètres n dans \mathbb{N} et $p \in [0, 1]$

- A 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 : $n \cdot p$
- Variance : $n \cdot p \cdot (1 - p)$
- Fonction caractéristique : $\phi(t) = (1 - p + p \cdot e^{it})^n$
- Intuition : somme de n lois de Bernoullis de même paramètre p .
- Signe particulier : la somme de deux variables aléatoires 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 lois 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 $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda$, alors $B(n, p_n)$ converge en loi vers une variable aléatoire de loi de Poisson $P(\lambda)$. Noter que seul le produit $n \cdot p_n$ compte pour ce passage à la limite ; d'où l'additivité des lois de Poisson quel que soient leurs paramètres.

☐ **Loi géométrique**

Proposition 71 (Loi géométrique) • Paramètre : $G(p)$ a un paramètre $p \in]0, 1]$

- A valeurs dans \mathbb{N}
- Loi : $P(X = k) = p \cdot (1 - p)^k$
- Espérance : $\frac{1-p}{p}$
- Variance : $\frac{1-p}{p^2}$
- Fonction caractéristique : $(\frac{p}{1-q \cdot e^{it}})^n$ FLEMMARD
- 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 géométrie $G(p)$.

☐ **loi binomiale négative**

Proposition 72 (Loi binomiale négative) • Paramètres : $B^-(n, p)$ a deux paramètres $n \in \mathbb{N}$ et $p \in]0, 1]$

- A valeurs dans \mathbb{N}
- Loi : $P(X = k) = C_{n+k-1}^{n-1} p^n \cdot (1 - p)^k$ pour tout $k \in \mathbb{N}$
- Espérance : $n \cdot \frac{1-p}{p}$
- Variance : $n \cdot \frac{1-p}{p^2}$
- Fonction caractéristique : $\phi(t) = (\frac{p}{1-q \cdot e^{it}})^n$ FLEMMARD
- Intuition : c'est un peu comme une série géométrique, à part que l'on attend d'avoir gagné n fois ; on compte le nombre d'échecs.
- Cas limite : $B^-(1, p) = G(p)$

☐ **Loi multinomiale**

Proposition 73 (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 $\sum_{i=1}^d p_i = 1$

- A valeurs $(n_1, \dots, n_d) \in [0, n]^d$, avec $\sum_{i=1}^d n_i = n$
- Loi : $P(X = (n_1, \dots, n_d)) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_d!}$ si $\sum_{i=1}^d n_i = n$ et 0 sinon
- Espérance : $(n \cdot p_1, n \cdot p_2, \dots, n \cdot p_d)$
- Matrice de covariance : $M_{i,j} = -n \cdot p_i \cdot p_j$ si $i \neq j$, $M_{i,i} = n \cdot p_i \cdot (1 - p_i)$
- Fonction caractéristique : $\phi(t) = FLEMMARD$
- Intuition : on tire au sort n fois un nombre entier entre 1 et d , et la i -ième composante représente le nombre de fois que l'on a tiré l'entier i .

2.8.4 Loi de Poisson

Proposition 74 (Loi de Poisson) • Paramètre : $\mathcal{P}(\lambda)$ a pour paramètre $\lambda \in \mathbb{R}$

- A valeurs dans \mathbb{N}
- Loi : $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$
- Espérance : λ
- Variance : λ
- Fonction caractéristique : $\phi(t) = e^{\lambda \cdot (e^{it} - 1)}$
- Intuition : cas limite de la loi binomiale (voir partie 2.8.3)
- 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 suit une loi $\mathcal{P}(\lambda)$, alors $\frac{X - \lambda}{\sqrt{\lambda}}$ converge en loi vers une loi normale $\mathcal{N}(0, 1)$ quand λ tend vers $+\infty$.

2.8.5 Loi hypergéométrique

Proposition 75 (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 $q \in \{0, 1, \dots, N\}$.

- A 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 $n - N \cdot (1 - p)$ et inférieur ou égal à n et à $N \cdot p$.
- Espérance : $n \cdot p$ (indépendante de N !)
- Variance : $\frac{N-n}{N-1} \cdot n \cdot p \cdot (1 - p)$
- Fonction caractéristique : FLEMMARD
- 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 : une suite de variables aléatoires suivant une loi hypergéométrique $H(N, n, p)$ converge en loi quand $N \rightarrow \infty$ vers une loi binomiale $B(n, p)$ (logique intuitivement !).

FLEMMARD + illustration en matlab d'une au moins de ces fonctions (toutes les caracs.)

2.9 Loi des grands nombres

FLEMMARD

Exemple Matlab : texte du programme lgn.m

```
function f = lgn(i,j,n,p)

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 2.1.

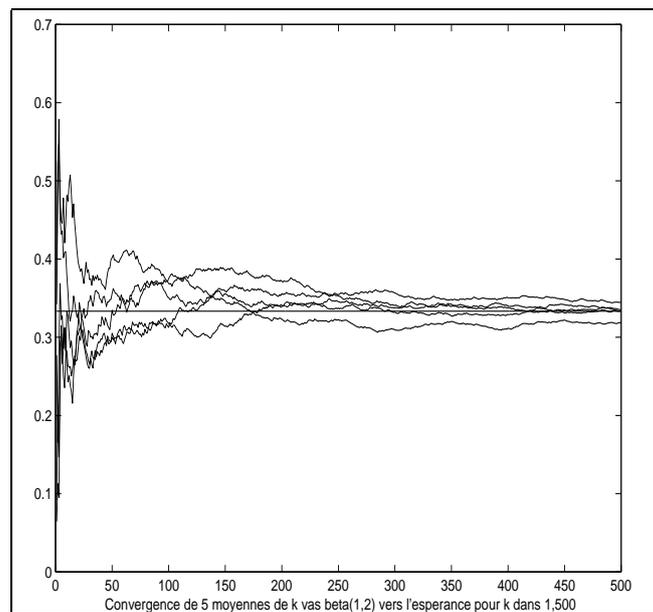


FIG. 2.1 – Loi des grands nombres

2.10 Théorème central limite

FLEMMARD citer comme application les grandes déviations et d'autres trucs. On donne ici deux exemples d'illustration, l'un en matlab, l'autre en Maple.

Exemple Matlab : texte du programme tcl.m

```
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. Je n'ai pas représenté la figure obtenue car il s'agit d'une séquence, qu'on ne peut rendre sur papier sans occuper une place abusive...

Exemple Maple

```
> restart;with(stats);
[anova, describe, fit, importdata, random, statevalf, statplots, transform]
> unif:= i -> random[uniform[0,1]](i);
      unif := randomuniform0,1
> normale:= i -> random[normald[0,1]](i);
      normale := randomnormald0,1
> fit[leastsquare[[x,y,z],z=d*x*y+e*x^2+f*y^2+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 = -1/2 xy - 23/6 x^2 + 1/6 y^2 + 125/6 x + 5/2 y - 160/9
> with(statplots);histogram([normale(50)],[normale(200)],
[normale(800)],[normale(6400)],numbars=40,area=1);
[boxplot, histogram, scatterplot, xscale, xshift, xyexchange, xzexchange, yscale, yshift,
yzerchange, zscale, zshift]
```

2.11 Inégalité de Cramer-Chernoff, grandes déviations

On pourra consulter le livre [18], lecture 16, avec profit. Une application possible des grandes déviations est illustrée en 2.12.5.

FLEMMARD

Le programme suivant illustre le résultat prédit par le résultat FLEMMARD ci-

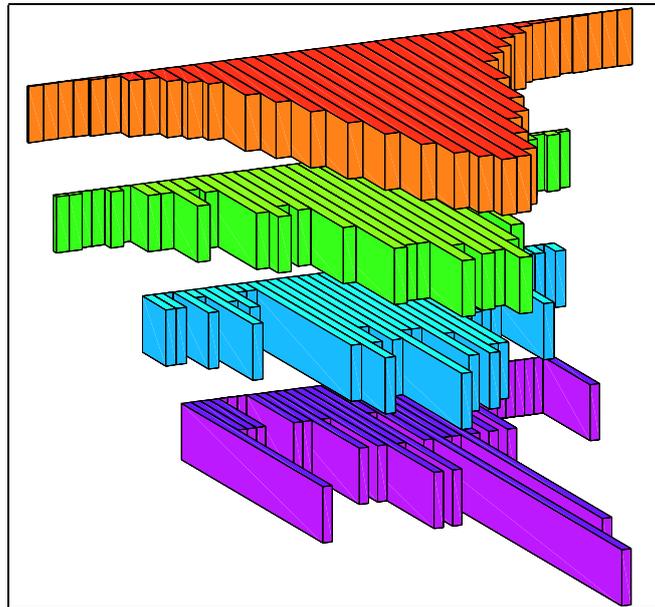


FIG. 2.2 – Théorème central limite : convergence vers la loi normale.

dessus.

Exemple Matlab : texte du programme gd.m
<pre> 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)); text(n/2,m(1,floor(n/2)),sprintf('%g',c)); text(n/2,m(2,floor(n/2)),sprintf('%g',2*c)); text(n/2,m(3,floor(n/2)),sprintf('%g',3*c)); text(n/2,m(4,floor(n/2)),sprintf('%g',4*c)); </pre>

Le résultat est illustré en figure 2.3 ; il faut noter le fait que la courbe est bien linéaire.

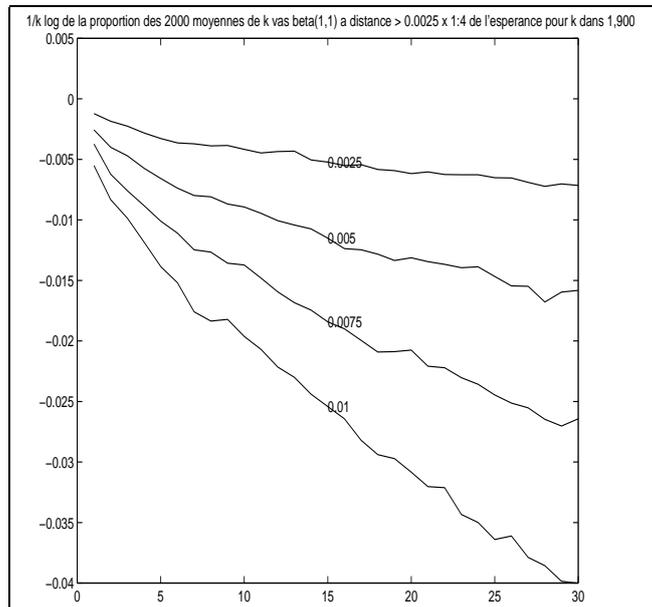


FIG. 2.3 – Grandes déviations

2.12 Applications des probabilités

2.12.1 Liste d'exemples simples

▣ Les singes tapant Shakespeare

On note SK la concaténation de toutes les œuvres de Shakespeare. Un singe est placé devant une machine à écrire ; il tape un caractère par seconde, chaque caractère ayant une probabilité > 0 d'être tapé à la seconde t .

→ la probabilité pour que le singe tape SK une infinité de fois est 1

Démonstration : Par le deuxième lemme de Borel-Cantelli.□

Si la suite de caractères de SK est désigné par $(u_n)_{n \in [1, N]}$, alors la probabilité pour que SK soit tapé au moins k fois en temps P est FLEMMARD.□

Démonstration : FLEMMARD□

2.12.2 Application des probabilités au calcul d'intégrales

On se donne une fonction $f \in L^1$ 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

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n f(X_i) - I\right| \leq \epsilon\right) \leq FLEMMARD$$

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

2.12.3 ABRACADABRA : application des martingales

Le résultat suivant est extrait de [21] (en tant qu'exercice non corrigé) :

Proposition 76 *Soit T la variable aléatoire égale au temps (en secondes) pendant lequel un singe tape sur une machine à écrire une lettre au hasard par seconde jusqu'à obtenir la suite "ABRACADABRA", chaque lettre ayant une probabilité $1/26$ d'être tirée au sort à chaque seconde. Alors $E(T) = 26^{11} + 26^4 + 1$.*

Démonstration : On imagine qu'un nouveau joueur intervient à chaque seconde $n \in [0, T]$, et parie sur la prochaine lettre fournie par le singe. Il mise un franc sur A , et gagne 26 francs s'il a raison. A la seconde suivante (alors même qu'un autre joueur arrive), s'il a perdu il arrête de jouer, et s'il a gagné il remise tout ce qu'il a gagné sur B , gagnant 26^2 francs s'il gagne, et perdant 26 francs s'il perd. Il continue ainsi jusqu'à gagner 26^{11} francs avec le mot complet ABRACADABRA ou jusqu'à perdre.

Il est clair que le gain total de l'ensemble des joueurs est une martingale (par le théorème 64), et que donc par le théorème d'arrêt éventuel de Doob (66) permet de conclure que l'espérance est nulle. On en déduit le résultat demandé en constatant que les pertes sont égales à $T - 26^{11} - 26^4 - 26$ (examiner les joueurs encore en jeu à l'instant T).□

2.12.4 Application au calcul de la longueur d'une courbe

FLEMMARD

2.12.5 Application à l'évaluation de la perte de précision dans un algorithme

On peut utiliser ici des résultats divers concernant la limite d'une somme de variables aléatoires identiquement distribuées ; grandes déviations, théorème central limite, inégalité de Bienaymé-Tchébychev 50 FLEMMARD...

On suppose qu'un programme effectue un calcul et qu'à chaque étape on ajoute un nombre, et que la perte de précision est la valeur absolue de la somme de variables aléatoires indépendantes identiquement distribuées.

A chaque étape, on a un u_n représentant la valeur obtenue après le n -ième calcul, et X_n la différence entre u_{n+1} et $f(u_n)$ la valeur que l'on obtiendrait si l'ordinateur avait une précision infinie à cette étape.

C'est-à-dire que, S_n représentant la somme des X_i pour i variant de 1 à n , les X_i étant des variables aléatoires identiquement distribuées, en supposant que l'espérance de X_i est nulle, on a les résultats suivants :

- $E(S_n) = 0$
- Par l'inégalité de Bienaymé-Tchébitchev, si la variance est $\sigma^2 < \infty$: $P(|S_n| \geq c) \leq \frac{n\sigma^2}{c^2}$
- Par le théorème central limite,

$$P(|S_n| > c\sqrt{n}\sigma) \rightarrow_{n \rightarrow \infty} 2P(N > c)$$

avec N variable aléatoire de loi normale $N(0, 1)$ (ceci permet donc la construction d'intervalles de confiance).

- Par les résultats sur les grandes déviations,

$$P(|S_n| > cn)$$

avec c positif décroît exponentiellement en n .

2.12.6 Application des probabilités à la géométrie euclidienne

Soient $(x_i)_{i \in [1, d]}$ une famille de d vecteurs de \mathbb{R}^d .

On se donne $(p_i)_{i \in [1, d]}$ une famille de réels dans $[0, 1]$.

Définissons $x = \sum_i p_i \cdot x_i$.

Montrons qu'il existe $(\epsilon_i)_{i \in [1, d]}$ une famille de d éléments de $\{0, 1\}$ tels que $\|w -$

$$\sum_i \epsilon_i \cdot x_i\| \leq \sqrt{\sum_{i \in [1, d]} \|u_i\|^2} / 2.$$

On définit (X_1, \dots, X_d) un vecteur aléatoire de loi

$$P(X = (x_1, \dots, x_d)) = \prod_{i \in [1, d]} p_i^{x_i} \cdot (1 - p_i)^{1 - x_i}$$

pour $(x_1, \dots, x_d) \in \{0, 1\}^d$

(c'est à dire, intuitivement, que la probabilité pour X_i d'être égal à 1 est p_i , et que les composantes sont indépendantes)

et on définit $X = \sum X_i \cdot x_i$; X est un vecteur aléatoire.

Calculons l'espérance de $\|X - w\|^2$.

$$\begin{aligned} & E(\|X - w\|^2) \\ &= E\left(\sum_{(i, j) \in [1, d]^2} (X_i - p_i) \cdot (X_j - p_j) \cdot \langle u_i, u_j \rangle\right) \\ &= \sum_{i=1}^d \|u_i\|^2 p_i \cdot (1 - p_i) \\ &= \sum_{i=1}^d \|u_i\|^2 / 4 \end{aligned}$$

D'où le résultat. \square

2.12.7 Probabilité pour que le rapport de #piles/(#piles+#faces) tende vers $\frac{1}{2}$

FLEMMARD

2.12.8 Proportion de diviseurs de n dans $[1, \iota]$

Ce résultat est extrait de [1].

Définition 77 On note $\mu(n)$ le nombre de diviseurs premiers de n .

Soit u_n une suite tendant vers $+\infty$. On définit E_n^i l'ensemble des $i \in [1, n]$ tels que

$|\mu(i) - \log(\log(n))|$ est supérieur ou égal à $a(n)\sqrt{\ln(\ln(n))}$. Alors l'objectif va être de prouver que :

$$\lim_{n \rightarrow \infty} \frac{|E_n^i|}{n} = 0$$

L'intérêt sera de prendre alors $a(n) = \sqrt{\ln(\ln(n))}$, et on arrivera à la conclusion que la proportion d'entiers i tels que $\mu(i)/\ln(\ln(i))$ soit plus loin de 1 que ϵ tendra vers 0 pour $n \rightarrow \infty$.

Montrons donc notre résultat.

- Pour cela on considère les lois de probabilités ($prob_n$) à valeurs dans \mathbb{N}^* pour $n \in \mathbb{N}$, définie par $prob_n(i) = 1/n$ si $i \leq n$ et 0 sinon.

- On définit ensuite $div(p, n)$, application de $P \times \mathbb{N}^*$ dans $\{0, 1\}$, avec P l'ensemble des nombres premiers, avec $div(p, n) = 1$ si $p|n$ et 0 sinon.

- On constate que $\mu(n) = \sum_{p \in P} (div(p, n))$.

- On constate aussi que $\frac{|E_n^i|}{n} = prob_n(|\mu - \ln(\ln(n))| \geq a(n) \cdot \ln(\ln(n)))$

- Déterminons maintenant l'espérance de $div(p, \cdot)$, pour la probabilité $prob_n$.

$$E(div(p, \cdot)) = \lfloor n/p \rfloor / n = 1/p + O(1/n)$$

- On détermine maintenant l'espérance de μ ; elle est somme des espérances des $div(p, \cdot)$, or le théorème des nombres premiers FLEMMARD affirme que

$$\sum_{p \in P \wedge p \leq n} 1/p = \ln(\ln(n)) + o(1)$$

pour $n \rightarrow \infty$

Donc toujours pour $prob_n$

$$E(\mu) = \sum_{p \in P \wedge p \leq n} 1/p + O(1/n)$$

$$= \ln(\ln(n)) + o(1)$$

- Déterminons maintenant la variance de $div(p, \cdot)$.

$$Var(div(p, \cdot)) = E((div(p, \cdot) - E(div(p, \cdot)))^2)$$

$$= E((div(p, \cdot) - 1/p + O(1/n))^2)$$

$$= \frac{1}{p} \cdot \frac{p-1}{p} + \frac{p-1}{p} \frac{1}{p} + O(1/n)$$

$$= \frac{2 \cdot (p-1)}{p^2}$$

- Déterminons maintenant la covariance de $div(p, \cdot)$ et $div(q, \cdot)$

$$Cov(div(p, \cdot), div(q, \cdot)) = E(div(p, \cdot) \cdot div(q, \cdot)) - E(div(p, \cdot))E(div(q, \cdot))$$

$$= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n}$$

$$\leq \frac{1}{pq} - (p-1/n) \cdot (q-1/n)$$

$$\leq \frac{1}{n}(1/p + 1/q)$$

- On peut donc maintenant calculer la variance de μ .

$$\begin{aligned} \text{Var}(\mu) &= \sum_{p \in P \wedge p \leq n} \text{Var}(\text{div}(p, \cdot)) \\ + 2 \sum_{(p,q) \in P^2 \wedge p,q \leq n} \text{Cov}(\text{div}(p, \cdot), \text{div}(q, \cdot)) \\ &\leq \ln(\ln(n)) + O(1) \end{aligned}$$

- On peut alors appliquer l'inégalité de Tchebitchev :

$$\text{prob}_n((X - E(x))^2 \geq t^2) \leq \frac{\text{Var}(\mu)}{t^2}$$

donc

$$\text{prob}_n(|X - \ln(\ln(n))| \geq t\sqrt{\ln(\ln(n))}) \leq 1/t^2$$

La preuve est ainsi complète !□

2.12.9 Processus de branchement

Je suis ici la démarche du chapitre 0 de [21], qui m'a semblé la plus intuitive ; on trouvera des résultats similaires dans [18], et des prolongements et illustrations dans d'autres livres cités par Williams dans [21] ; Feller, Ross.

On se donne une variable aléatoire X à valeurs dans \mathbb{N} . Intuitivement, X correspond au nombre d'enfants d'un animal donné.

On suppose que $P(X = 0) > 0$.

On note f la **fonction génératrice** de X , c'est-à-dire

$$f(\theta) = E(\theta^X) = \sum_{k \geq 0} P(X = k)\theta^k$$

On considère une suite double de variables aléatoires identiquement distribuées $X_{n,m}$ toutes distribuées comme X .

On définit $Z_0 = 1$, et $Z_{n+1} = \sum_{i \in [1, Z_n]} X_{n,i}$.

On note f_n la **fonction génératrice** de Z_n , c'est-à-dire $f_n(\theta) = E(\theta^{Z_n}) = \sum_{k \geq 0} P(Z_n = k)\theta^k$.

Intuitivement, Z_n est le nombre d'individus d'une espèce à l'instant n , descendant d'un même individu, qui est seul à l'instant $n = 0$ (puisque $Z_0 = 1$).

On définit aussi Π la probabilité d'extinction, définie par $\Pi = P(\exists n / Z_n = 0)$, et Π_n la probabilité d'extinction avant l'instant n , définie par $\Pi_n = P(Z_n = 0)$.

Lemme 78 Pour θ dans $[0, 1[$

$$f_n(\theta) = f^n(\theta)$$

c'est-à-dire, par définition de $f^n(\theta)$,

$$= \underbrace{f \circ f \circ \dots \circ f}_{n \text{ fois}}$$

Démonstration :

Le cas $n = 0$ est clair, $n = 1$ aussi, on procède ensuite par récurrence ; il suffit donc de montrer que $f_{n+1}(\theta) = f_n(f(\theta))$.

Or,

$$\begin{aligned} f_{n+1}(\theta) &= E(\theta^{Z_{n+1}}) \\ &= E(E(\theta^{Z_{n+1}}|Z_n)) \end{aligned}$$

par les propriétés de l'espérance conditionnelle,

Or

$$\begin{aligned} E(\theta^{Z_{n+1}}|Z_n = k) &= E(\theta^{X_{n,1}+X_{n,2}+\dots+X_{n,k}}) \\ &= \prod_{l=1}^k E(\theta^{X_{n,l}}) = E(\theta^X)^k \end{aligned}$$

vu l'indépendance des $\theta^{X_{n,m}}$

$$= f(\theta)^k$$

$$f^{n+1}(\theta) = E(\theta^{Z_{n+1}}) = \sum_k E(\theta^{Z_{n+1}}|Z_n = k)P(Z_n = k) = f_n(f(\theta))$$

D'où le résultat.□

Théorème 79 Si $E(X) > 1$ alors la probabilité d'extinction Π est l'unique racine de l'équation $\Pi = f(\Pi)$ située entre 0 et 1 strictement, et sinon, $\Pi = 1$.

Démonstration :

$\pi = \limsup_n \pi_n$ par le théorème de convergence monotone ??.

Par le lemme 78, $\pi_n = f(\pi_{n-1})$. Par continuité de f , $\pi = f(\pi)$.

Il ne reste plus qu'à considérer le graphe de f , convexe, croissante, vérifiant $f(0) > 0$, pour conclure...□FLEMMARD faire un dessin

FLEMMARD on peut dire plus de choses sur les processus (i ?) de branchement

2.12.10 Calcul de surface minimale

On se donne un compact K de \mathbb{R}^2 , et δK son contour. On suppose donnée une fonction g définie sur δK . Soit E l'ensemble des applications de K dans \mathbb{R} . Chaque f appartenant à E définit une surface, l'ensemble des $\{(x, y, f(x, y)) / (x, y) \in K\}$. On cherche parmi E la fonction définissant la surface minimale. On admet le fait que la fonction vérifiant cette propriété est celle dont le laplacien est nulle, et qu'elle 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 :

Exemple Matlab : front.m
<pre>function f=front(x,y) a=abs(x-round(x)); b=abs(y-round(y)); if (a<b) f=1; else f=0; end;</pre>

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 (ie les 4 directions sont équiprobables). On définit un temps d'arrêt T égal au nombre d'étapes avant que la marche aléatoire atteigne δK , ie 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 : lapla.m

function v=lapla(n,e)
u=zeros(n+1,n+1);
for i=0:n,
for j=0:n,
disp(sprintf('%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;
surfl(u)
shading interp
colormap autumn
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 [2.5](#), à gauche. En remplaçant la fonction "front.m" par "cos(4*atan((x-0.5)/(y-0.5)))", on obtient la figure [2.5](#), à droite.

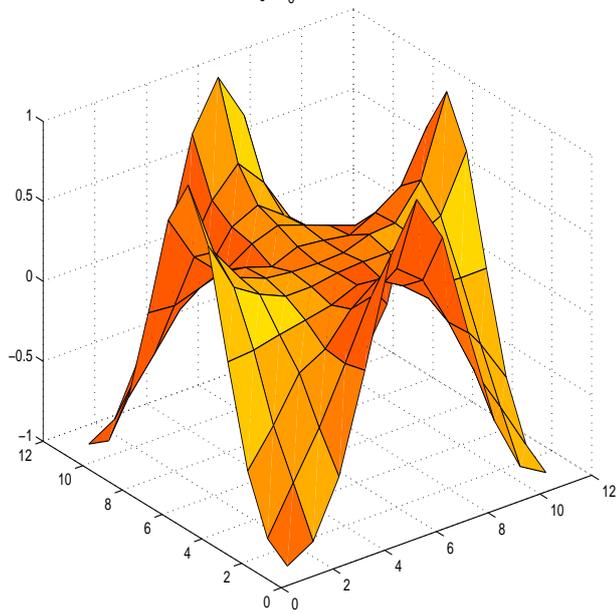
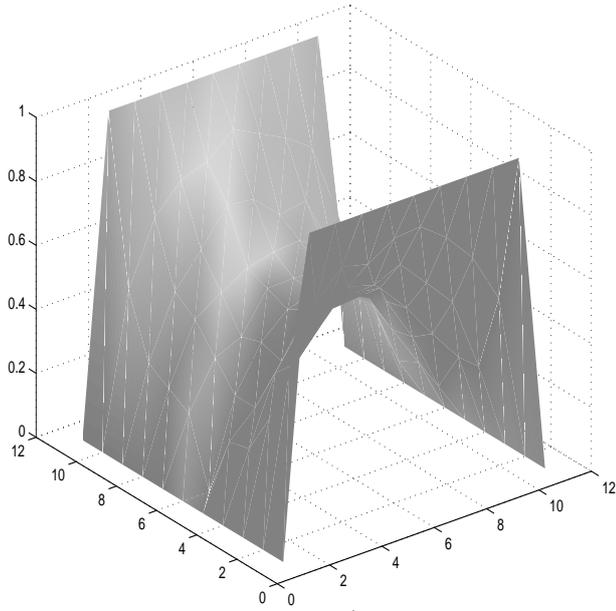


FIG. 2.4 – Approximation de surface minimale obtenue par "lapla.m".

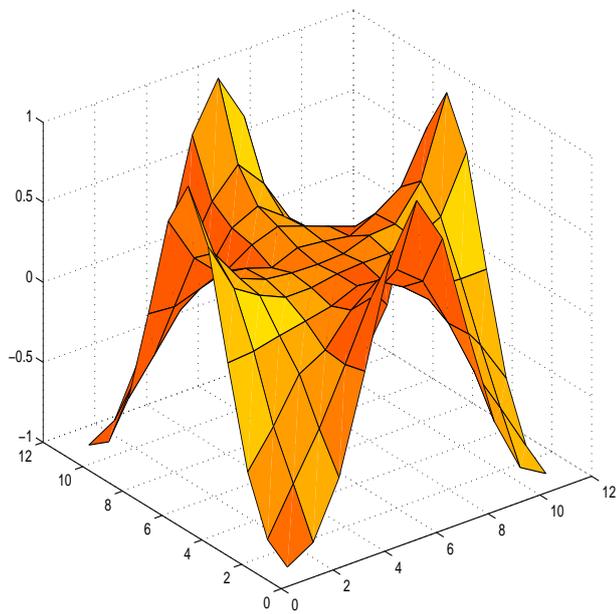


FIG. 2.5 – Approximation de surface minimale obtenue par "lapla.m".

Chapitre 3

Statistique

3.1 Quelques notions élémentaires

3.1.1 Définitions

Définition 80 On appelle **moyenne arithmétique** de n entiers 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} .

Définition 81 On appelle **moyenne géométrique** de n entiers x_1, \dots, x_n la quantité $\sqrt[n]{\prod_i x_i}$.

Définition 82 On appelle **moyenne harmonique** des x_i l'inverse de la moyenne arithmétique des inverses des x_i .

Définition 83 On appelle **moyenne quadratique** des x_i la racine carrée de la moyenne arithmétique des carrés des x_i .

Définition 84 On appelle **médiane** d'une mesure finie sur un espace ordonné un élément x tel que la mesure de $\{y/y > x\}$ est égale à la mesure de $\{y/y < x\}$. Cette notion est définie lorsque la mesure est finie.

Définition 85 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.

Définition 86 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. Bien évidemment il est nécessaire que la mesure soit finie pour cela. On définit de même des **quartiles**, des **déciles**. On appelle **interquartile** la différence entre le troisième et le premier quartile.

Définition 87 On appelle **mode** ou **dominante** d'une distribution la valeur x telle que la densité de probabilité en x soit maximale. S'il y a plusieurs modes la distribution est dite **plurimodale**.

Définition 88 On appelle **déviaton** de x_i la valeur $x_i - \bar{x}$.

Définition 89 On appelle **écart moyen** la moyenne des $|x_i - \bar{x}|$; c'est donc $\overline{|x_i - \bar{x}|}$.

Définition 90 On appelle **variance** la moyenne des $(x_i - \bar{x})^2$; on la note souvent V .

Définition 91 On appelle **écart type** ou **écart quadratique moyen** la racine carrée de la variance. On le note souvent σ ; $\sigma = \sqrt{V}$.

Définition 92 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.

Définition 93 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.

Définition 94 On appelle **moment d'ordre p** des x_i par rapport à y la moyenne des $(x_i - y)^p$. Pour $p = 1$ il s'agit donc de la moyenne (arithmétique), pour $p = 2$ il s'agit de la variance.

3.1.2 Propriétés

- 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.

3.2 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. Les livres bien faits d'introduction abondent FLEMMARD.

Nous nous contenterons ici de donner un cas d'applications, et de citer d'autres cadres d'applications.

Soit X_1, \dots, X_n indépendantes identiquement distribués. On suppose le théorème central limite ?? 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 du tirage. 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**.

On peut citer les développements suivants :

- le cas des petits échantillons ($n < 30$). Il faut alors utiliser la loi de Student.
- 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 de voir pour cela que $P(N > t) = \frac{1}{2}P(|N| > t)$ pour toute variable aléatoire N symétrique, et en particulier donc la loi normale.
- le cas de X_i à valeur dans $\{0, 1\}$.
- le cas où l'on n'étudie pas la moyenne des X_i mais leur max.
- le cas de X_i non réellement indépendants.
- le cas de X_i non identiquement distribués.
- l'évaluation de la loi de m par la fameuse technique du bootstrap, très ingénieuse.
- on a fait l'approximation que pour $n > 30$, la distribution était environ celle de la loi normale. On peut se passer de cette approximation, et donner une borne absolue à l'écart à la loi normale.

Ces études et d'autres encore constituent la théorie des tests et font appel à des variantes difficiles du théorème central limite. Les trois premiers points sont utiles pour faire bonne impression à un jury d'agrégation...

Bibliographie

- [1] P. BARBE, M. LEDOUX, *Probabilité*, BELIN, 1998.
- [2] H. BRÉZIS, *Analyse fonctionnelle*, MASSON, 1983.
- [3] H. CARTAN, *Calcul différentiel*, FLEMMARD.
- [4] A. CHAMBERT-LOIR, S. FERMIGIER, V. MAILLOT, *Exercices de mathématiques pour l'agrégation, Analyse 1*, MASSON, 1997.
- [5] A. CHAMBERT-LOIR, S. FERMIGIER, *Exercices de mathématiques pour l'agrégation, Analyse 2*, MASSON, 1995.
- [6] A. CHAMBERT-LOIR, S. FERMIGIER, *Exercices de mathématiques pour l'agrégation, Analyse 3*, MASSON, 1996.
- [7] P.G. CIARLET, *Introduction à l'analyse numérique matricielle et à l'optimisation*, DUNOD, 1998.
- [8] F. COMBES *Algèbre et géométrie*, BRÉAL, 1998. <
- [9] J.-P. DEMAILLY, *Analyse numérique et équations différentielles*, PRESSES UNIVERSITAIRES DE GRENOBLE, 1996.
- [10] W. GIORGI, *Thèmes mathématiques pour l'agrégation*, MASSON, 1998.
- [11] A. GRAMAIN, *Intégration*, HERMANN 1988, PARIS.
- [12] J.-L. KRIVINE, *Introduction to axiomatic set theory*, D. REIDEL PUBLISHING COMPANY, DORDRECHT-HOLLAND.
- [13] S. LANG, *Real analysis*, ADDISON-WESLEY PUBLISHING COMPANY, 1969.
- [14] D. PERRIN, *Cours d'algèbre*, ELLIPSES 1996.
- [15] A. POMMELLET, *Cours d'analyse*, ELLIPSES 1994.
- [16] W. RUDIN, *Analyse réelle et complexe*, MASSON 1992.
- [17] R. SMULLYAN, *Théorie de la récursion pour la métamathématique*, FLEMMARD.
- [18] Y.G. SINAI *Probability theory - An introduction course*, SPRINGER TEXT-BOOK, 1992.
- [19] P. TAUVEL, *Mathématiques générales pour l'agrégation*, MASSON, 1997.
- [20] J. VAUTHIER, J.J. PRAT, *Cours d'analyse mathématiques de l'intégration*, MASSON, 1994.
- [21] D. WILLIAMS, *Probability with martingales*, CAMBRIDGE UNIVERSITY PRESS, 1991.
- [22] C. ZUILY, H. QUEFFÉLEC, *Éléments d'analyse pour l'intégration*, MASSON, 1995.