LES PROBABILITÉS POUR LES OPTIONS B, C ET D

PRÉPARATION À L'AGRÉGATION EXTERNE DE MATHÉMATIQUES DE L'UNIVERSITÉ RENNES 1 [1]
ANNÉE 2009/2010

1. ESPACE PROBABILISÉ - VARIABLE ALÉATOIRE

1.1 ESPACE PROBABILISÉ

Lors d'un lancer de dé, on peut raisonnablement convenir que les 2 seules issues possibles sont "Pile" ou "Face". L'ensemble des résultats possibles de l'expérience, ou *univers*, est alors $\Omega = \{\text{Pile}, \text{Face}\}$. Si la pièce est équilibrée, on peut affecter à cet ensemble une *probabilité* $\mathbb{P}$ telle que pour chaque $\omega \in \Omega$, appelé *éventualité*, $\mathbb{P}(\{\omega\}) = 1/2$.

Afin de rendre compte de situations plus complexes, on doit considérer le cas où l'univers Ω associé à l'expérience n'est pas dénombrable. Pour des raisons techniques, on est alors contraint d'associer à Ω une tribu de parties $\mathscr{F}$, et le couple $(\Omega, \mathscr{F})$ s'appelle *espace probabilisable*. Définie à partir de l'univers, la tribu $\mathscr{F}$ représente les issues possibles de l'expérience ; un élément de $\mathscr{F}$ est donc naturellement appelé *événement*. Si $\mathbb{P}$ est une probabilité sur $(\Omega, \mathscr{F})$, i.e. une mesure de masse 1 sur $(\Omega, \mathscr{F})$, le triplet $(\Omega, \mathscr{F}, \mathbb{P})$ s'appelle *espace probabilisé*.

De manière implicite, on se place dans toute la suite sur l'espace probabilisé $(\Omega, \mathscr{F}, \mathbb{P})$. Les propriétés ci-dessous sont des conséquences immédiates de la définition d'une probabilité :

- ▷ $\mathbb{P}(\emptyset) = 0$;
- ▷ $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, pour 2 événements A, B [EGALITÉ DE POINCARÉ] ;
- ▷ $\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$ pour un événement A ;
- ▷ $\mathbb{P}(A) \le \mathbb{P}(B)$ et $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ pour 2 événements A, B tels que $A \subset B$;
- ▷ $\mathbb{P}\big(\cup_n A_n\big) \le \sum_n \mathbb{P}(A_n)$, pour une suite d'événements $(A_n)_n$ [INÉGALITÉ DE BOOLE].

1.2 VARIABLE ALÉATOIRE ET LOI

Si l'on veut décrire la durée de vie d'une ampoule par exemple, le concept d'espace probabilisé n'est pas adapté. On introduit alors la notion de *variable aléatoire* :

Définition. *On appelle variable aléatoire à valeurs dans $\mathbb{R}^d$ une application mesurable de $(\Omega, \mathscr{F})$ dans $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. On utilise l'abbréviation v.a., ou v.a.r. pour désigner une v.a. à valeurs réelles ($d = 1$).*

Pour $A \in \mathscr{B}(\mathbb{R}^d)$, on peut ainsi calculer la probabilité de l'événement $X^{-1}(A)$, traditionnellement noté $\{X \in A\}$; il s'agit donc de calculer $\mathbb{P}(\{X \in A\})$. Cette notation est écourtée en $\mathbb{P}(X \in A)$.

Si X est une v.a.r. associée à la durée de vie de l'ampoule, cela signifie que pour chaque éventualité ω, $X(\omega)$, appelé *réalisation de X*, est la durée de vie pour la configuration ω de l'univers. Il faut noter que l'univers Ω a changé de rôle : il ne représente plus l'ensemble des résultats possibles de l'expérience (qui est ici $\mathbb{R}^+$), mais plutôt des conditions expérimentales, qui mènent à telle ou telle valeur pour la durée de vie de l'ampoule. Dans le cadre de cette expérience, il n'est pas possible de décrire l'application X. On se tourne alors vers une notion plus faible, celle de *loi (de probabilité)*.

Définition. *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. La loi de X, notée $\mathbb{P}_X$, est la mesure image de $\mathbb{P}$ par X. Plus concrètement, $\mathbb{P}_X$ est définie par : $\mathbb{P}_X(A) = \mathbb{P} \circ X^{-1}(A) = \mathbb{P}(X \in A)$, $\forall A \in \mathscr{B}(\mathbb{R}^d)$.*

Si $X = (X_1, \cdots, X_d)^t$ est une v.a. à valeurs dans $\mathbb{R}^d$, la loi de chaque coordonnée X_i s'appelle *loi marginale* et on appelle quelquefois *loi jointe* la loi de X, afin de lui donner un statut à part. On notera à ce sujet les faits suivants :

- ▷ La connaissance de la loi jointe permet de retrouver les lois marginales (cf `[O, p. 12]`), car, par exemple, pour chaque $A \in \mathscr{B}(\mathbb{R})$, $\mathbb{P}(X_1 \in A) = \mathbb{P}(X \in A_1)$, si $A_1 = \{x \in \mathbb{R}^d : x_1 \in A\}$;
- ▷ La connaissance des lois marginales ne permet pas -sans autre information- de retrouver la *loi jointe* ;

1. Benoît Cadre - ENS Cachan Bretagne

- ▷ La loi d'une permutation des coordonnées de X n'est en général pas la loi de X, et ceci même si les lois marginales sont identiques.

Pour illustrer simultanément les 2 dernières assertions, on peut considérer le contre-exemple suivant. Sur l'espace probabilisé $(\Omega,\mathbb{P})$, où $\Omega=\{a,b,c\}$ et $\mathbb{P}$ est la probabilité uniforme sur Ω i.e. $\mathbb{P}(\{\omega\})=1/3$ pour tout $\omega\in\Omega$, on note X_1 et X_2 les v.a. telles que $X_1(a)=0$, $X_1(b)=1$, $X_1(c)=2$, $X_2(a)=2$, $X_2(b)=0$ et $X_2(c)=1$. Alors, X_1 et X_2 suivent la même loi uniforme sur $\{1,2,3\}$ i.e. $\mathbb{P}(X_1=i)=\mathbb{P}(X_2=i)=1/3$ pour $i\in\{1,2,3\}$, mais la loi de (X_1,X_2) est différente de celle de (X_2,X_1).

Les exemples de lois usuelles sont absolument continues par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ ou la mesure de comptage sur $\mathbb{Z}^d$, que l'on appelle dans ce contexte des *mesures dominantes*. Le *théorème de Radon-Nikodym* `[O, p. 6]` nous donne alors une fonction f_X (notation adoptée pour toute la suite), appelée *densité de la loi de X par rapport à la mesure dominante* ν, telle que f_X est ν-p.p. positive et $\int f_X d\nu=1$.

Exemples de lois de v.a.r. `[FF1, chap. 7 et 14]`.

- ▷ LOI UNIFORME $\mathscr{U}(D)$, avec $D\subset\mathbb{N}$ fini. C'est la loi sur D, dont la densité par rapport à la mesure de comptage sur D vaut $1/\mathrm{card}(D)$.
- ▷ LOI BINOMIALE $\mathscr{B}(n,p)$, avec $n\in\mathbb{N}$ et $p\in[0,1]$. C'est la loi sur $\{0,\cdots,n\}$, dont la densité par rapport à la mesure de comptage sur $\{0,\cdots,n\}$ vaut $C_n^k p^k(1-p)^{n-k}$. Cas particulier : LOI DE BERNOULLI $B(p)=B(1,p)$.
- ▷ LOI DE POISSON $\mathscr{P}(\lambda)$, avec $\lambda\geq 0$. C'est la loi sur $\mathbb{N}$, dont la densité par rapport à la mesure de comptage sur $\mathbb{N}$ vaut $\exp(-\lambda)\lambda^k/k!$. Pour une v.a. X à valeurs dans $\mathbb{N}$ telle que $\mathbb{P}(X=k)\neq 0\ \forall k\in\mathbb{N}$, on a l'équivalence (cf `[FF1, p. 76]`) : $\exists\lambda\geq 0$ avec $X\sim\mathscr{P}(\lambda)\Leftrightarrow\mathbb{P}(X=n)/\mathbb{P}(X=n-1)=\lambda/n\ \forall n\geq 1$.
- ▷ LOI GÉOMÉTRIQUE $\mathscr{G}(p)$, avec $p\in[0,1]$. C'est la loi sur $\mathbb{N}^*$, dont la densité par rapport à la mesure de comptage sur $\mathbb{N}^*$ vaut $p(1-p)^{k-1}$.
- ▷ LOI UNIFORME $\mathscr{U}([a,b])$, avec $a<b$. C'est la loi sur $[a,b]$, dont la densité par rapport à la mesure de Lebesgue sur $[a,b]$ vaut $1/(b-a)$.
- ▷ LOI NORMALE, OU GAUSSIENNE $\mathscr{N}(m,\sigma^2)$, avec $m\in\mathbb{R}$ et $\sigma>0$. C'est la loi sur $\mathbb{R}$, dont la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ vaut $1/(\sigma\sqrt{2\pi})\exp\big(-(x-m)^2/(2\sigma^2)\big)$.
- ▷ LOI GAMMA $\gamma(a,\lambda)$, avec $a,\lambda>0$. C'est la loi sur $\mathbb{R}^+$, dont la densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^+$ vaut $(\lambda/\Gamma(a))\exp(-\lambda x)(\lambda x)^{a-1}$, avec $\Gamma(a)=\int_0^\infty\exp(-x)x^{a-1}dx$. Cas particulier : LOI EXPONENTIELLE $\mathscr{E}(\lambda)=\gamma(1,\lambda)$. Pour une v.a. X à valeurs dans $\mathbb{R}^+$ qui possède une densité par rapport à la mesure de Lebesgue, on a l'équivalence (cf `[FF1, p. 184]`) : $\exists\lambda>0$ avec X de loi $\mathscr{E}(\lambda)\Leftrightarrow\mathbb{P}(X>x+y)=\mathbb{P}(X>x)\mathbb{P}(X>y)\ \forall x,y>0$.

1.3 THÉORÈME DE TRANSFERT - ESPÉRANCE ET VARIANCE

Revenons au problème de la modélisation de la durée de vie d'une ampoule, représentée par une v.a.r. X. Dans le cadre d'un modèle, on a fixé une loi de probabilité suivie par X, par exemple une loi $\mathscr{E}(\lambda)$. On note alors $X\sim\mathscr{E}(\lambda)$. Une caractéristique importante de l'ampoule est sa durée de vie moyenne. Il s'agit d'une moyenne, pondérée par la loi de probabilité suivie par X. Le *théorème de transfert* permet de formaliser ce point de vue :

Théorème de transfert `[FF1, p.130]`. *Soit X une v.a. à valeurs dans $\mathbb{R}^d$, et g une fonction borélienne de $\mathbb{R}^d$ dans $\mathbb{R}$. Alors, si l'une des expressions $\int_\Omega g\circ X d\mathbb{P}$ ou $\int_{\mathbb{R}^d} g d\mathbb{P}_X$ existe, il en est de même pour l'autre et l'on a :*

$$\int_\Omega g\circ X\,d\mathbb{P}=\int_{\mathbb{R}^d} g\,d\mathbb{P}_X.$$

Pour simplifier, on note $g(X)=g\circ X$ et, sous les conditions d'intégrabilité adéquates :

$$\mathbb{E}(g(X))=\int_\Omega g(X)\,d\mathbb{P}.$$

En particulier, $\mathbb{E}(\mathbf{1}_A(X))=\mathbb{P}(X\in A)$ pour tout borélien A. Par ailleurs, on a $g(X)\in\mathbb{L}^1(\mathbb{P})$ si, et seulement si $g\in\mathbb{L}^1(\mathbb{P}_X)$. Dans les 2 cas fondamentaux (discret et à densité), la formule de transfert s'énonce ainsi :

$$\begin{aligned}\mathbb{E}(g(X)) &= \int_{\mathbb{R}^d} g(x)f(x)dx,\ \text{si } \mathbb{P}_X(dx)=f(x)dx;\\ &= \sum_{k\in\mathbb{Z}^d} g(k)\,\mathbb{P}(X=k),\ \text{si } \mathbb{P}_X=\sum_{k\in\mathbb{Z}^d}\mathbb{P}(X=k)\,\delta_k.\end{aligned}$$

Ce théorème est aussi l'occasion de définir des quantités fondamentales, les *moment*, *variance*, *covariance* et enfin *matrice de variance* :

- ▷ MOMENT D'ORDRE p DE LA V.A.R. X. Si $X \in \mathbb{L}^p(\mathbb{P})$, son *moment d'ordre* p est $\mathbb{E}(X^p)$. L'*espérance, ou moyenne* de X est $\mathbb{E}(X)$;
- ▷ VARIANCE DE LA V.A.R. X. Si $X \in \mathbb{L}^2(\mathbb{P})$, la *variance* de X est :

$$\mathrm{var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2.$$

  Elle représente l'écart des valeurs de X par rapport à sa moyenne. Noter que $\mathbb{E}(X)$ existe car $\mathbb{L}^2(\mathbb{P}) \subset \mathbb{L}^1(\mathbb{P})$, et que si $a \in \mathbb{R}$: $\mathrm{var}(aX) = a^2 \mathrm{var}(X)$. Enfin, $\mathbb{E}(X)$ est, pour le critère $\mathbb{L}^2(\mathbb{P})$, la meilleure approximation de X par une constante.
- ▷ COVARIANCE ENTRE LES V.A.R. X ET Y. Si $X,Y \in \mathbb{L}^2(\mathbb{P})$, la *covariance* entre X et Y est :

$$\mathrm{cov}(X,Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

  Elle joue un rôle crucial dans la notion d'*indépendance* (cf. Section 3). On notera en particulier les relations $\mathrm{cov}(X,X) = \mathrm{var}(X)$, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) - \mathrm{cov}(X,Y)$ et $\mathrm{cov}(X+Z,Y) = \mathrm{cov}(X,Y) + \mathrm{cov}(Z,Y)$ si Z est une v.a.r. de carré intégrable.
- ▷ ESPÉRANCE ET MATRICE DE VARIANCE D'UN VECTEUR ALÉATOIRE. Un *vecteur aléatoire* n'est rien d'autre qu'une v.a. $X = (X_1, \cdots, X_d)^t$ à valeurs dans $\mathbb{R}^d$. Sous les conditions d'intégrabilité adéquates, son *espérance* $\mathbb{E}(X)$ est le vecteur des espérances des coordonnées, et sa *matrice de variance* est définie par $\mathbb{V}(X) = (\mathrm{cov}(X_i,X_j))_{i,j}$. Noter que $\mathbb{V}(X)$ est une matrice symétrique positive.

Quelques calculs d'espérances et variances.

- ▷ Si $X \sim \mathscr{B}(n,p)$, $\mathbb{E}(X) = np$ et $\mathrm{var}(X) = np(1-p)$;
- ▷ Si $X \sim \mathscr{P}(\lambda)$, $\mathbb{E}(X) = \mathrm{var}(X) = \lambda$;
- ▷ Si $X \sim \mathscr{N}(m,\sigma^2)$, $\mathbb{E}(X) = m$ et $\mathrm{var}(X) = \sigma^2$;
- ▷ Si $X \sim \gamma(a,\lambda)$, $\mathbb{E}(X) = a/\lambda$ et $\mathrm{var}(X) = a/\lambda^2$;
- ▷ Si $X \sim \mathscr{G}(p)$ avec $p > 0$, $\mathbb{E}(X) = 1/p$ et $\mathrm{var}(X) = (1-p)/p^2$.

Si l'espérance et la variance ne caractérisent pas la loi d'une v.a., elles contribuent néanmoins au contrôle des déviations de la v.a. En effet, si $X \in \mathbb{L}^1(\mathbb{P})$ est une v.a.r. , on a

$$\forall t > 0 \ : \ \mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}(|X|)}{t} \qquad \text{[Inégalité de Markov]}.$$

Si $X \in \mathbb{L}^2(\mathbb{P})$, on peut estimer plus précisémment la probabilité de déviation de X par rapport à sa moyenne :

$$\forall t > 0 \ : \ \mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{var}(X)}{t^2} \qquad \text{[Inégalité de Bienaymé-Tchebitchev]}.$$

2. CALCUL DE LOIS : FONCTION DE RÉPARTITION, FONCTION CARACTÉRISTIQUE ET TRANSFORMÉE DE LAPLACE

2.1 LE THÉORÈME DE LA LOI IMAGE

Si X possède une densité f sur $\mathbb{R}$ par rapport à la mesure de Lebesgue, et si $\varphi : \mathbb{R} \to \mathbb{R}$ est un difféomorphisme de classe $\mathscr{C}^1$, on a d'après le théorème de transfert, pour tout borélien A :

$$\mathbb{P}(\varphi(X) \in A) = \int_{\mathbb{R}} \mathbf{1}_A(\varphi(x)) \mathbb{P}_X(dx) = \int_{\mathbb{R}} \mathbf{1}_A(\varphi(x)) f(x) dx = \int_A f \circ \varphi^{-1}(y) \big| \big(\varphi^{-1}\big)'(y) \big| dy.$$

Ainsi, $\varphi(X)$ possède une densité par rapport à la mesure de Lebesgue qui est $f \circ \varphi^{-1}(y) |(\varphi^{-1})'(y)|$. On peut généraliser ce résultat :

Théorème de la loi image. `[FF1, p. 195]` *Soit X une v.a. à valeurs dans un ouvert U de $\mathbb{R}^d$, et $\varphi : U \to V$ un difféomorphisme de classe $\mathscr{C}^1$. Si X possède une densité f, alors $Y = \varphi(X)$ possède une densité, qui vaut :*

$$f \circ \varphi^{-1} \, |\mathrm{Jac}_{\varphi^{-1}}| \, \mathbf{1}_V.$$

Le calcul de la loi image se ramène donc dans ce cas à un calcul d'intégrale multiple.

Exemples.

- ▷ Si X est une v.a.r. de densité f par rapport à la mesure de Lebesgue, la v.a.r. $1/X$ possède une densité, qui est $(1/x^2)f(1/x)$. En particulier, si X suit la loi de *Cauchy*, i.e. X est à densité $f_X(x) = 1/(\pi(1+x^2))$ par rapport à la mesure de Lebesgue sur $\mathbb{R}$, alors $1/X \sim X$;
- ▷ Si $m \in \mathbb{R}$ et $\sigma > 0$, $\mathscr{N}(m,\sigma^2) = m + \sigma\mathscr{N}(0,1)$ au sens suivant : $X \sim m+\sigma Y$, si $X \sim \mathscr{N}(m,\sigma^2)$ et $Y \sim \mathscr{N}(0,1)$.

2.2 La fonction de répartition

Note préliminaire. Nous ne parlerons ici que de fonction de répartition pour une v.a.r., car elle est essentiellement utilisée dans ce contexte. Néanmoins, on peut, au prix de quelques complications techniques supplémentaires, définir cette notion pour des v.a. à valeurs dans $\mathbb{R}^d$ (cf `[FF1, p. 142]`).

Définition. *On appelle fonction de répartition (f.r.) toute fonction $F : \mathbb{R} \to [0,1]$ croissante, continue à droite, et telle que $\lim_{-\infty} F = 0$, $\lim_{+\infty} F = 1$.*

Pour une v.a.r. X, la fonction F_X définie par $F_X(x) = \mathbb{P}(X \leq x)$ est une f.r., appelée *fonction de répartition de X* (la notation F_X est adoptée pour toute la suite). Si X possède une densité f par rapport à la mesure de Lebesgue, la dérivée de F_X vaut f ; autrement dit, F_X caractérise la loi, représentée ici par f. Ce phénomène se généralise à tous les types de lois, comme le montre le résultat suivant :

Théorème `[FF1, p. 48-49]`. *A toute fonction de répartition correspond une et une seule mesure de probabilité sur $(\mathbb{R},\mathscr{B}(\mathbb{R}))$. En d'autres termes, 2 v.a.r. ont même loi si, et seulement si, elles ont même f.r.*

On souhaite calculer la loi de X^2, avec $X \sim \mathscr{N}(0,1)$. Ce calcul ne peut pas se faire en utilisant le théorème de la loi image. En revanche, l'usage de la f.r. permet de trouver la loi de X^2 : la dérivée de la f.r. de X^2 est la densité de la loi $\gamma(1/2,1/2)$. Donc $X^2 \sim \gamma(1/2,1/2)$.

Générer une réalisation d'une v.a.r. Supposons que l'on veuille générer une réalisation d'une v.a.r. X. Pour simplifier, on suppose que F_X est strictement croissante (cf `[O, p. 29]` pour le cas général). On sait générer une réalisation u d'une v.a.r. $U \sim \mathscr{U}[0,1]$ (générateur aléatoire). En remarquant que $F_X^{-1}(U) \sim X$ car $F_{F_X^{-1}(U)} = F_X$, on en déduit que $F_X^{-1}(u)$ est une réalisation de X.

2.3 La fonction caractéristique

Définition. *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. On appelle fonction caractéristique (f.c.) de X, et on note φ_X, la transformé de Fourier de $\mathbb{P}_X$. Autrement dit, pour tout $t \in \mathbb{R}^d$:*

$$\varphi_X(t) = \hat{\mathbb{P}}_X(t) = \mathbb{E}\big(\exp(i < t, X >)\big).$$

Une f.c. est définie et continue partout, et bornée par 1. De plus, si X est une v.a. à valeurs dans $\mathbb{R}^d$, on a pour tout $A \in \mathscr{M}_{d,k}$ et $b \in \mathbb{R}^k$,

$$\varphi_{AX+b}(t) = \exp(i < b, t >)\varphi_X(A^\star t),\ \forall t \in \mathbb{R}^k.$$

Théorème d'unicité `[FF1, p. 166] ou [O, p. 201]`. *La fonction caractéristique d'une v.a. à valeurs dans $\mathbb{R}^d$ caractérise sa loi. En d'autres termes, 2 v.a. à valeurs dans $\mathbb{R}^d$ ont même loi si, et seulement si, elles ont même f.c.*

La loi de X est symétrique (i.e. $X \sim -X$) si, et seulement si, φ_X ne prend que des valeurs réelles. Enfin, toute combinaison convexe de f.c. est une f.c.

Il est donc essentiel de calculer les f.c. des lois usuelles. Faisons-le pour $X \sim \mathscr{N}(0,1)$. Comme la loi $\mathscr{N}(0,1)$ est symétrique,

$$\varphi_X(t) = \mathbb{E}\big(\cos(tX)\big) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \cos(tx)e^{-x^2/2}dx.$$

On montre alors que $\varphi_X'(t) = -t\varphi_X(t)$. Comme $\varphi_X(0) = 1$, on obtient $\varphi_X(t) = e^{-t^2/2}$. En utilisant l'égalité $\mathscr{N}(m,\sigma^2) = m + \sigma\mathscr{N}(0,1)$ (cf section 2.1), on en déduit la f.c. de la loi $\mathscr{N}(m,\sigma^2)$.

Fonctions caractéristiques de lois classiques `[FF1, chap. 7 et 14].`

- ▷ LOI BINOMIALE $\mathscr{B}(n,p)$, avec $n \in \mathbb{N}$ et $p \in [0,1]$. Sa f.c. vaut $(1-p+pe^{it})^n$.
- ▷ LOI DE POISSON $\mathscr{P}(\lambda)$, avec $\lambda > 0$. Sa f.c. vaut $\exp(\lambda(e^{it}-1))$.
- ▷ LOI GÉOMÉTRIQUE $\mathscr{G}(p)$, avec $p \in [0,1]$. Sa f.c. vaut $pe^{it}/(1-(1-p)e^{it})$.
- ▷ LOI EXPONENTIELLE $\mathscr{E}(\lambda)$, avec $\lambda > 0$. Sa f.c. vaut $\lambda/(\lambda - it)$.
- ▷ LOI NORMALE, OU GAUSSIENNE $\mathscr{N}(m,\sigma^2)$, avec $m \in \mathbb{R}$ et $\sigma > 0$. Sa f.c. vaut $\exp(itm - \sigma^2 t^2/2)$.
- ▷ LOI GAMMA $\gamma(a,\lambda)$, avec $a, \lambda > 0$. Sa f.c. vaut $(\lambda/(\lambda - it))^a$.

Passons maintenant en revue d'autres propriétés essentielles de la f.c. Dans ce qui suit, on considère le cas d'une v.a.r. Le cas des v.a. à valeurs dans $\mathbb{R}^d$ se traite de manière similaire, au prix de quelques complications techniques supplémentaires.

Théorème. *Soit X une v.a.r.*

(1). `[O, p. 203]`*. Si φ_X est Lebesgue-intégrable, alors X possède une densité par rapport à la mesure de Lebesgue, qui vaut en chaque x :*

$$\frac{1}{2\pi}\int_{\mathbb{R}} \varphi_X(t)\exp(-ixt)dt.$$

(2). `[O, p. 208]`*. Si $X \in \mathbb{L}^n(\mathbb{P})$, alors $\varphi \in \mathscr{C}^n$ et pour chaque $\ell = 1, \cdots, n$: $\varphi_X^{(\ell)}(0) = i^\ell \mathbb{E}(X^\ell)$.*

(3). `[O, p. 208]`*. Si φ_X est k fois dérivable en 0 ($k \geq 2$), alors $X \in \mathbb{L}^{2[k/2]}(\mathbb{P})$. De plus, pour chaque $\ell = 1, \cdots, 2[k/2]$: $\varphi_X^{(\ell)}(0) = i^\ell \mathbb{E}(X^\ell)$.*

(4). `[O, p. 214]`*. Si $X \in \mathbb{L}^n(\mathbb{P})$ pour tout n, et si $\limsup_n \|X\|_{\mathbb{L}^n(\mathbb{P})}/n = 1/R < \infty$, alors φ_X est développable en série entière au voisinage de tout réel, le rayon de convergence étant supérieur à R/e. D'après (2), on a donc le développement :*

$$\varphi_X(t) = \sum_{\ell \geq 0} \frac{(it)^\ell}{\ell!}\mathbb{E}(X^\ell),\ \forall t \in \left] -\frac{R}{e}, \frac{R}{e} \right[.$$

Difficile, dans cette section, de ne pas évoquer la notion de *vecteur (aléatoire) gaussien*. La notion de v.a. gaussienne est essentielle en probabilités (cf section 6.2).

Définition. *On dit que X est un vecteur gaussien de $\mathbb{R}^d$, et on note $X \sim N_d(m,\Sigma)$, si il existe $m \in \mathbb{R}^d$ et $\Sigma \in \mathscr{M}_d(\mathbb{R})$ symétrique positive tels que la fonction caractéristique φ_X de X s'écrit :*

$$\varphi_X(u) = \exp\Big(i < u, m > -\frac{1}{2}u^t\Sigma u\Big),\ \forall u \in \mathbb{R}^d.$$

Remarques `[DD1, p. 130-132].`

- ▷ Cette définition inclut notamment le cas où X suit une loi de Dirac en m (cas $\Sigma = 0$) ;
- ▷ Si $X \sim N_d(m,\Sigma)$, alors $\mathbb{E}(X) = m$ et $\mathbb{V}(X) = (\mathrm{cov}(X_i,X_j))_{i,j=1,\cdots,d} = \Sigma$. En particulier, un vecteur gaussien admet un moment d'ordre 2 et, de manière plus générale, des moments de tous ordres ;
- ▷ Un vecteur gaussien admet une densité par rapport à la mesure de Lebesgue si, et seulement si sa matrice de variance est inversible.

2.2 LA TRANSFORMÉE DE LAPLACE

Définition. *Soit X une v.a. à valeurs dans $\mathbb{R}^d$. On appelle transformée de Laplace de X, et on note $\mathscr{L}_X$, la fonction définie par*

$$\mathscr{L}_X(t) = \mathbb{E}\big(\exp(< t, X >)\big),$$

pour chaque $t \in \mathbb{R}^d$ tel que l'intégrale soit définie.

La transformée de Laplace jouit de propriétés très similaires à la f.c. En particulier, elle caractérise la loi et les moments d'une v.a. peuvent être déduits de dérivations successives de la transformée de Laplace. Son inconvénient majeur est qu'elle n'est en général pas définie sur tout $\mathbb{R}^d$, contrairement à la f.c.

Pour une v.a. X à valeurs dans $\mathbb{R}^d$, on note $I_X = \{t \in \mathbb{R}^d : \mathscr{L}_X(t) \text{ est définie}\}$.

Proposition `[DD1, p. 73]`*. Soit X une v.a. à valeurs dans $\mathbb{R}^d$. Alors, l'ensemble I_X est convexe, et la fonction $\mathscr{L}_X$ est de classe $\mathscr{C}^\infty$ dans l'intérieur de I_X. De plus, la fonction $\ln \mathscr{L}_X$ est convexe sur I_X ; elle est même strictement convexe si X ne suit pas une loi de Dirac.*

Si la transformée de Laplace de X est définie sur un ouvert contenant 0, X possède des moments de tous ordres. En ce sens, l'utilisation de la transformée de Laplace pour caractériser une loi est donc moins pratique que l'utilisation de la f.c. En fait, l'intérêt de la transformée de Laplace en probabilités est à rechercher ailleurs.

Si X est une v.a.r., la *transformée de Cramer* de X (cf `[DD1, p. 78]`) est la fonction h_X définie par :

$$h_X(u) = \sup_{t \in I_X} \big(ut - \ln \mathscr{L}_X(t)\big).$$

Théorème `[DD1, p. 79]`. *Soit X une v.a.r. telle que l'intérieur de I_X soit un voisinage de 0. Alors, $h_X(t) > 0$ si $t \neq \mathbb{E}(X)$, et :*

$$\mathbb{P}(X \geq t) \leq \exp(-h_X(t)) \text{ pour } t > \mathbb{E}(X) \quad \text{et} \quad \mathbb{P}(X \leq t) \leq \exp(-h_X(t)) \text{ pour } t < \mathbb{E}(X).$$

Ces inégalités améliorent considérablement l'inégalité de Markov. Cependant, elles ne sont valables que pour des lois admettant des moments de tous ordres.

Nous reviendrons sur l'utilisation de la transformée de Laplace pour une somme de v.a.r. dans la section 4.1.

3. LA NOTION D'INDÉPENDANCE

3.1 Indépendance d'événements et de classes d'événements

Le concept d'indépendance est l'une des spécificités majeures des probabilités par rapport à l'analyse.

Définition. *On dit que des événements $A_1, \cdots, A_n$ sont indépendants si pour tout $I \subset \{1, \cdots, n\}$, on a*

$$\mathbb{P}\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} \mathbb{P}(A_i).$$

Tout événement est donc indépendant de Ω et $\emptyset$. De plus, si A et B sont indépendants, il en va de même pour A et B^c. On trouvera bien d'autres propriétés de cette nature dans [FF1, p. 58].

Exemple. Le modèle probabiliste associé à l'expérience aléatoire de 2 lancers indépendants d'un dé équilibré est $(\Omega, \mathbb{P})$, avec $\Omega = \{1, \cdots, 6\}^2$ et $\mathbb{P}(\{\omega\}) = 1/6$ pour chaque éventualité ω. Les événements A : "le 1er lancer donne un nombre pair" et B : "le 2nd lancer donne un nombre impair" sont indépendants. Il est important de noter que l'indépendance des événements 2 à 2 n'est pas suffisante : si C est l'événement "les 2 lancers donnent des nombres de même parité", alors les 3 événements sont indépendants 2 à 2, mais A, B, C ne constitue pas un système d'événements indépendants, car $\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Le concept d'indépendance s'exporte naturellement à une suite infinie.

Définition. *Soit $(A_n)_{n \geq 1}$ une suite d'événements. On dit que les événements $A_1, A_2, \cdots$ sont indépendants si toute suite finie extraite de la suite $(A_n)_{n \geq 1}$ est constituée d'événements indépendants.*

Pour les classes d'événements, la notion d'indépendance est définie comme suit :

Définition. *Les classes d'événements $\mathscr{C}_1, \mathscr{C}_2, \cdots$ sont dites indépendantes si, pour tous $A_1 \in \mathscr{C}_1, A_2 \in \mathscr{C}_2, \cdots$, les événements $A_1, A_2, \cdots$ sont indépendants.*

En particulier, cette définition généralise la notion d'*indépendance de tribus*. Le résultat ci-dessous, sur lequel on serait tenté de jeter un coup d'oeil distrait, est d'utilité constante dans la manipulation des v.a. indépendantes :

Théorème `[FF1, p. 60]`. *Soient $\pi_1, \pi_2, \cdots$ des classes d'événements qui sont indépendantes. Si, pour chaque $i \geq 1$, π_i est stable par intersection finie, les tribus $\sigma(\pi_1), \sigma(\pi_2), \cdots$ sont indépendantes.*

3.2 INDÉPENDANCE DE VARIABLES ALÉATOIRES

L'idée est la suivante : deux v.a. sont indépendantes si les tribus qu'elles engendrent le sont. On se permet dans ce qui suit d'aborder d'emblée le cas d'une suite éventuellement infinie de v.a. Cela suppose que l'on puisse définir une notion de *mesure produit* sur $(\mathbb{R}^d)^{\mathbb{N}}$ par exemple, ce que nous admettrons ici. La construction d'une telle probabilité fait l'objet du *théorème de prolongement de Kolmogorov*, qui ne figure pas au programme de l'agrégation.

Définition. *Les v.a. $(X_n)_{n\geq 1}$ à valeurs dans $\mathbb{R}^d$ sont indépendantes si les tribus $\sigma(X_1), \sigma(X_2), \cdots$ sont indépendantes.*

On en déduit par un argument habituel de densité que les v.a. $(X_n)_{n\geq 1}$ à valeurs dans $\mathbb{R}^d$ sont indépendantes si, seulement si, pour tout $I \subset \mathbb{N}^\star$ fini et toute famille de fonctions mesurables bornées (ou positives) $(g_i)_{i\in I}$ définies sur $\mathbb{R}^d$ et à valeurs dans $\mathbb{R}$ (cf `[O, p. 42]`) :

$$\mathbb{E}\Big(\prod_{i\in I} g_i(X_i)\Big) = \prod_{i\in I} \mathbb{E}\big(g_i(X_i)\big).$$

L'indépendance entre v.a. est une propriété de la loi. Considérons le cas de 2 v.a. (extension immédiate au cas de n v.a.). Les v.a. X_1 et X_2 sont indépendantes (on notera cette propriété $X_1 \perp\!\!\!\perp X_2$) si, et seulement si, $\mathbb{P}_{(X_1,X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$ (cf `[FF1, p. 132]` ou `[O, p. 41]`). Selon la nature de la loi du couple (discrète ou à densité), on en déduit les critères d'indépendance suivants (cf `[O, p. 43]`) :

- ▷ Si $\forall i$, X_i possède une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, on a : $X_1 \perp\!\!\!\perp X_2 \Leftrightarrow$

$$f_{(X_1,X_2)}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \text{ pour Lebesgue-presque tout } (x_1,x_2);$$

- ▷ Si $\forall i$, X_i est à valeurs dans un ensemble discret D, on a : $X_1 \perp\!\!\!\perp X_2 \Leftrightarrow$

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2), \quad \forall x_1, x_2 \in D.$$

Remarquer ici l'usage de la virgule, qui remplace le symbole "$\cap$". Combiné au théorème de la loi image, on peut utiliser ces caractérisations de l'indépendance pour en déduire facilement la loi de certaines transformations de v.a. A titre d'exemple, si $X \perp\!\!\!\perp Y$ avec $X \sim Y \sim \mathscr{N}(0,1)$, X/Y suit la loi de Cauchy.

Par ailleurs, le dernier théorème de la section 3.1 montre que des v.a.r. X_1 et X_2 sont indépendantes dès que pour chaque $s,t \in \mathbb{R}$:

$$\mathbb{P}(X_1 \leq s, X_2 \leq t) = \mathbb{P}(X_1 \leq s)\mathbb{P}(X_2 \leq t).$$

Cette propriété s'étend facilement au cas d'une suite de v.a.r. De manière plus générale, on a un critère d'indépendance entre v.a. qui est relativement simple :

Théorème. *Soit $(X_i)_{i\geq 1}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$, et pour chaque $i \geq 1$, $\pi_i \subset \mathscr{B}(\mathbb{R}^d)$ une classe stable par intersection finie et telle que $\sigma(\pi_i) = \mathscr{B}(\mathbb{R}^d)$. Si les classes d'événements $X_1^{-1}(\pi_1), X_2^{-1}(\pi_2), \cdots$ sont indépendantes, alors les v.a. $(X_i)_{i\geq 1}$ sont indépendantes.*

Quelques exemples incontournables.

- ▷ Soient $X \perp\!\!\!\perp Y$, avec $X \sim \mathscr{N}(0,1)$ et $Y \sim \mathscr{N}(0,1)$. Notons (R,Θ) la représentation polaire du vecteur aléatoire $(X,Y)^t$. Alors, $R \perp\!\!\!\perp \Theta$, $R^2 \sim \mathscr{E}(1/2)$ et $\Theta \sim \mathscr{U}[0,2\pi]$ (cf `[FF1, p. 198]`).
- ▷ Soient $X \perp\!\!\!\perp Y$, avec $X \sim \mathscr{E}(\lambda_1)$ et $Y \sim \mathscr{E}(\lambda_2)$. Alors, $\inf(X,Y) \sim \mathscr{E}(\lambda_1 + \lambda_2)$. `[Utilisation de la f.r.]`
- ▷ Soient $X_1, \cdots, X_n$ des v.a.r. indépendantes et de même loi, de densité f par rapport à la mesure de Lebesgue, et de f.r. F. On note σ une permutation aléatoire de $\{1, \cdots, n\}$ (en fait la seule p.s., car $\mathbb{P}(X_i = X_j) = 0$ si $i \neq j$) telle que $X_{\sigma(1)} \leq \cdots \leq X_{\sigma(n)}$. Alors, la v.a. $X_{\sigma(k)}$ admet une densité, qui vaut (cf `[O, p. 79]`) :

$$nf(x) C_{n-1}^{k-1} F(x)^{k-1} \big(1 - F(x)\big)^{n-k}, \quad x \in \mathbb{R}.$$

  La suite $X_{\sigma(1)}, \cdots, X_{\sigma(n)}$ est la *statistique d'ordre* associée à la suite $X_1, \cdots, X_n$.
- ▷ Soient X, Y des v.a.r. avec $X \perp\!\!\!\perp Y$. On suppose que le vecteur aléatoire $M = (X,Y)^t$ possède une densité par rapport à la mesure de Lebesgue, et que cette densité est strictement positive en tout point. Alors, on a l'équivalence : $\exists \sigma > 0$, tel que $X \sim Y \sim \mathscr{N}(0,\sigma^2) \Leftrightarrow$ la densité de M est radiale (cf `[FF1, p. 157]`).
- ▷ Soient $(X_n)_{n\geq 1}$ des v.a.r. indépendantes et de même loi. On note $G_n = \text{card}\{i \in \{1, \cdots, n\} : X_i \leq 1\}$ et $D_n = n - G_n$. Soit $N \sim \mathscr{P}(k)$, avec $N \perp\!\!\!\perp (X_n)_{n\geq 1}$. Alors, $D_N \sim \mathscr{P}(k\mathbb{P}(X_1 > 1))$, $G_N \sim \mathscr{P}(k\mathbb{P}(X_1 \leq 1))$, et surtout $D_N \perp\!\!\!\perp G_N$. `[Une curiosité !]`

Mentionnons enfin un dernier résultat important (conséquence du théorème de Fubini) :

Théorème `[FF1, p. 132]`. *Soient $X_1, \cdots, X_n$ des v.a.r. intégrables. Alors :*

$$\mathbb{E}\Big[\prod_{i=1}^{n} X_i\Big] = \prod_{i=1}^{n} \mathbb{E}[X_i].$$

3.3 Indépendance et fonction caractéristique

Comme la loi d'une v.a. est caractérisée par sa f.c., il est naturel de voir l'indépendance, qui est une propriété de loi, via une propriété de la f.c.

Théorème `[O, p.205]`. *Soient $X_1, \cdots, X_n$ des v.a.r. Elles sont indépendantes si, et seulement si,*

$$\varphi_{(X_1,\cdots,X_n)}(t_1,\cdots,t_n) = \varphi_{X_1}(t_1)\cdots,\varphi_{X_n}(t_n), \quad \forall t_1,\cdots,t_n \in \mathbb{R}.$$

On montre facilement avec ce critère que si X et Y sont des v.a.r. avec $X \perp\!\!\!\perp Y$ et $X \sim Y \sim \mathscr{N}(0,1)$, alors, $(X+Y)/\sqrt{2}$ et $(X-Y)/\sqrt{2}$ sont indépendantes et de même loi $\mathscr{N}(0,1)$.

Si X et Y sont des v.a.r. indépendantes et de carrés intégrables, $\mathrm{cov}(X,Y) = 0$. Mais la réciproque est en général fausse (prendre $X \sim \mathscr{N}(0,1)$ et $Y = X^2$). Les 2 résultats qui suivent établissent des réciproques, dans des contextes particuliers (leurs preuves utilisent la f.c.) :

Théorème.
- ▷ `[O, p. 225]`. *Soient X, Y des v.a.r. bornées, i.e. il existe $C > 0$ tel que $\mathbb{P}(|X| \le C) = \mathbb{P}(|Y| \le C) = 1$. Alors, $X \perp\!\!\!\perp Y \Leftrightarrow$* $\mathrm{cov}(X^k, Y^\ell) = 0$, $\forall k, \ell \in \mathbb{N}$.
- ▷ `[O, p. 252]`. *Soit $X = (X_1, \cdots, X_d)^t$ un vecteur gaussien de $\mathbb{R}^d$. Alors, les composantes de X sont des v.a.r. indépendantes $\Leftrightarrow \mathbb{V}(X)$ est diagonale.*

Si X n'est pas un vecteur gaussien, le dernier résultat est faux. En effet, soient $Z \perp\!\!\!\perp \varepsilon$, avec $Z \sim \mathscr{N}(0,1)$ et $\mathbb{P}(\varepsilon = \pm 1) = 1/2$. Si $Y = \varepsilon Z \sim \mathscr{N}(0,1)$, le vecteur aléatoire $(Z,Y)^t$ n'est pas gaussien. Or, $\mathrm{cov}(Z,Y) = 0$, mais Z et Y ne sont pas indépendantes.

Citons enfin un dernier résultat remarquable, quoique anecdotique, dont la preuve utilise la f.c. La seule référence pour la preuve (`[O, p. 280]`) est beaucoup trop longue, et peut être considérablement simplifiée (cf fascicule d'exercice de M. Gradinaru pour le Master 1).

Théorème `[O, p. 280]`. *Soient X et Y des v.a.r. indépendantes et de même loi de carré intégrable. Alors, X et Y suivent des lois gaussiennes $\Leftrightarrow X+Y \perp\!\!\!\perp X-Y$.*

Arguments de la preuve. On peut supposer que les 2 v.a. sont *centrées* (i.e. d'espérance nulle) et *réduites* (i.e. de variance 1). Si X et Y suivent une loi gaussienne, un calcul de f.c. montre que $X+Y \perp\!\!\!\perp X-Y$. Supposons maintenant que $X+Y \perp\!\!\!\perp X-Y$, et notons φ la f.c. de X. Alors, $\varphi(2t) = \varphi(t)^3\varphi(-t)$. Puisque $\varphi(0) = 1$, on en déduit par l'absurde que φ ne s'annulle jamais. Par ailleurs, comme $X \in \mathbb{L}^2(\mathbb{P})$, on sait (dernier théorème de la section 2.1) que $\varphi(t) = 1 - t^2/2 + o(t^2)$ au voisinage de 0. Par suite, si $\psi(t) = \varphi(t)/\varphi(-t)$, on a $\psi(t) = 1 + o(t^2)$ au voisinage de 0. Comme $\psi(2t) = \psi(t)^2$ et $\psi(0) = 1$, cela entraîne que $\psi \equiv 1$, i.e. $\varphi(t) = \varphi(-t)$. Ainsi, φ est à valeurs réelles positives et vérifie $\varphi(t) = \varphi(t/2)^4$. Comme φ est de classe $\mathscr{C}^2$ avec $\varphi(0) = 1$, $\varphi'(0) = 0$ et $\varphi''(0) = -1$, on en déduit que $\varphi(t) = e^{-t^2/2}$. •

3.4 Le lemme de Borel-Cantelli

Pour une suite d'événements $(A_n)_n$, on note :

$$\limsup_n A_n = \bigcap_n \bigcup_{k \ge n} A_k \quad \text{et} \quad \liminf_n A_n = \bigcup_n \bigcap_{k \ge n} A_k.$$

L'ensemble $\limsup_n A_n$ contient les éventualités qui sont dans une infinité des A_n, alors que $\liminf_n A_n$ est l'ensemble des éventualités qui, à partir d'un certain rang, sont dans tous les A_n. Du coup, $\liminf_n A_n \subset \limsup_n A_n$.

Lemme de Borel-Cantelli `[O, p. 51]`. *Soit $(A_n)_n$ une suite d'événements.*

(i) $\sum_n \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(\limsup_n A_n) = 0$;

(ii) Si les événements $(A_n)_n$ sont indépendants, $\sum_n \mathbb{P}(A_n) = \infty \Rightarrow \mathbb{P}(\limsup_n A_n) = 1$.

Dans le cadre du lemme de Borel-Cantelli, il est difficile de ne pas évoquer les 2 applications qui suivent.

Jeu de pile ou face. On joue à pile ou face une infinité de fois. L'événement "N piles consécutifs" apparaît une infinité de fois p.s. d'après le lemme de Borel-Cantelli.

Marche aléatoire sur $\mathbb{Z}$. Soient $(X_i)_{i\geq 1}$ des v.a. indépendantes et de même loi telle que $\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = -1) = p$. On pose $S_0 = 0$ et $S_n = X_1 + \cdots + X_n$ pour $n \geq 1$. La suite de v.a. $(S_n)_{n\geq 1}$, ou *processus stochastique*, est la *marche aléatoire sur $\mathbb{Z}$ issue de 0*. D'après la formule de Stirling,

$$\mathbb{P}(S_{2n} = 0) = C_{2n}^n p^n (1-p)^n \sim \frac{1}{\sqrt{\pi n}} \big(4p(1-p)\big)^n.$$

On s'intéresse au nombre de retour de $(S_n)_{n\geq 0}$ en 0 (cf `[D, p. 188]`).

- ▷ Cas $p \neq 1/2$. D'après le lemme de Borel-Cantelli, $\mathbb{P}(\limsup_n \{S_n = 0\}) = 0$;
- ▷ Cas $p = 1/2$. Pour $m \geq 0$ et $k \geq 1$, notons $A_{m,k} = \{S_m = 0, S_n \neq 0 \forall n \geq m+k\}$. Par indépendance des accroissements S_m et $S_n - S_m$ si $n \geq m$:

$$\begin{aligned} \mathbb{P}(A_{m,k}) &= \mathbb{P}(S_m = 0, S_n - S_m \neq 0, \forall n \geq m+k) \\ &= \mathbb{P}(S_m = 0)\,\mathbb{P}(S_n - S_m \neq 0, \forall n \geq m+k) = \mathbb{P}(S_m = 0)\,\mathbb{P}(S_j \neq 0, \forall j \geq k). \end{aligned}$$

Une éventualité est dans au plus k événements $(A_{m,k})_{m\geq 0}$, donc $\sum_{m\geq 0} \mathbf{1}_{A_{m,k}} \leq k$. Par suite,

$$k \geq \sum_{m\geq 0} \mathbb{P}(A_{m,k}) \geq \sum_{m\geq 0} \mathbb{P}(S_m = 0)\,\mathbb{P}(S_j \neq 0, \forall j \geq k).$$

Comme $\sum_{m\geq 0} \mathbb{P}(S_m = 0) = \infty$, on a forcément $\mathbb{P}(S_j \neq 0, \forall j \geq k) = 0$, et ceci pour tout $k \geq 1$. On en déduit que $\mathbb{P}(\liminf_n \{S_n \neq 0\}) = 0$, soit $\mathbb{P}(\limsup_n \{S_n = 0\}) = 1$.

3.5 La loi du $0-1$ de Kolmogorov

Pour une suite de tribus $(\mathscr{A}_n)_{n\geq 1}$, on note $\mathscr{A}_\infty$ la *tribu asymptotique* qui lui est associée, i.e.

$$\mathscr{A}_\infty = \bigcap_{n\geq 1} \sigma\Big(\bigcup_{k\geq n} \mathscr{A}_k\Big).$$

Un événement fait partie de la tribu asymptotique s'il est dans chaque tribu $\sigma(\cup_{k\geq n}\mathscr{A}_k)$.

Loi du 0-1 de Kolmogorov `[O, p. 50]`. *Si les tribus $(\mathscr{A}_n)_{n\geq 1}$ sont indépendantes, alors $\forall A \in \mathscr{A}_\infty$: $\mathbb{P}(A) \in \{0,1\}$.*

Considérons maintenant le cas particulier suivant : soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes à valeurs dans $\mathbb{R}$, et $\mathscr{A}_k = \sigma(X_k)$. La tribu asymptotique vaut :

$$\mathscr{A}_\infty = \bigcap_{n\geq 1} \sigma(X_k, k \geq n).$$

Les v.a. $\limsup_n X_n$ et $\liminf_n X_n$, qui sont $\mathscr{A}_\infty$-mesurable, sont donc p.s. constantes. De plus, $\{(X_n)_{n\geq 1}$ converge dans $\mathbb{R}\} \in \mathscr{A}_\infty$. De nouveau, comme les $(X_n)_{n\geq 1}$ sont indépendantes, cet événement est de probabilité 0 ou 1 d'après la loi du 0-1. De même, $\{\sum_n X_n$ converge$\} \in \mathscr{A}_\infty$ est de probabilité 0 ou 1.

Séries entières à coefficients aléatoires. En poursuivant dans le même contexte, considérons la série entière à coefficients aléatoires $\sum_{n\geq 1} X_n z^n$. Son rayon de convergence R, qui est une v.a. $\mathscr{A}_\infty$-mesurable d'après la règle d'Hadamard, est donc p.s. constant. Si $X_1 \in \mathbb{L}^1(\mathbb{P})$, $R \geq 1$ p.s. car, par convergence monotone :

$$\mathbb{E}\Big(\Big|\sum_{n\geq 1} X_n z^n\Big|\Big) \leq \mathbb{E}(|X_1|) \sum_{n\geq 1} |z|^n.$$

De plus, si $\mathbb{P}(X_1 = 0) < 1$, il existe $\alpha > 0$ tel que $\sum_n \mathbb{P}(|X_n| \geq \alpha) = \infty$. Comme les v.a. $(X_n)_n$ sont indépendantes, on en déduit du lemme de Borel-Cantelli que $\mathbb{P}(\limsup_n \{|X_n| \geq \alpha\}) = 1$. Du coup, presque sûrement, il existe une suite $(n_k)_k$ telle que $|X_{n_k}| \geq \alpha$ pour chaque k. On utilise maintenant la règle d'Hadamard :

$$R^{-1} = \limsup_n |X_n|^{1/n} \geq \lim_k \alpha^{1/n_k} = 1,$$

pour en déduire que $R \leq 1$. Sous les hypothèses mentionnées jusqu'ici, le rayon de convergence de la série entière à coefficients aléatoires vaut 1. On peut même montrer qu'il n'y a pas convergence sur le bord du disque de convergence.

4. SOMMES DE VARIABLES ALÉATOIRES INDÉPENDANTES

4.1 QUELQUES EXEMPLES DE SOMMES DE VARIABLES ALÉATOIRES INDÉPENDANTES

Soient $X_1, \cdots, X_n$ des v.a.r. indépendantes, et M_n leur somme : $M_n = X_1 + \cdots + X_n$. Si les v.a. sont de carré intégrables, alors, du fait de leur indépendance, $\mathrm{cov}(X_i, X_j) = 0$ si $i \neq j$ et donc :

$$\mathrm{var}(M_n) = \sum_{i=1}^{n} \mathrm{var}(X_i).$$

L'utilisation de la f.c. permet souvent de trouver facilement la loi de la somme de v.a. indépendantes, dont les lois font partie d'une même famille. Dans les cas standards, on peut ainsi retrouver :

▷ Si $\forall i, X_i \sim B(n_i, p)$, $M_n \sim \mathscr{B}(\sum_{i=1}^n n_i, p)$;
▷ Si $\forall i, X_i \sim \mathscr{P}(\lambda_i)$, $M_n \sim \mathscr{P}(\sum_{i=1}^n \lambda_i)$;
▷ Si $\forall i, X_i \sim \mathscr{N}(m_i, \sigma_i^2)$, $M_n \sim \mathscr{N}(\sum_{i=1}^n m_i, \sum_{i=1}^n \sigma_i^2)$;
▷ Si $\forall i, X_i \sim \gamma(a_i, \lambda)$, $M_n \sim \gamma(\sum_{i=1}^n a_i, \lambda)$.

Identité de Wald `[FF2, p. 23]`. *Soient $(X_i)_{i\geq 1}$ des v.a. indépendantes et de même loi intégrable, et S une v.a. intégrable à valeurs dans $\mathbb{N}^\star$ telle que pour chaque $k \geq 1$: $\{S = k\} \in \sigma(X_1, \cdots, X_k)$ (temps d'arrêt). Alors,*

$$\mathbb{E}\Big(\sum_{i=1}^{S} X_i\Big) = \mathbb{E}(X_1)\,\mathbb{E}(S).$$

Ruine du joueur. Soient $(\xi_i)_{i\geq 1}$ des v.a.r. indépendantes, et de même loi de *Rademacher*, i.e. $\mathbb{P}(\xi_1 = \pm 1) = 1/2$. On note $G_0 = 0$ et pour $n \geq 1$, $G_n = \xi_1 + \cdots + \xi_n$. Comme $(\xi_1 + 1)/2 \sim \mathscr{B}(1/2)$, on a $(G_n + n)/2 \sim \mathscr{B}(n, 1/2)$. La suite de v.a. $(G_n)_{n\geq 0}$, appelée *marche aléatoire simple symétrique sur* $\mathbb{Z}$, modélise le gain d'un joueur qui affronte une banque. En effet, G_n représente le *gain* d'un joueur à l'issue de la n-ième partie de pile ou face, selon la règle : il met en jeu a euros au jeu de pile ou face ; il perd 1 euro à chaque pile, et il en gagne 1 à chaque face ; s'il arrive à b euros (censé représenter la fortune totale de la banque), il a gagné le jeu, et s'il arrive à $-a$ euro, c'est une défaite. La v.a. $T = \inf\{n \geq 0 : G_n \in \{-a, b\}\}$ représente le numéro de la dernière partie du jeu. Toute la question est de savoir si $\mathbb{P}(T < \infty) = 1$, i.e. le jeu s'arrête-t-il ? Dans l'affirmative, il faut calculer la loi de la v.a. G_T, afin de répondre à la question : quelle est la probabilité que le joueur gagne ?

▷ La réponse à la 1ère question est affirmative, car pour chaque $k \in \mathbb{Z}$, $\mathbb{P}(G_n = k) \to 0$ lorsque $n \to \infty$ (formule de Stirling) et pour chaque $n \geq 1$, $\{T = \infty\} \subset \{G_n \in \{-a+1, \cdots, b-1\}\}$. Une autre preuve sera apportée avec le *théorème central limite* (cf section 6.2). Noter que, en particulier, $\mathbb{P}(G_T \in \{-a, b\}) = 1$.
▷ Pour la 2ème question, on remarque que pour tous $n, k \geq 1$, $\{\inf(T, n) = k\} \in \sigma(\xi_1, \cdots, \xi_k)$. Donc, d'après l'identité de Wald, $\mathbb{E}(G_{\inf(n,T)}) = 0$ pour chaque $n \geq 1$. Comme $G_{\inf(n,T)} \to G_T$ p.s. et $|G_{\inf(n,T)}| \leq \max(a, b)$, on en déduit avec le théorème de Lebesgue que $\mathbb{E}(G_T) = 0$. Par suite, $\mathbb{P}(G_T = -a) = b/(a+b)$.

4.2 UTILISATION DE LA TRANSFORMÉE DE LAPLACE POUR LES SOMMES DE VARIABLES ALÉATOIRES INDÉPENDANTES

Les résultats ci-dessous prolongent le théorème principal de la section 2.2. Ce sont aussi des exemples d'utilisation de la transformée de Laplace en probabilités.

Théorème de grandes déviations `[DD2, p. 174]`. *Soient $(X_i)_{i\geq 1}$ des v.a.r. indépendantes et de même loi, et $S_n = X_1 + \cdots + X_n$. On suppose que l'intérieur de I_{X_1} est un voisinage de 0. Soit $\varepsilon > 0$. Alors, il existe $a > 0$ tel que pour chaque $n \geq 1$:*

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}(X_1)\Big| \geq \varepsilon\Big) \leq e^{-na}.$$

La moyenne de Césarro S_n/n converge donc -en un sens que nous préciserons (cf section 5)- vers $\mathbb{E}(X_1)$ à vitesse exponentielle, lorsque $n \to \infty$.

Théorème de Chernoff `[DD2, p. 174]`. *Soient $(X_i)_{i\geq 1}$ des v.a.r. indépendantes et de même loi, et $S_n = X_1 + \cdots + X_n$. On suppose que l'intérieur de I_{X_1} est un voisinage de 0. Notons Ψ la dérivée de $\ln \mathscr{L}_{X_1}$, $\alpha = \inf\{\Psi(t), t \in I_{X_1}\}$ et $\beta = \inf\{\Psi(t), t \in I_{X_1}\}$. Alors,*

(i) $(1/n)\ln\mathbb{P}(S_n \geq nt) \to -h_{X_1}(t)$ *si* $\beta > t > \mathbb{E}(X_1)$ *;*
(ii) $(1/n)\ln\mathbb{P}(S_n \leq nt) \to -h_{X_1}(t)$ *si* $\alpha < t < \mathbb{E}(X_1)$*.*

Comme $h_{X_1}(t) > 0$ si $t \neq \mathbb{E}(X_1)$, le théorème de Chernoff nous donne des évaluations très fines de la déviation de la moyenne S_n/n par rapport à sa moyenne $\mathbb{E}(X_1)$. Si l'on compare au résultat donné par l'inégalité de Bienaymé-Tchebychev, on a pour chaque $\varepsilon > 0$, en notant $t = \varepsilon + \mathbb{E}(X_1)$:

$$\mathbb{P}(S_n \geq nt) \leq \mathbb{P}(|S_n - n\mathbb{E}(X_1)| \geq n\varepsilon) \leq \frac{\mathrm{var}(X_1)}{\varepsilon^2}.$$

On peut donc juste montrer que $\limsup_n (1/n)\ln\mathbb{P}(S_n \geq nt) \leq 0$. Sans commentaire ...

4.3 Loi de la somme de variables aléatoires indépendantes et produit de convolution

On rappelle que le *produit de convolution* de 2 mesures de probabilité μ et ν sur $\mathbb{R}^d$, noté $\mu \star \nu$, est défini par la propriété :

$$\int_{\mathbb{R}^d} f\, d(\mu \star \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x+y)\, \mu \otimes \nu(dx, dy),$$

pour chaque fonction $f : \mathbb{R}^d \to \mathbb{R}$ borélienne positive (cf [O, p. 63]). Noter que $\mu \star \nu$ est une probabilité. Par ailleurs, l'opération $\star$ est commutative, et la mesure de Dirac en 0 est élément neutre.

Pour 2 v.a. X et Y indépendantes et à valeurs dans $\mathbb{R}^d$, on vérifie que $\hat{\mathbb{P}}_{X+Y} = \varphi_{X+Y} = \varphi_X \varphi_Y = \hat{\mathbb{P}}_X \hat{\mathbb{P}}_Y$. Par ailleurs, cette dernière quantité est aussi la transformée de Fourier de la loi $\mathbb{P}_X \star \mathbb{P}_Y$. Par injectivité, on en déduit le résultat :

Théorème [O, p. 63]. *Soient X, Y des v.a. indépendantes à valeurs dans $\mathbb{R}^d$. Alors, $\mathbb{P}_{X+Y} = \mathbb{P}_X \star \mathbb{P}_Y$.*

En particulier, si la loi de X est *diffuse*, i.e. $\mathbb{P}_X$ ne charge pas les points, alors la loi de $X+Y$ est diffuse. En effet, pour chaque $t \in \mathbb{R}^d$, d'après le théorème de Fubini-Tonelli :

$$\mathbb{P}(X+Y=t) = \int_{\mathbb{R}^d} \mathbf{1}_{\{t\}}\, d(\mathbb{P}_X \star \mathbb{P}_Y) = \int_{\mathbb{R}^d} \mathbb{P}_X(\{t-y\})\mathbb{P}_Y(dy) = 0.$$

La loi de la somme de 2 v.a. indépendantes étant représentée par le produit de convolution des lois, on peut donc écrire, d'après les résultats de la section 4.1, $\mathscr{P}(\lambda_1) \star \mathscr{P}(\lambda_2) = \mathscr{P}(\lambda_1 + \lambda_2)$, $\mathscr{N}(m_1, \sigma_1^2) \star \mathscr{N}(m_2, \sigma_2^2) = \mathscr{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$, $\mathscr{B}(n_1, p) \star \mathscr{B}(n_2, p) = \mathscr{B}(n_1 + n_2, p)$ ...

En pratique, le produit de convolution entre mesures ne se prête pas très bien au calcul de la loi de la somme de v.a. indépendantes. On lui préférera souvent un calcul basé sur la notion de f.c., sauf éventuellement lorsque les v.a. possèdent des densités par rapport à la mesure de Lebesgue, ou si elles ont des lois discrètes. Dans ces cas, le produit de convolution prend une forme simple, par exemple :

Théorème [O, p. 64]. *Soient X et Y des v.a. indépendantes à valeurs dans $\mathbb{R}^d$. Si elles possèdent chacune une densité par rapport à la mesure de Lebesgue, alors $X+Y$ possède une densité par rapport à la mesure de Lebesgue, et*

$$f_{X+Y}(z) = f_X \star f_Y(z) = \int_{\mathbb{R}^d} f_X(x) f_Y(z-x)\,dx = \int_{\mathbb{R}^d} f_Y(x) f_X(z-x)\,dx.$$

5. CONVERGENCES STOCHASTIQUES

5.1 Les différentes notions de convergence

Définition. *Soient $(X_n)_n$, X des v.a. à valeurs dans $\mathbb{R}^d$, et $\|.\|$ une norme sur $\mathbb{R}^d$. Alors,*

(1). $(X_n)_n$ converge p.s. vers X si $\mathbb{P}(X_n \to X) = 1$. On note $X_n \overset{p.s.}{\to} X$;

(2). $(X_n)_n$ converge en probabilité vers X si $\mathbb{P}(\|X_n - X\| \geq \varepsilon) \to 0$, $\forall \varepsilon > 0$. On note $X_n \overset{\mathbb{P}}{\to} X$;

(3). $(X_n)_n$ converge dans $\mathbb{L}^p(\mathbb{P})$ vers X si $\mathbb{E}(\|X_n - X\|^p) \to 0$. On note $X_n \overset{\mathbb{L}^p}{\to} X$;

(4). $(X_n)_n$ converge en loi vers X si $\int_{\mathbb{R}^d} f\, d\mathbb{P}_{X_n} \to \int_{\mathbb{R}^d} f\, d\mathbb{P}_X$, $\forall f : \mathbb{R}^d \to \mathbb{R}$ continue bornée. On note $X_n \overset{\mathscr{L}}{\to} X$.

On trouvera dans [FF1, chap. 16] les relations entre ces différents modes de convergence, ainsi que des contre-exemples. Mentionnons seulement que les faits suivants :

- ▷ La convergence p.s. et la convergence $\mathbb{L}^p(\mathbb{P})$ entraînent la convergence en probabilité, et cette dernière entraîne la convergence en loi ;
- ▷ Une suite converge en probabilité si, et seulement si, de toute suite d'entiers, on peut extraire une sous-suite pour laquelle la convergence a lieu p.s. Contrairement aux autres notions de convergence, la convergence p.s. n'est donc pas associée à une métrique ;
- ▷ Si une suite converge en loi vers une v.a. *dégénérée*, i.e. une v.a. de loi de Dirac, alors la convergence a lieu en probabilité.

Par ailleurs, les convergences p.s., en probabilité et en loi sont stables sous l'action des fonctions continues : si $X_n \to X$ p.s. (resp. en loi, en probabilité), et si f est une fonction continue, alors $f(X_n) \to f(X)$ p.s. (resp. en loi, en probabilité). En revanche, la convergence $\mathbb{L}^p(\mathbb{P})$ ne respecte pas ce principe.

Dans les espaces de probabilité, le théorème de Lebesgue admet une version sous des conditions plus faibles :

Théorème de Lebesgue. *Soient $(X_n)_n$ et X des v.a. à valeurs dans $\mathbb{R}^d$ telles que $X_n \overset{P}{\to} X$ et $\sup_n \|X_n\| \in \mathbb{L}^1(\mathbb{P})$. Alors, $X_n \overset{\mathbb{L}^1}{\to} X$.*

5.2 La convergence presque sûre

La convergence p.s., n'étant pas associée à une métrique, est sans doute la plus obscure. Pour illustrer ceci, considérons une suite d'événements indépendants $(A_n)_n$, tel que $\mathbb{P}(A_n) = 1/n$. On a $\mathbf{1}_{A_n} \overset{\mathbb{P}}{\to} 0$, mais $(\mathbf{1}_{A_n})_n$ ne converge pas p.s. En effet, pour chaque $\varepsilon \in]0,1]$, la série de terme général $\mathbb{P}(\mathbf{1}_{A_n} \geq \varepsilon)$ diverge et donc $\mathbb{P}(\limsup_n \{\mathbf{1}_{A_n} \geq \varepsilon\}) = 1$ d'après le lemme de Borel-Cantelli. En clair, $\mathbb{P}$-presque toute éventualité est dans une infinité des événements $(\{\mathbf{1}_{A_n} \geq \varepsilon\})_n$, ce qui montre que $(\mathbf{1}_{A_n})_n$ ne converge pas p.s. vers 0.

Le lemme de Borel-Cantelli nous donne un critère simple de convergence p.s. Ce critère nous montre que, d'une certaine manière, la convergence p.s. correspond à une convergence en probabilité "suffisamment rapide".

Théorème [FF1, p. 211]. *Soient $(X_n)_n$, X des v.a. à valeurs dans $\mathbb{R}^d$. Alors, $X_n \overset{p.s.}{\to} X$ si, pour chaque $\varepsilon > 0$, la série de terme général $\mathbb{P}(\|X_n - X\| \geq \varepsilon)$ est convergente.*

Si $(A_n)_n$ est une suite d'événements tels que $\mathbb{P}(A_n) = 1/n^2$, on a donc $\mathbf{1}_{A_n} \overset{p.s.}{\to} 0$.

Par ailleurs, dans le contexte du théorème de grandes déviations de la section 4.1, on a $S_n/n \overset{p.s.}{\to} \mathbb{E}(X_1)$. C'est une forme "forte" de la loi des grands nombres, la convergence ayant lieu p.s. Cependant, le contexte est assez restrictif, puisque X_1 admet des moments de tous ordres.

5.3 La convergence en probabilité

On peut donner une forme "faible" à la loi des grands nombres, mais sous des conditions plus générales. Soit $(X_i)_{i\geq 1}$ une suite de v.a.r. indépendantes et de même loi, et $S_n = X_1 + \cdots + X_n$. Si $X_1 \in \mathbb{L}^2(\mathbb{P})$, alors d'après l'inégalité de Bienaymé-Tchebychev, on a pour chaque $\varepsilon > 0$:

$$\mathbb{P}\Big(\Big|\frac{S_n}{n} - \mathbb{E}(X_1)\Big| \geq \varepsilon\Big) \leq \frac{\mathrm{var}(S_n/n)}{\varepsilon^2} = \frac{\mathrm{var}(X_1)}{n\varepsilon^2}.$$

Par conséquent, $S_n/n \overset{\mathbb{P}}{\to} \mathbb{E}(X_1)$: c'est la *loi faible des grands nombres*.

Une preuve probabiliste du théorème de Weierstrass. Poursuivons avec les mêmes notations. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. Fixons $x \in [0,1]$, et supposons que $X_1 \sim \mathscr{B}(x)$. Puisque $\mathbb{E}(X_1) = x$ et f est continue bornée, $\mathbb{E}(f(S_n/n)) \to f(x)$ d'après le théorème de Lebesgue et la loi faible des grands nombres. Comme $S_n \sim \mathscr{B}(n,x)$, on a

$$\mathbb{E}\Big(f\Big(\frac{S_n}{n}\Big)\Big) = \sum_{k=0}^{n} C_n^k f\Big(\frac{k}{n}\Big) x^k (1-x)^{n-k},$$

c'est-à-dire que f a été approchée ponctuellement par une suite de polynômes. On peut montrer avec un peu plus de travail que la convergence est en fait uniforme (cf `[FF1, p. 234]`) : cette méthode constitue une démonstration efficace du *théorème de Weierstrass*.

5.4 La convergence en loi

Tout d'abord, le cas où les v.a. sont discrètes :

Théorème `[O, p. 317]`. *Soient $(X_n)_n$, X des v.a. à valeurs dans $\mathbb{Z}^d$. Alors : $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)\ \forall k \in \mathbb{Z}^d \Leftrightarrow X_n \xrightarrow{\mathscr{L}} X$.*

Il peut être utile de disposer de la version de la convergence en loi ci-dessous lorsque l'on a affaire à des suites de v.a. qui sont définies par des sup ou des inf, car le calcul de la f.r. est souvent simple :

Théorème `[O, p. 318]`. *Soient $(X_n)_n$, X des v.a.r. Alors, $F_{X_n} \to F_X$ en tout point de continuité de $F_X \Leftrightarrow X_n \xrightarrow{\mathscr{L}} X$.*

Soit $(X_i)_{i\geq 1}$ une suite de v.a. indépendantes et de même loi $\mathscr{E}(1)$. On note $M_n = \sup\{X_i,\, i \in \{1,\cdots,n\}\}$. Pour chaque $x \in \mathbb{R}$:

$$\mathbb{P}(M_n - \ln n \leq x) = \big(F_{X_1}(x + \ln n)\big)^n = \Big(1 - \frac{1}{n}e^{-x}\Big)^n \to \exp(-e^{-x}).$$

Comme la fonction limite est une f.r., elle est associée à une loi (on peut noter aussi que cette loi est à densité par rapport à la mesure de Lebesgue). La suite $(M_n - \ln n)_n$ converge donc en loi.

Le résultat ci-dessous est particulièrement adapté à des suites de v.a. qui sont définies par des sommes de v.a. indépendantes :

Théorème de P. Lévy `[FF1, p. 219]`. *Soient $(X_n)_n$, X des v.a. à valeurs dans $\mathbb{R}^d$. Alors : $\varphi_{X_n} \to \varphi_X$ ponctuellement $\Leftrightarrow X_n \xrightarrow{\mathscr{L}} X$.*

Un simple calcul de f.c. montre alors que $\mathscr{B}(n,p_n) \xrightarrow{\mathscr{L}} \mathscr{P}(\lambda)$, si $np_n \to \lambda$. Noter ici l'abus de notation consistant à remplacer les v.a. par leurs lois. Plus généralement, un phénomène aléatoire qui peut se représenter comme une superposition d'événements rares et indépendants suit approximativement une loi de Poisson :

Théorème des événements rares `[O, p. 321]`. *Soit, pour chaque n, $\{A_{n,j},\, j \in M_n\}$ une famille finie d'événements indépendants. On pose $p_{n,j} = \mathbb{P}(A_{n,j})$, et on note*

$$S_n = \sum_{j \in M_n} \mathbf{1}_{A_{n,j}}.$$

On suppose que $M_n \nearrow \infty$, $\max_{j\in M_n} p_{n,j} \to 0$ et $\sum_{j\in M_n} p_{n,j} \to \lambda$. Alors, $X_n \xrightarrow{\mathscr{L}} \mathscr{P}(\lambda)$.

Noter l'usage de la notation hybride "$X_n \xrightarrow{\mathscr{L}} \mathscr{P}(\lambda)$".

Le "problème des rencontres" constitue un autre exemple où la loi de Poisson intervient en tant que limite d'une suite de v.a. La preuve présentée dans `[FF1, p. 259]` utilise le théorème de P. Lévy, et donc la f.c.

5.5 Convergence de sommes et de produits de variables aléatoires

Soient $(X_n)_n$, $(Y_n)_n$, X et Y des v.a.r. Si $X_n \to X$ et $Y_n \to Y$ p.s. ou en probabilité, alors la suite de vecteurs aléatoires $(X_n,Y_n)^t$ converge vers $(X,Y)^t$ selon le même mode de convergence. Comme ces 2 formes de convergence sont stables sous l'action des fonctions continues, on peut en déduire des résultats du type $X_n + Y_n \to X + Y$ p.s. ou en probabilité.

Cependant, ce n'est plus vrai avec la convergence en loi : le fait que les suites $(X_n)_n$ et $(Y_n)_n$ convergent en loi vers X et Y n'entraîne pas que la suite des vecteurs aléatoires $(X_n,Y_n)^t$ converge (sauf si $X_n \perp\!\!\!\perp Y_n$) : la raison principale est que la connaissance des lois *marginales* de X_n et Y_n ne permet en général pas de retrouver la loi du vecteur aléatoire $(X_n, Y_n)^t$. Heureusement, des résultats partiels existent quand même dans ce sens :

Lemme de Slutsky `[0, p. 347]`. *Soient* $(X_n)_n$, $(Y_n)_n$, X *des v.a.r et* $y \in \mathbb{R}$. *Si* $X_n \xrightarrow{\mathscr{L}} X$ *et* $Y_n \xrightarrow{P} y$, *alors* $X_n + Y_n \xrightarrow{\mathscr{L}} X + y$ *et* $X_n Y_n \xrightarrow{\mathscr{L}} Xy$.

A titre d'exemple, considérons le cas d'une suite de v.a.r. $(X_n)_n$ telle que, pour une suite croissante $(v_n)_n$, on ait $v_n X_n \xrightarrow{\mathscr{L}} X$. Observons que $X_n \xrightarrow{P} 0$. Soit maintenant $f : \mathbb{R} \to \mathbb{R}$ une fonction borélienne nulle en 0, et dérivable en 0. Alors, d'après le lemme de Slutsky :

$$v_n f(X_n) = \frac{f(X_n)}{X_n} v_n X_n \xrightarrow{\mathscr{L}} f'(0)X.$$

Cette remarque porte le nom de δ*-méthode*.

6. LA LOI FORTE DES GRANDS NOMBRES ET LE THÉORÈME DE LA LIMITE CENTRALE

6.1 LA LOI FORTE DES GRANDS NOMBRES

Ainsi que nous l'avons déjà vu dans les sections 5.2 et 5.3, le titre de "loi des grands nombres" est en général conféré à un théorème limite portant sur une moyenne de Césarro de v.a. On y ajoute le qualificatif de "fort" lorsque la convergence à lieu p.s. (cf section 5.2).

Il existe de nombreuses formes de la loi forte des grands nombres (cf `[0, chap. 10]`). Nous mentionnons ici la forme la plus habituelle. Dorénavant, pour une suite $(X_i)_{i\geq 1}$ de v.a., on note

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

sa moyenne de Césarro, ou *moyenne empirique*, si l'on considère, par abus, $X_1, \cdots, X_n$ comme des *observations* d'un phénomène aléatoire.

Loi forte des grands nombres de Kolmogorov-Khintchine `[0, p. 113]`. *Soit* $(X_i)_{i\geq 1}$ *une suite de v.a. à valeurs dans* $\mathbb{R}^d$, *supposées indépendantes et de même loi. Les 2 assertions suivantes sont équivalentes :*

(i) il existe $m \in \mathbb{R}$ *tel que* $\bar{X}_n \xrightarrow{p.s.} m$ *;*

(ii) $X_1 \in \mathbb{L}^1(\mathbb{P})$.

Si (ii) est vérifiée, $m = \mathbb{E}(X_1)$ *et la convergence à lieu dans* $\mathbb{L}^1(\mathbb{P})$.

On retrouve en statistique une application importante de la loi forte des grands nombres. Supposons que l'on ait observé une réalisation $(x_1, \cdots, x_n)$ de $(X_1, \cdots, X_n)$, i.e. pour une éventualité ω, $x_i = X_i(\omega)$. Une question fondamentale de la statistique est : comment retrouver la loi, supposée commune, des v.a. à partir des observations $x_1, \cdots, x_n$? Placons-nous dans le cas de v.a.r., afin de faire intervenir leur f.r. commune notée F. Si l'on a aucune idée de la loi de ces v.a.r., on peut chercher à l'estimer avec la *fonction de répartition empirique* F_n, définie pour chaque $x \in \mathbb{R}$ par :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}.$$

La loi des grands nombres suggère que p.s. cette suite de fonctions converge vers F. En effet, d'après le théorème de Kolmogorov-Khintchine, $F_n(x) \xrightarrow{p.s.} F(x)$. Cependant, l'ensemble de probabilité nulle sur lequel cette convergence n'a pas lieu dépend de x, et x est dans un ensemble non dénombrable. On ne peut donc pas conclure directement à la convergence p.s. de F_n vers F, et ceci quelque soit le mode de convergence. C'est l'objet du résultat ci-dessous :

Théorème de Glivenko-Cantelli `[0, p. 116]`. *Avec les notations ci-dessus, on a :*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s.} 0.$$

Puisque la convergence ponctuelle des f.r. est un critère de convergence des lois associées, on déduit du théorème de Glivenko-Cantelli que :

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \xrightarrow{\mathscr{L}} \mathbb{P}_{X_1}\Big) = 1.$$

C'est la version unidimensionnelle du *théorème fondamental de la Statistique*.

Test de Kolmogorov-Smirnov. La *statistique* issue du théorème de Glivenko-Cantelli, $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)|$, est *libre*, i.e. sa loi ne dépend pas de la loi de X_1 (cf `[O, p. 119]`). En effet, dans le cas simple où F est inversible :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{F(X_i)\leq F(x)\}} - F(x) \right| = \sup_{y\in[0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{F(X_i)\leq y\}} - y \right|.$$

Comme $F(X_1) \sim \mathscr{U}[0,1]$ (cf section 2.2), la loi est libre : c'est la loi de *Kolmogorov-Smirnov*. Cette remarque est à l'origine du *test de Kolmogorov-Smirnov* : on choisit une fonction F qui semble représentative de la loi des observations ; si la statistique de Glivenko-Cantelli, évaluée sur les observations, est dans une zone de forte probabilité pour la loi de Kolmogorov-Smirnov, on décide que F est bien la f.r. des v.a.

6.2 Le théorème central limite

Le théorème central limite vient en renfort de la loi des grands nombres, en ce sens qu'il en précise la vitesse de convergence. Comme pour la loi des grands nombres, on en trouve de nombreuses versions. Les preuves reposent le plus souvent sur le théorème de P. Lévy (cf section 5.4).

Théorème central limite `[O, p. 325]`. *Soit $(X_i)_{i\geq 1}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$, supposées indépendantes et de même loi de carré intégrable. Alors,*

$$\sqrt{n}\big(\bar{X}_n - \mathbb{E}(X_1)\big) \xrightarrow{\mathscr{L}} N_d(0, \mathbb{V}(X_1)).$$

Il est essentiel de comprendre pourquoi ce théorème est "central" : sous réserve que les v.a. soient de carré intégrables et indépendantes, la loi de leur somme est proche d'une loi gaussienne. C'est ce *principe d'invariance* qui confère à ce théorème un caractère central. L'autre appellation de loi "normale" pour désigner une loi gaussienne est donc amplement méritée ! La loi limite étant libre, les applications de ce théorème sont souvent de nature statistique, par exemple le théorème de Pearson (cf `[O, p. 326]`). Mentionnons ici d'autres types d'applications.

Calcul approché d'intégrales. On cherche une valeur approchée de

$$I = \int_{\mathbb{R}^d} g(x) f(x) dx,$$

avec $g, f : \mathbb{R}^d \to \mathbb{R}$ de produit Lebesgue-intégrable, et f une densité de probabilité. Un générateur aléatoire peut calculer une réalisation d'une suite de v.a. indépendantes distribuées selon la densité f. Notons $(X_i)_{i\geq 1}$ les v.a. dont ces réalisations sont issues. On peut donc calculer une réalisation de la moyenne de Césarro :

$$I_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i).$$

La loi forte des grands nombres nous donne $I_n \xrightarrow{p.s.} \mathbb{E}(g(X_1)) = I$. Si la réalisation est dans l'ensemble de probabilité 1 dans laquelle cette convergence à lieu (ce à quoi on peut légitimement s'attendre ...), on a ainsi une approximation de I. Cependant, ce résultat ne nous donne aucune information sur la précision de l'approximation. Ici intervient le théorème central limite. Supposons que $g(X_1) \in \mathbb{L}^2(\mathbb{P})$. Alors, $\sqrt{n}(I_n - I) \xrightarrow{\mathscr{L}} \mathscr{N}(0, \mathrm{var}(g(X_1)))$. En utilisant la caractérisation de la convergence en loi via la f.r. (cf section 5.4), cela nous donne, pour chaque $x \geq 0$:

$$\mathbb{P}\Big(I \in \Big[I_n - x\sqrt{\frac{\mathrm{var}(g(X_1))}{n}}, I_n + x\sqrt{\frac{\mathrm{var}(g(X_1))}{n}}\Big]\Big) \to \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-x^2/2} dx.$$

Choisissons pour x une valeur de sorte que le terme de droite soit égal à α (proche de 1). L'intervalle

$$\Big[I_n - x\sqrt{\frac{\mathrm{var}(g(X_1))}{n}}, I_n + x\sqrt{\frac{\mathrm{var}(g(X_1))}{n}}\Big]$$

est un *intervalle de confiance asymptotique au niveau* α pour la valeur de I, en ce sens que I a (asymptotiquement) α chances de se trouver dans cet intervalle. L'inconvénient majeur de cette méthode est que l'intervalle dépend de la valeur de $\mathrm{var}(g(X_1))$, qu'on ne sait a priori pas calculer. Son avantage est que la taille de l'intervalle de confiance ne dépend pas de la dimension.

Ruine du joueur (suite de la section 4.1). Pour chaque n, on a si $s=\max(a,b)$:

$$\mathbb{P}(T=\infty)=\mathbb{P}\Big(\bigcap_n \big\{G_n \in \{-a+1,\cdots,b-1\}\big\}\Big) \le \mathbb{P}\Big(\Big|\frac{G_n}{\sqrt{n}}\Big| \le \frac{s}{\sqrt{n}}\Big) \le \mathbb{P}\Big(\Big|\frac{G_n}{\sqrt{n}}\Big| \le \frac{s}{\sqrt{k}}\Big), \quad \forall k \le n.$$

Lorsque $n \to \infty$, le terme de droite tend vers $\mathbb{P}(|\mathcal{N}(0,1)| \le s/\sqrt{k})$. En faisant ensuite tendre k vers ∞, on obtient $\mathbb{P}(T=\infty)=0$, i.e. p.s. le jeu a un nombre fini de tours.

BIBLIOGRAPHIE

- [D] - R. Durrett (1996) - *Probability : Theory and Examples - Second Edition*, Duxbury Press.
- [DD1] - D. Dacunha-Castelle & M. Duflo (1994) - *Probabilités et Statistiques, 1. Problèmes à temps fixe - 2ème édition*, Masson.
- [DD2] - D. Dacunha-Castelle & M. Duflo (1993) - *Probabilités et Statistiques, 2. Problèmes à temps mobile - 2ème édition*, Masson.
- [FF1] - D. Foata & A. Fuchs (1998) - *Calcul des probabilités, Cours, exercices et problèmes corrigés*, Dunod.
- [FF2] - D. Foata & A. Fuchs (2002) - *Processus stochastiques - Processus de Poisson, chaînes de Markov et martingales, Cours, exercices et problèmes corrigés*, Dunod.
- [O] - J.-Y. Ouvrard (2000) - *Probabilités 2*, Cassini.