

Probabilités générales

Magalie Fromont

Références :

- *Calcul des probabilités*, D. Foata et A. Fuchs,
- *Probability and random processes*, G. Grimmett and D. Stirzaker,
- *Cours et exercices de probabilités appliquées*, M. Lefebvre,
- *Probabilités, analyse des données et statistique*, G. Saporta,
- *Théorie des probabilités en vue des applications statistiques*, P. Tassi et S. Legait,
- *Cours de probabilités*, A. Monfort.

1 Introduction - rappels

Introduction. Les probabilités à l'ENSAI, pourquoi ?

1.1 Définitions - Notations (rappels)

- univers des possibles ou ensemble fondamental Ω , épreuves $\omega \in \Omega$, tribu \mathcal{A} sur Ω , événements, mesure de probabilité P sur (Ω, \mathcal{A}) , probabilité d'un événement A de \mathcal{A} $P(A)$,
- expérience aléatoire modélisée par un espace de probabilité (Ω, \mathcal{A}, P) (pas nécessairement unique, dépend des hypothèses faites sur l'expérience, ou permet d'en poser).

1.2 Exemples

Les grands classiques du pile ou face, du lancé de dé, des boules dans une urne... Expériences aléatoires plus complexes : sondages, études de fiabilité, durées de vie, ... \Rightarrow Nécessité d'introduire la notion de variable aléatoire.

2 Variables aléatoires réelles

On considère une expérience aléatoire modélisée par un espace de probabilité (Ω, \mathcal{A}, P) .

2.1 Définitions - premiers exemples

Définition 1 (Variable aléatoire réelle) On appelle variable aléatoire réelle (v.a.r.) toute application $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout borélien B de \mathbb{R} ($B \in \mathcal{B}(\mathbb{R})$), l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ appartient à \mathcal{A} .

Propriétés immédiates

Une variable aléatoire réelle est une application $(\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout réel b , $X^{-1}(]-\infty, b]) \in \mathcal{A}$. Les propriétés algébriques usuelles ainsi que la composition par une fonction réelle mesurable conservent la notion de variable aléatoire (ainsi, si X et Y sont deux variables aléatoires réelles, λ un nombre réel, λX , $X + Y$, XY , $|X|$, $\sqrt{|X|}$, $1/X$, X^n , e^X ... sont des variables aléatoires réelles).

Notation. Si $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire réelle, l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ est un événement généralement noté $\{X \in B\}$ et dont la probabilité notée $P(X \in B)$ se lit "probabilité que X soit dans B ". L'ensemble B peut prendre diverses formes et on écrit par exemple, $P(X = b)$ au lieu de $P(X \in \{b\})$, $P(X \leq b)$ au lieu de $P(X \in]-\infty, b])$...

Ces probabilités sont lues "probabilité que X soit égal à b ", "probabilité que X soit inférieur ou égal à b "... Si $X_1, X_2 : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont deux variables aléatoires réelles, la probabilité $P(X_1^{-1}(]-\infty, b_1]) \cap X_2^{-1}(]-\infty, b_2])$) est notée $P(X_1 \leq b_1, X_2 \leq b_2)$ et lue "probabilité que X_1 soit inférieur ou égal à b_1 et que X_2 soit inférieur ou égal à b_2 ".

Exemples. Exemples liés aux expériences aléatoires ci-dessus. Exemple de l'indicatrice d'un événement.

2.2 Loi d'une variable aléatoire réelle

La notion de variable aléatoire réelle permet de probabiliser l'espace mesurable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On a en effet la proposition suivante.

Proposition 1 Soit $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une variable aléatoire réelle. Alors l'application $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ définie pour tout borélien B de \mathbb{R} par

$$P_X(B) = P(X^{-1}(B)),$$

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

Définition 2 (Loi d'une v.a.r.) La mesure de probabilité P_X ainsi définie est appelée loi de probabilité de la variable aléatoire X .

On dit encore que X suit la loi de probabilité P_X , et on peut noter $X \sim P_X$.

Remarques importantes. L'espace $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ est un nouvel espace de probabilité. De plus, pour toute mesure de probabilité Q sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, il existe toujours un espace de probabilité (Ω, \mathcal{A}, P) et une variable aléatoire réelle $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $P_X = Q$ (il suffit de prendre $(\Omega, \mathcal{A}, P) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), Q)$ et pour X l'application identique de Ω).

On peut ainsi parler de variable aléatoire X ayant une loi de probabilité P_X sans spécifier l'espace de probabilité (Ω, \mathcal{A}, P) sur lequel X est définie à la base.

En revanche, il est clair que X n'est pas l'unique v.a.r. telle que $P_X = Q$. Autrement dit, une v.a.r. n'est pas déterminée par sa loi...

2.3 Fonction de répartition - fonction de masse - fonction de densité

2.3.1 La fonction de répartition d'une v.a.r.

Définition 3 (Fonction de répartition d'une v.a.r.) Si $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire réelle on appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie pour tout $x \in \mathbb{R}$ par $F_X(x) = P_X(-\infty, x] = P(X \leq x)$.

Proposition 2 La fonction de répartition d'une v.a.r. X satisfait les propriétés suivantes :

1. Pour tout $x \in \mathbb{R}$, $0 \leq F_X(x) \leq 1$.
2. F_X est une fonction croissante, continue à droite en tout point x de \mathbb{R} .
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

La preuve est laissée en exercice (pas difficile).

Proposition 3 Pour toute fonction réelle F d'une variable réelle vérifiant les propriétés 1., 2., 3. de la Proposition ??, il existe une unique loi de probabilité Q sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $F(x) = Q(-\infty, x]$ pour tout réel x .

Conséquence : La loi d'une v.a.r. est entièrement déterminée par sa fonction de répartition.

Une preuve de l'existence. (Importante pour la simulation de certaines v.a.r.) On va montrer que si F est une fonction réelle d'une variable réelle vérifiant les propriétés 1., 2., 3. de la Proposition ??, il existe une variable aléatoire X telle que $F = F_X$.

On considère U une variable aléatoire de loi uniforme sur $[0, 1]$, c'est-à-dire telle que pour tout $[a, b] \subset [0, 1]$, $P(U \in [a, b]) = b - a$.

Cas particulier où F est continue et strictement croissante. Dans ce cas, la fonction F est une bijection de \mathbb{R} dans $[0, 1]$, et on peut considérer sa réciproque F^{-1} . Si l'on pose $X = F^{-1}(U)$, alors pour tout réel x ,

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(F^{-1}(U) \leq x) \\ &= P(U \leq F(x)) \\ &= F(x). \end{aligned}$$

Cas général. On définit l'inverse généralisée de F . On pose ainsi pour tout $u \in [0, 1]$, $F^{-1}(u) = \text{Inf}\{x \in \mathbb{R}, F(x) \geq u\}$, et on vérifie que pour $u \in [0, 1]$,

$$F^{-1}(u) \leq x \Leftrightarrow F(x) \geq u.$$

Soit $A = \{x \in \mathbb{R}, F(x) \geq u\}$. Si $x_0 \in A$, et $x_1 > x_0$, alors $x_1 \in A$. Par conséquent, A est de la forme $]a, +\infty[$ ou $[a, +\infty[$. Mais si on prend une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de A qui tend en décroissant vers a , comme $F(x_n) \geq u$ et F est continue à droite, on a $F(a) \geq u$. Donc $a \in A$, et A est de la forme $[a, +\infty[$. $F^{-1}(u)$ est un minimum et $A = [F^{-1}(u), +\infty[$. On reprend la fin de la preuve précédente pour conclure.

2.3.2 Fonction de masse et classification des v.a.r.

Définition 4 (Fonction de masse d'une v.a.r.) Si $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une variable aléatoire réelle, la fonction $\pi_X : x \in \mathbb{R} \mapsto P(X = x)$ est appelée la fonction de masse de X .

On montre à l'aide des propriétés de base des mesures de probabilité que pour tout $x \in \mathbb{R}$, $\pi_X(x) = F_X(x) - F_X(x-)$. D'où la :

Proposition 4 Si X est une v.a.r., alors l'ensemble (noté D_X) des points de discontinuité de sa fonction de répartition F_X vérifie $D_X = \{x, \pi_X(x) > 0\}$. De plus, cet ensemble est fini ou dénombrable.

Définition 5 (loi diffuse, loi discrète) Si $D_X = \emptyset$, on dit que la loi de X est diffuse.

Si $P(X \in D_X) = \sum_{x \in D_X} \pi_X(x) = 1$, on dit que la loi de X est discrète (et que son support est $S_X = D_X$).

Il est facile de montrer que si $a, b \in \mathbb{R}$, $P(X \in]a, b]) = F_X(b) - F_X(a)$, $P(X \in [a, b]) = F_X(b) - F_X(a) + \pi_X(a)$, $P(X \in]a, b[) = F_X(b) - \pi_X(b) - F_X(a)$, $P(X \in [a, b[) = F_X(b) - \pi_X(b) - F_X(a) + \pi_X(a)$. Donc si la loi de X est diffuse, $P(X \in]a, b]) = P(X \in [a, b]) = P(X \in]a, b[) = P(X \in [a, b[) = F_X(b) - F_X(a)$.

Proposition 5 Il existe des v.a.r. dont la loi de probabilité Q n'est ni diffuse ni discrète. On peut alors montrer qu'il existe $\alpha \in]0, 1[$, Q_1 une loi discrète, Q_2 une loi diffuse tels que $Q = \alpha Q_1 + (1 - \alpha)Q_2$. La loi Q est dite mixte.

Pour la preuve, on renvoie au polycopié de Michel Carbon.

Les v.a.r. de lois discrètes ont été traitées en détail dans le cours de Statistique 1. On s'intéresse ici un peu plus particulièrement à des v.a.r de lois diffuses particulières : les v.a.r. de lois absolument continues (les autres ont des lois dites *singulières*.)

2.3.3 Loi absolument continue - densité

Définition 6 (V.a.r de loi absolument continue) La loi de probabilité P_X d'une v.a.r. est dite absolument continue si elle est absolument continue par rapport à la mesure de Lebesgue λ .

Théorème 1 (Théorème de Radon-Nikodym) La loi de probabilité P_X d'une v.a.r. est absolument continue si et seulement si il existe une fonction f_X positive et intégrable telle que pour tout $x \in \mathbb{R}$,

$$F_X(x) = \int_{]-\infty, x]} f_X(t) d\lambda(t). \tag{1}$$

Définition 7 On appelle densité de la loi P_X (ou par abus densité de X) toute fonction f_X qui vérifie (??).

On a les propriétés suivantes :

1. Si (??) est vérifiée, F_X est presque partout dérivable avec $F_X' = f_X$ presque partout.
2. Si (??) est vérifiée, il est évident que $\int_{\mathbb{R}} f_X(t) d\lambda(t) = 1$.
3. La fonction f_X est unique à une λ -équivalence près.
4. Important en pratique : Si la fonction de répartition F_X admet en tout point de \mathbb{R} une dérivée continue, alors la loi de X est absolument continue. Attention : la réciproque est fautive ! Il est seulement garanti que si P_X est absolument continue, alors F_X est continue.
5. Une loi absolument continue est diffuse. Son support est défini par $S_X = \{x \in \mathbb{R}, f_X(x) > 0\}$.

Remarque : On a vu que f_X est unique à une équivalence par rapport à la mesure de Lebesgue près. Donc, à une λ -équivalence près, la fonction f_X caractérise la loi.

Proposition 6 Toute fonction f réelle positive, d'une variable réelle, intégrable (au sens de Lebesgue) et telle que $\int_{\mathbb{R}} f(t)d\lambda(t) = 1$, est la densité d'une loi de probabilité.

Une dernière remarque (*Last but not least...*) : Si P_X est absolument continue, pour tout borélien B de \mathbb{R} , on a d'après (??), $P_X(B) = P(X \in B) = \int_B f_X d\lambda$. **Interprétation graphique de la densité à comparer avec l'interprétation de l'histogramme d'une loi discrète.** Pour une loi P_X discrète de support $S_X \subset \mathbb{Z}$, et toute partie B de S_X , si μ désigne la mesure de comptage sur \mathbb{Z} , on a $P_X(B) = \sum_{x \in B} \pi_X(x) = \int_B \pi_X d\mu$. On peut donc regrouper l'étude des lois discrètes et celle des lois absolument continues, en considérant la notion générale d'absolue continuité vue en cours d'intégration (Statistique 1), et en considérant les fonctions de masse des lois discrètes comme des densités par rapport à des mesures de comptage. On donnera néanmoins ici, à la suite de chaque résultat, les différentes expressions dans chacun des cas.

2.4 Espérance mathématique - Moments d'une v.a.r.

2.4.1 Définitions, premières propriétés

Pré-requis fondamental : Intégration d'une v.a.r.

Rappel du cours de Statistique 1 : Théorème de transfert.

On considère ici un espace de probabilité (Ω, \mathcal{A}, P) , et une v.a.r. X définie sur (Ω, \mathcal{A}, P) , de loi de probabilité P_X .

Théorème 2 (Théorème de transfert) Soit φ une fonction mesurable réelle définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. $\varphi(X)$ est P -intégrable si et seulement si φ est P_X -intégrable, et l'on a :

$$\int_{\Omega} \varphi(X(\omega))dP(\omega) = \int_{\mathbb{R}} \varphi(x)dP_X(x).$$

Définition 8 (Espérance mathématique) Si X est à valeurs positives ou si X est P -intégrable, on appelle espérance mathématique de X , notée $E[X]$, l'intégrale de X par rapport à P :

$$E[X] = \int X dP = \int_{\Omega} X(\omega)dP(\omega) = (\text{Th. de transfert}) \int_{\mathbb{R}} x dP_X(x).$$

Remarques.

1. Le théorème de transfert permet de calculer l'espérance mathématique de X sans connaître l'espace de probabilité sous-jacent.
2. L'espérance mathématique vérifie toutes les propriétés de l'intégrale, en particulier la linéarité et la monotonie.
3. Interprétation de $E[X]$ comme indicateur de centralité. On parle souvent de valeur "moyenne" de X ...
4. Si $A \in \mathcal{A}$, on a $P(A) = E[\mathbb{I}_A]$.
5. Important : Soit φ une fonction mesurable réelle définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors $\varphi(X)$ est aussi une v.a.r. définie sur l'espace de probabilité (Ω, \mathcal{A}, P) . Par conséquent, si elle est positive ou P -intégrable, elle possède une espérance, qui d'après le théorème de transfert, est égale à

$$E[\varphi(X)] = \int \varphi(X)dP = \int_{\mathbb{R}} \varphi(x)dP_X(x).$$

Calcul effectif dans le cas d'une v.a.r. de loi discrète ou absolument continue.

1. Si P_X est une loi discrète de fonction de masse π_X , de support S_X ($P_X = \sum_{x \in S_X} \pi_X(x)\delta_x$), si X est positive ou si $\sum_{x \in S_X} |x|\pi_X(x) < +\infty$, alors $E[X]$ existe, et

$$E[X] = \sum_{x \in S_X} x\pi_X(x).$$

- Si φ est une fonction réelle mesurable sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, si φ est positive ou si $\sum_{x \in S_X} |\varphi(x)|\pi_X(x) < +\infty$, alors $E[\varphi(X)]$ existe et

$$E[\varphi(X)] = \sum_{x \in S_X} \varphi(x)\pi_X(x).$$

2. Si P_X est une loi absolument continue, de densité f_X , si X est positive ou si $x \mapsto |x|f_X(x)$ est λ -intégrable, alors $E[X]$ existe et

$$E[X] = \int_{\mathbb{R}} x f_X(x) d\lambda(x).$$

Si φ est une fonction réelle mesurable sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, si φ est positive ou si $x \mapsto |\varphi(x)|f_X(x)$ est λ -intégrable, alors $E[\varphi(X)]$ existe et

$$E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f_X(x) d\lambda(x).$$

Définition 9 Si $|E[X]| < +\infty$, on dit que X a une espérance finie.

Si $E[X]$ existe et $E[X] = 0$, X est dite centrée.

Si $E[X]$ existe, on dit qu'on centre ou recentre X lorsqu'on lui retranche $E[X]$.

On a, de façon évidente d'après les propriétés de l'intégrale, $|E[X]| \leq E[|X|]$. En fait, on a une inégalité plus générale.

Proposition 7 (Inégalité de Jensen) Si X est une v.a.r. à valeurs dans un intervalle $]a, b[$, d'espérance finie, si ϕ est une fonction réelle convexe sur $]a, b[$, alors

$$\phi(E[X]) \leq E[\phi(X)].$$

On introduit ici d'autres caractéristiques de la loi de X , qui rendent compte de sa dispersion, par exemple les moments, ou les moments centrés.

Proposition 8 Soit $r, r' \in \mathbb{R}$, $0 < r < r'$. Si $E[|X|^{r'}] < +\infty$, alors $E[|X|^r] < +\infty$.

Preuve. Pour $x \geq 1$, on a $x^r = x^{r'} x^{r-r'} \leq x^{r'}$, et pour $0 < x < 1$, $x^r < 1$. Par conséquent, pour tout $x > 0$, on a $x^r \leq x^{r'} + 1$. On a donc en particulier $|X|^r \leq 1 + |X|^{r'}$, et comme $E[1 + |X|^{r'}] = 1 + E[|X|^{r'}]$, le résultat suit par monotonie de l'espérance.

Définition 10 (Moments) Soit $a, r \in \mathbb{R}$. Si X est positive ou si $E[|X|^r] < +\infty$, on appelle moment d'ordre r de X le nombre $E[X^r]$, et moment absolu d'ordre r de X le nombre $E[|X|^r]$.

Si $(X - a)^r$ est positive ou si $E[|X - a|^r] < +\infty$, on appelle moment centré en a d'ordre r de X le nombre $E[(X - a)^r]$. Lorsque $E[|X|] < +\infty$ et $a = E[X]$, on parle du moment centré de X .

On s'intéresse en particulier aux moments d'ordre 2.

Définition 11 (Variance, écart-type) Le moment centré d'ordre 2 de X est appelé la variance de X , et noté $Var(X)$:

$$Var(X) = E[(X - E[X])^2].$$

Sa racine carrée positive est appelée l'écart-type de X , noté $\sigma(X)$.

Propriétés immédiates.

1. Si $\alpha \in \mathbb{R}$, $Var(\alpha X) = \alpha^2 Var(X)$.
2. Si $a \in \mathbb{R}$, $Var(a + X) = Var(X)$.
3. $Var(X) = 0$ si et seulement si X est une v.a.r. presque certaine ($X = E[X]$ p.s.).

Définition 12 Si $E[X] = 0$ et $Var(X) = 1$, on dit que X est centrée réduite. Si $E[|X|] < +\infty$ et $Var(X) < +\infty$, on dit qu'on centre et on réduit X si l'on calcule $\frac{X - E[X]}{\sigma(X)}$.

Il résulte de la Proposition ?? que si X admet un moment d'ordre 2 fini, alors elle a une espérance mathématique finie.

Proposition 9 *X a un moment d'ordre 2 fini si et seulement si son espérance mathématique $E[X]$ et sa variance $Var(X)$ existent et sont finies, et on a :*

$$Var(X) = E[X^2] - (E[X])^2.$$

Preuve. Si X a un moment d'ordre 2 fini, $E[X]$ est finie et on a $(X - E[X])^2 = X^2 - 2XE[X] + (E[X])^2$, d'où $Var(X) = E[X^2] - (E[X])^2$. Pour la réciproque, on remarque que si $E[|X|] < \infty$, $Var(X) < +\infty$, comme $E[X^2] = E[(X - E[X] + E[X])^2] = E[(X - E[X])^2] + (E[X])^2 + E[(X - E[X])E[X]] = E[(X - E[X])^2] + (E[X])^2 < +\infty$.

Remarque : cette proposition nous donne une nouvelle méthode de calcul de la variance, souvent plus aisée à utiliser.

Proposition 10 *Si $E[X^2] < +\infty$, pour tout réel a , $E[(X - a)^2] \geq Var(X)$.*

Preuve. Il suffit d'écrire $E[(X - a)^2] = E[((X - E[X]) + (E[X] - a))^2]$ et de développer.

Interprétation en termes de dispersion.

2.4.2 Inégalités faisant intervenir les moments

Les moments permettent, comme on l'a dit ci-dessus, de donner une indication sur la dispersion d'une variable. Cette dispersion peut être précisée par exemple à l'aide des inégalités suivantes.

Théorème 3 (Inégalité de Markov) *Si X est positive, pour tout réel $a > 0$,*

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Preuve. On a $E[X] = \int_{\mathbb{R}_+} x dP_X(x) = \int_{[0,a[} x dP_X(x) + \int_{[a,+\infty[} x dP_X(x)$, d'où $E[X] \geq \int_{[a,+\infty[} x dP_X(x) \geq a \int_{[a,+\infty[} dP_X(x)$.

Théorème 4 (Inégalité de Bienaymé-Chebychev) *Si $E[X^2] < +\infty$, alors pour tout réel $a > 0$,*

$$P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}.$$

Preuve. Il suffit d'appliquer l'inégalité de Markov à la variable aléatoire $(X - E[X])^2$.

Remarque : Ces deux inégalités ont un très grand intérêt théorique. Cependant, l'inégalité de Bienaymé-Chebychev ne fournit pas en général une bonne majoration de $P(|X - E[X]| \geq a)$. On verra en TD une autre inégalité découlant de l'inégalité de Markov, l'inégalité de Chernov, qui en donne une meilleure approximation.

2.5 Médiane - Quantiles d'une var

L'espérance est-elle toujours le meilleur indicateur de centralité d'une v.a.r. ? On verra en TD un exemple où ce n'est clairement pas le cas. Il est parfois nécessaire d'étudier la médiane.

Définition 13 (Médiane) *On appelle médiane de X tout nombre M tel que*

$$P(X \leq M) \geq 1/2 \text{ et } P(X \geq M) \geq 1/2.$$

Plus généralement, on définit les quantiles de X .

Définition 14 (Quantiles) *Pour $\alpha \in]0, 1[$, on appelle quantile d'ordre α ou α -quantile de X tout nombre q_α tel que*

$$P(X \leq q_\alpha) \geq \alpha \text{ et } P(X \geq q_\alpha) \geq 1 - \alpha.$$

Interprétation graphique. Exemples.

Remarque : Si F_X est la fonction de répartition de X , si F_X^{-1} désigne son inverse généralisée, alors un quantile d'ordre α de X est donné par $F_X^{-1}(\alpha)$. Et en particulier, si F_X est continue et strictement croissante, X admet pour unique quantile d'ordre α $F_X^{-1}(\alpha)$, c'est-à-dire le nombre q_α tel que $P(X \leq q_\alpha) = \alpha$.

2.6 Calcul de lois

Dans ce paragraphe, on considère toujours l'espace de probabilité (Ω, \mathcal{A}, P) , et une v.a.r. X définie sur (Ω, \mathcal{A}, P) , dont on connaît la loi de probabilité P_X . On cherche ici à répertorier les méthodes de détermination de la loi d'une variable aléatoire définie comme $\varphi(X)$, φ étant une fonction réelle mesurable.

2.6.1 Première méthode. Utilisation de la fonction de répartition.

On a vu que la fonction de répartition caractérise entièrement la loi d'une variable aléatoire, donc on peut chercher à exprimer par exemple $F_{\varphi(X)}$ en fonction de F_X (c.f. exos de TD). Cette méthode est valable pour tout type de v.a.r. Cependant, lorsqu'on connaît le type de la variable $\varphi(X)$, on préfère caractériser la loi par des fonctions plus explicites (fonction de masse pour une loi discrète, fonction de densité pour une loi absolument continue).

Pour une v.a.r. de loi discrète, la fonction de masse s'obtient à partir de la fonction de répartition très facilement ($\pi_X(x) = F_X(x) - F_X(x-)$).

Pour une v.a.r. de loi absolument continue, la fonction de densité s'obtient également à partir de la fonction de répartition par dérivation $f_X = (F_X)' \lambda - p.p.$

Il est en fait souvent plus facile de déterminer directement les fonctions de masse et de densité sans passer par la détermination de la fonction de répartition.

2.6.2 Deuxième méthode. Détermination directe de la fonction de masse dans le cas d'une loi discrète.

On cherche ici à exprimer directement $P(\varphi(X) = y)$ pour tout $y \in \varphi(S_X)$, et on utilise pour cela les propriétés de base des probabilités.

2.6.3 Troisième méthode. Détermination directe de la fonction de densité dans le cas d'une loi absolument continue.

Cette méthode est basée sur le théorème suivant.

Théorème 5 Soit X une v.a.r. dont la loi est absolument continue, de densité f . Alors pour toute fonction réelle mesurable bornée h , on a

$$E[h(X)] = \int_{\mathbb{R}} h(x)f(x)d\lambda(x). \quad (2)$$

Réciproquement, si pour toute fonction h réelle mesurable bornée, l'équation (??) est vérifiée, alors f est la densité de la loi de X .

Il s'agit donc ici de trouver une fonction g telle que pour toute fonction borélienne bornée

$$E[h(\varphi(X))] = \int_{\mathbb{R}} h(y)g(y)d\lambda(y).$$

Par identification, on pourra alors affirmer que g est la densité de $\varphi(X)$.

Comment trouver cette fonction ?

Un cas simple est celui où X admet pour densité $f_X \mathbb{I}_U$, où U est un ouvert de \mathbb{R} , et où φ est un difféomorphisme de U sur un ouvert V .

On peut en effet écrire, puisque $h \circ \varphi$ est une fonction réelle mesurable,

$$\begin{aligned} E[h(\varphi(X))] &= \int_{\mathbb{R}} h \circ \varphi(x) dP_X(x) \\ &= \int_U h(\varphi(x)) f_X(x) d\lambda(x) \\ &= \int_V h(y) f_X(\varphi^{-1}(y)) |(\varphi^{-1})'(y)| d\lambda(y), \end{aligned}$$

la dernière ligne étant obtenue par le théorème de changement de variables. Ainsi, la densité de $\varphi(X)$ est donnée par $y \mapsto f_X(\varphi^{-1}(y)) |(\varphi^{-1})'(y)| \mathbb{I}_V(y)$.

Dans les autres cas, on peut toujours découper le support de X en morceaux sur lesquels φ est un difféomorphisme, ou revenir à la première méthode...

Exemples.

3 Vecteurs aléatoires

On généralise dans ce chapitre les notions introduites dans le chapitre précédent.

On considère une expérience aléatoire modélisée par un espace de probabilité (Ω, \mathcal{A}, P) .

Remarque importante : Par commodité d'écriture, un vecteur $(x_1, \dots, x_n)'$ sera noté ici (x_1, \dots, x_n) (on ne précisera pas la transposée) sauf dans quelques cas où cela sera précisé.

3.1 Définitions - premiers exemples

Pour tout $i \in \{1, \dots, n\}$ ($n \in \mathbb{N}^*$), on note π_i la projection canonique $\mathbb{R}^n \rightarrow \mathbb{R}$, $x = (x_1, \dots, x_n) \mapsto x_i$.

Définition 15 (Vecteur aléatoire, variables aléatoires marginales) On appelle variable aléatoire à valeurs dans \mathbb{R}^n ou vecteur aléatoire à n dimensions toute application $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ telle que pour tout borélien B de \mathbb{R}^n ($B \in \mathcal{B}(\mathbb{R}^n)$), l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ appartient à \mathcal{A} .

Si pour tout $i = 1, \dots, n$, $X_i = \pi_i(X)$ est la i ème coordonnée de X dans \mathbb{R}^n (de telle sorte que $X = (X_1, \dots, X_n)$), X_1, \dots, X_n sont des variables aléatoires réelles, appelées variables aléatoires marginales.

Lorsque $n = 2$, $X = (X_1, X_2)$ est aussi appelé couple aléatoire.

Notation. Si $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ est un vecteur aléatoire, si B est un borélien de \mathbb{R}^n , l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ est un événement généralement noté $\{X \in B\}$ et dont la probabilité notée $P(X \in B)$ se lit "probabilité que X soit dans B ". Suivant les différentes formes de B , on adopte une terminologie similaire à celle des variables aléatoires réelles.

Exemples.

3.2 Loi conjointe, lois marginales

La notion de vecteur aléatoire permet de probabiliser l'espace mesurable $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. On a en effet la proposition suivante.

Proposition 11 Soit $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ un vecteur aléatoire à n dimensions. Alors l'application $P_X : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, 1]$ définie pour tout borélien B de \mathbb{R}^n par

$$P_X(B) = P(X^{-1}(B)),$$

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

Définition 16 (Loi conjointe d'un vecteur aléatoire) La mesure de probabilité P_X ainsi définie est appelée loi de probabilité ou loi conjointe du vecteur aléatoire X .

On dit encore que X suit la loi de probabilité P_X ou a pour loi conjointe P_X , et on peut noter $X \sim P_X$.

Remarques. L'espace $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_X)$ est un nouvel espace de probabilité. De plus, pour toute mesure de probabilité Q sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, il existe toujours un espace de probabilité (Ω, \mathcal{A}, P) et un vecteur aléatoire $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ tel que $P_X = Q$.

Comme pour les variables aléatoires réelles, on peut donc aussi parler de vecteur aléatoire X ayant une loi conjointe P_X sans spécifier l'espace de probabilité (Ω, \mathcal{A}, P) sur lequel X est défini à la base.

La loi conjointe d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ détermine les lois des variables marginales X_1, \dots, X_n .

Proposition 12 La loi conjointe P_X du vecteur aléatoire à n dimensions $X = (X_1, \dots, X_n)$ permet de déterminer les lois (dites marginales) P_{X_1}, \dots, P_{X_n} des variables aléatoires marginales X_1, \dots, X_n . On a en effet pour tout borélien B de \mathbb{R} , $P_{X_i}(B) = P_X(\pi_i^{-1}(B))$ ($i \in \{1, \dots, n\}$).

Remarque importante : la connaissance des lois marginales ne détermine pas la loi conjointe P_X . Tout va dépendre des "relations" pouvant exister entre les variables marginales X_1, \dots, X_n .

3.3 Fonctions de répartition - densités

Définition 17 (Fonction de répartition conjointe) Si $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ est un vecteur aléatoire à n dimensions, on appelle fonction de répartition conjointe de la loi de probabilité de X ou (par abus) fonction de répartition conjointe de X la fonction $F_X : \mathbb{R}^n \rightarrow [0, 1]$ définie par $F_X(x_1, \dots, x_n) = P_X(\prod_{i=1}^n]-\infty, x_i]) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$.

Les fonctions de répartition conjointes ne sont pas très souvent utilisées. On cite donc ici sans les démontrer les trois propositions suivantes.

Proposition 13 La fonction de répartition conjointe F_X d'un vecteur aléatoire X satisfait les propriétés suivantes :

1. Pour tout $x \in \mathbb{R}^n$, $0 \leq F_X(x) \leq 1$.
2. F_X est une fonction croissante, continue à droite en chacun de ses arguments x_1, \dots, x_n .
3. $F_X(x)$ tend vers 0 lorsque l'un des arguments x_i tend vers $-\infty$, et tend vers 1 lorsque tous les x_i tendent vers $+\infty$.

Proposition 14 La fonction de répartition conjointe caractérise la loi conjointe.

Proposition 15 La fonction de répartition conjointe d'un vecteur aléatoire $X = (X_1, \dots, X_n)$ détermine les fonctions de répartition F_{X_1}, \dots, F_{X_n} des variables aléatoires marginales X_1, \dots, X_n (appelées fonctions de répartition marginales). En effet, pour tout $i \in \{1, \dots, n\}$,

$$F_{X_i}(x_i) = P(X_i \leq x_i) = \lim_{x_1, \dots, x_{i-1}, x_{i+1}, x_n \rightarrow +\infty} F_X(x_1, \dots, x_n) = F_X(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty).$$

Par commodité d'écriture, on ne considère dans toute la suite de ce paragraphe que des couples aléatoires ($n = 2$). Toutes les notions vues ici s'étendent néanmoins aux vecteurs aléatoires à n dimensions (n quelconque).

Définition 18 (Loi discrète) La loi conjointe P_X d'un couple aléatoire $X = (X_1, X_2)$ est dite discrète s'il existe un sous-ensemble S_X de \mathbb{R}^2 fini ou dénombrable tel que pour tout $(x_1, x_2) \in S_X$, $P_X(x_1, x_2) > 0$ et $P_X(S_X) = \sum_{(x_1, x_2) \in S_X} P_X(x_1, x_2) = 1$. L'ensemble S_X est alors appelé support de la loi conjointe de X .

Remarque : un couple aléatoire a une loi conjointe discrète si et seulement si ses lois marginales sont discrètes.

Proposition 16 Si $X = (X_1, X_2)$ est un couple aléatoire de loi conjointe P_X discrète, si S_{X_1} désigne le support de X_1 , S_{X_2} celui de X_2 , alors $S_X \subset S_{X_1} \times S_{X_2}$ et les lois marginales de X_1 et X_2 sont données par :

$$P_{X_1}(x_1) = \sum_{x_2 \in S_{X_2}} P_X(x_1, x_2) \quad \forall x_1 \in S_{X_1}, \quad P_{X_2}(x_2) = \sum_{x_1 \in S_{X_1}} P_X(x_1, x_2) \quad \forall x_2 \in S_{X_2}.$$

Exemple. On lance successivement trois pièces de monnaie (équilibrées) et on désigne respectivement par X et Y les variables aléatoires représentant le nombre de *faces* apparues sur les deux premières pièces et le nombre de *piles* sur les deux dernières.

La loi conjointe du couple (X, Y) et les lois marginales de X et Y sont données dans le tableau suivant :

$Y \backslash X$	0	1	2	
0	0	1/8	1/8	1/4
1	1/8	2/8	1/8	1/2
2	1/8	1/8	0	1/4
	1/4	1/2	1/4	1

Définition 19 (Loi absolument continue) La loi conjointe P_X d'un couple aléatoire $X = (X_1, X_2)$ est dite absolument continue si elle est absolument continue par rapport à la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

Théorème 6 (Théorème de Radon-Nikodym) La loi conjointe P_X d'un couple aléatoire $X = (X_1, X_2)$ est absolument continue si et seulement s'il existe une fonction f_X positive et intégrable telle que pour tout borélien B de \mathbb{R}^2 ,

$$P_X(B) = \int \int_B f_X(x_1, x_2) d\lambda(x_1) d\lambda(x_2). \tag{3}$$

Définition 20 (Densité conjointe) On appelle densité conjointe de la loi P_X ou (par abus) densité conjointe de X toute fonction f_X qui vérifie (??).

Une densité conjointe f_X vérifie les propriétés suivantes :

1. $\int \int_{\mathbb{R}^2} f_X(x_1, x_2) d\lambda(x_1) d\lambda(x_2) = 1$.
2. Si F_X désigne la fonction de répartition conjointe de X ,

$$F_X(x_1, x_2) = \int \int_{]-\infty, x_1] \times]-\infty, x_2]} f_X(x_1, x_2) d\lambda(x_1) d\lambda(x_2).$$

3. Si f_X est continue en (x_1, x_2) , alors $f_X(x_1, x_2) = \frac{\partial^2 F_X}{\partial x_1 \partial x_2}(x_1, x_2)$.
4. Si la fonction de répartition F_X est de classe \mathcal{C}^2 sur \mathbb{R}^2 , alors la loi de X est absolument continue et $f_X(x_1, x_2) = \frac{\partial^2 F_X}{\partial x_1 \partial x_2}(x_1, x_2)$. Attention : la réciproque est fautive !
5. A une équivalence près pour la mesure de Lebesgue sur \mathbb{R}^2 , la fonction f_X caractérise la loi conjointe P_X .

Proposition 17 Toute fonction f_X de \mathbb{R}^2 dans \mathbb{R} , positive, intégrable, et telle que

$$\int \int_{\mathbb{R}^2} f_X(x_1, x_2) d\lambda(x_1) d\lambda(x_2) = 1,$$

est la densité conjointe d'une loi absolument continue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$.

Proposition 18 Si le couple $X = (X_1, X_2)$ admet une loi conjointe absolument continue, alors ses lois marginales sont absolument continues et sa densité conjointe détermine les densités des lois marginales (appelées densités marginales) f_{X_1} et f_{X_2} de la façon suivante :

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) d\lambda(x_2), \quad f_{X_2}(x_2) = \int_{\mathbb{R}} f_{(X_1, X_2)}(x_1, x_2) d\lambda(x_1).$$

Preuve : Théorème de Fubini-Tonnelli.

Exemple. Soit (X, Y) un couple aléatoire dont la densité conjointe est définie par :

$$f(x, y) = \begin{cases} xy/2 & \text{si } 0 \leq y \leq x \leq 2 \\ 0 & \text{sinon.} \end{cases}$$

1. La fonction de répartition du couple (X, Y) est donnée par les équations suivantes. Si $x \leq 0$ ou si $y \leq 0$, $F(x, y) = 0$. Si $0 \leq y \leq x \leq 2$, $F(x, y) = \frac{1}{2} \int_0^y v (\int_v^x u du) dv = \frac{1}{4} \int_0^y v(x^2 - v^2) dv = \frac{y^2}{16} (2x^2 - y^2)$. Si $0 \leq y \leq 2 \leq x$, $F(x, y) = F(2, y) = \frac{y^2}{16} (8 - y^2)$. Si $0 < x \leq y$ et $x \leq 2$, $F(x, y) = F(x, x) = \frac{x^4}{16}$. Enfin, si $x \geq 2$, $y \geq 2$, $F(x, y) = 1$.
2. Les densités marginales s'obtiennent par exemple par intégration de la densité conjointe : $f_X(x) = \frac{x^3}{4} \mathbb{I}_{[0,2]}(x)$, et $f_Y(y) = \frac{y}{4} (4 - y^2) \mathbb{I}_{[0,2]}(y)$ ou à partir de la fonction de répartition conjointe (plus difficile).
3. On peut calculer par exemple $P(Y \geq X^2) : P(Y \geq X^2) = \int \int \frac{xy}{2} \mathbb{I}_{0 \leq y \leq x \leq 2} \mathbb{I}_{y \geq x^2} dx dy = \int \int \frac{xy}{2} \mathbb{I}_{0 \leq x \leq 1} \mathbb{I}_{x^2 \leq y \leq x} dx dy$ (faire un graphe), d'où $P(Y \geq X^2) = \frac{1}{2} \int_0^1 x (\int_{x^2}^x y dy) dx = 1/48$.

3.4 Espérance mathématique - Moments d'un vecteur aléatoire

On considère un vecteur aléatoire $X = (X_1, \dots, X_n)$ défini sur (Ω, \mathcal{A}, P) de loi conjointe notée P_X .

Rappel : Théorème de transfert pour les vecteurs aléatoires.

Théorème 7 (Théorème de transfert) Soit φ une fonction mesurable de $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors $\varphi(X)$ est une variable aléatoire réelle P -intégrable si et seulement si φ est P_X -intégrable, et l'on a :

$$E[\varphi(X)] = \int_{\Omega} \varphi(X(\omega)) dP(\omega) = \int \dots \int_{\mathbb{R}^n} \varphi(x_1, \dots, x_n) dP_X(x_1, \dots, x_n).$$

3.4.1 Espérance mathématique

Définition 21 (Espérance mathématique) On dit que $X = (X_1, \dots, X_n)$ possède une espérance mathématique si ses variables marginales X_1, \dots, X_n possèdent une espérance mathématique, et on note

$$E[X] = (E[X_1], \dots, E[X_n]).$$

Les propriétés de l'espérance mathématique pour les vecteurs aléatoires découlent donc directement de celles de l'espérance mathématique des variables aléatoires réelles. En particulier, on a le résultat de linéarité suivant :

Proposition 19 Si A est une matrice réelle de taille $p \times n$, alors

$$E[AX] = AE[X].$$

On utilise les termes de centrage et recentrage comme pour les variables aléatoires réelles.

3.4.2 Variations-covariances

Définition 22 (Variations-covariances) Un vecteur aléatoire $X = (X_1, \dots, X_n)$ dont toutes les variables marginales possèdent un moment d'ordre 2 (i.e. $E[X_i^2] < +\infty \forall i = 1, \dots, n$) est dit du second ordre.

On appelle matrice des moments du second ordre de X la matrice carrée M_X de taille $n \times n$ de terme général $m_{i,j} = E[X_i X_j]$.

On appelle matrice des variations-covariances de X la matrice carrée Σ_X de taille $n \times n$ de terme général $\sigma_{i,j}^2 = E[(X_i - E[X_i])(X_j - E[X_j])] = E[X_i X_j] - E[X_i]E[X_j]$. $\sigma_{i,j}^2$ s'appelle la covariance de X_i et X_j , et on la note $Cov(X_i, X_j)$.

Remarques :

1. L'existence des $m_{i,j}$ et $\sigma_{i,j}^2$ est assurée par l'inégalité de Cauchy-Schwarz :

Théorème 8 (Inégalité de Cauchy-Schwarz) Si (X_1, X_2) est un couple aléatoire, alors

$$(E[X_1 X_2])^2 \leq E[X_1^2] E[X_2^2].$$

2. Les matrices M_X et Σ_X sont des matrices symétriques.

3. Pour $i = 1, \dots, n$, $Cov(X_i, X_i) = Var(X_i)$ donc les éléments diagonaux de la matrice de variances-covariances sont les variances des variables marginales.

Définition 23 Un vecteur aléatoire du second ordre est dit réduit si $\Sigma_X = I$.

Pour les deux propositions suivantes, on précise à nouveau la transposée.

En admettant que l'on définit une matrice aléatoire de la même façon qu'un vecteur aléatoire, on a le résultat suivant.

Proposition 20 Si $X = (X_1, \dots, X_n)'$ est un vecteur aléatoire du second ordre,

$$M_X = E[XX'],$$

et

$$\Sigma_X = E[(X - E[X])(X - E[X])'].$$

Des propriétés de linéarité de l'espérance, on déduit la proposition suivante.

Proposition 21 Si $X = (X_1, \dots, X_n)'$ est un vecteur aléatoire du second ordre et A une matrice réelle de taille $p \times n$, on a

$$M_{AX} = AM_X A',$$

et

$$\Sigma_{AX} = A\Sigma_X A'.$$

Proposition 22 Une matrice symétrique Σ est la matrice de variances-covariances d'un vecteur aléatoire si et seulement si elle est positive.

Preuve. On considère un vecteur aléatoire $U = (U_1, \dots, U_n)$ de matrice de variances-covariances $\Sigma_U = I$ (on verra après le paragraphe sur l'indépendance qu'un tel vecteur existe). On sait que toute matrice symétrique positive Σ s'écrit sous la forme $\Sigma = P\Delta P'$, P étant une matrice orthogonale (matrice de changement de base), Δ une matrice diagonale positive (composée des valeurs propres de Σ). On peut donc introduire $\Sigma^{1/2} = P\Delta^{1/2}P'$, où $\Delta^{1/2}$ est la matrice diagonale, dont les éléments diagonaux sont les racines carrées positives des éléments diagonaux de Δ . Le vecteur $X = \Sigma^{1/2}U$ a pour matrice de variances-covariances Σ .

Proposition 23 Soit X un vecteur aléatoire dont la matrice de variances-covariances Σ_X est inversible. Le vecteur $Y = \Sigma_X^{-1/2}X$ est réduit, et le vecteur $\Sigma_X^{-1/2}(X - E[X])$ est centré réduit.

3.4.3 Somme de variables aléatoires réelles

Une fois les covariances définies, on peut étudier un peu plus précisément les caractéristiques des sommes de n variables aléatoires réelles.

Proposition 24 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire du second ordre.

Alors

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n],$$

et

$$Var(X_1 + \dots + X_n) = \sum_{i=1}^n Var(X_i) + \sum_{1 \leq i \neq j \leq n} Cov(X_i, X_j) = \sum_{1 \leq i, j \leq n} Cov(X_i, X_j).$$

3.5 Corrélation, indépendance

On considère un espace de probabilité (Ω, \mathcal{A}, P) .

3.5.1 Le coefficient de corrélation linéaire

Définition 24 Soit (X_1, X_2) un couple aléatoire défini sur (Ω, \mathcal{A}, P) , du second ordre et tel que $Var(X_1)Var(X_2) > 0$. On appelle coefficient de corrélation (linéaire) du couple (X_1, X_2) le nombre

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}} = E \left[\frac{X_1 - E[X_1]}{\sqrt{Var(X_1)}} \frac{X_2 - E[X_2]}{\sqrt{Var(X_2)}} \right].$$

Définition 25 Soit (X_1, X_2) un couple aléatoire défini sur (Ω, \mathcal{A}, P) , du second ordre. On dit que X_1 et X_2 sont corrélées si $Cov(X_1, X_2) \neq 0$, non corrélées si $Cov(X_1, X_2) = 0$.

Proposition 25 On a $|\rho(X_1, X_2)| \leq 1$.

Preuve. Inégalité de Cauchy-Schwarz, dont on redonne la preuve dans ce cas particulier. Pour tout $\lambda \in \mathbb{R}$,

$$0 \leq E[(X_1 - E[X_1] + \lambda X_2 - \lambda E[X_2])^2] = Var(X_1) + 2\lambda Cov(X_1, X_2) + \lambda^2 Var(X_2).$$

On a un trinôme du second degré en λ , qui est de signe constant, donc son discriminant est négatif ou nul, c'est-à-dire $Cov^2(X_1, X_2) \leq Var(X_1)Var(X_2)$ ou encore $\rho^2(X_1, X_2) \leq 1$

Proposition 26 Soit (X_1, X_2) un couple aléatoire défini sur (Ω, \mathcal{A}, P) , du second ordre, et tel que $Var(X_1)Var(X_2) > 0$. Si $|\rho(X_1, X_2)| = 1$, alors les variables marginales X_1 et X_2 sont liées par une relation affine.

Preuve. On reprend l'idée de la preuve précédente (cas d'égalité dans l'inégalité de Cauchy-Schwarz).

On suppose $\rho(X_1, X_2) = 1$. Pour tout $\lambda \in \mathbb{R}$,

$$E \left[\left(\frac{X_1 - E[X_1]}{\sqrt{Var(X_1)}} + \lambda \frac{X_2 - E[X_2]}{\sqrt{Var(X_2)}} \right)^2 \right] = 1 + 2\lambda + \lambda^2 = (1 + \lambda)^2.$$

En prenant $\lambda = -1$, on obtient que

$$\frac{X_1 - E[X_1]}{\sqrt{Var(X_1)}} = \frac{X_2 - E[X_2]}{\sqrt{Var(X_2)}} \text{ p.s.}$$

Lorsque $\rho(X_1, X_2) = -1$, on prend $\lambda = 1$, et on obtient que

$$\frac{X_1 - E[X_1]}{\sqrt{Var(X_1)}} = -\frac{X_2 - E[X_2]}{\sqrt{Var(X_2)}} \text{ p.s.}$$

3.5.2 Variables (ou vecteurs) aléatoires indépendant(e)s

Rappels : événements indépendants, classes d'événements indépendants.

Deux événements A_1 et A_2 de \mathcal{A} sont dits indépendants si $P(A_1 \cap A_2) = P(A_1)P(A_2)$.

Soit $(A_i)_{i \in I}$ une suite (finie ou infinie) d'événements distincts. Les événements $A_i, i \in I$, sont dits (mutuellement) indépendants si pour toute collection finie A_{i_1}, \dots, A_{i_k} d'événements distincts, on a

$$P(A_{i_1} \dots A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

Deux classes d'événements \mathcal{C}_1 et \mathcal{C}_2 sont indépendantes si pour tout événement A_1 de \mathcal{C}_1 , et tout événement A_2 de \mathcal{C}_2 , A_1 et A_2 sont indépendants.

Soit $(\mathcal{C}_i)_{i \in I}$ une suite (finie ou infinie) de classes d'événements de \mathcal{A} . On dit que les classes \mathcal{C}_i , $i \in I$, sont (mutuellement) indépendantes si pour toute suite finie $\mathcal{C}_{i_1}, \dots, \mathcal{C}_{i_k}$ extraite de la suite $(\mathcal{C}_i)_{i \in I}$ et pour toute suite d'événements $A_{i_1} \in \mathcal{C}_{i_1}, \dots, A_{i_k} \in \mathcal{C}_{i_k}$, on a

$$P(A_{i_1} \dots A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

Théorème : Si les classes $(\mathcal{C}_i)_{i \in I}$ sont indépendantes et stables par intersection alors les tribus engendrées par les $(\mathcal{C}_i)_{i \in I}$ sont indépendantes.

Rappel : tribu engendrée par une variable aléatoire ou un vecteur aléatoire X . $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}^n))$.

On peut donc prolonger la notion d'indépendance aux variables et aux vecteurs aléatoires.

Définition 26 Deux variables aléatoires réelles ou deux vecteurs aléatoires X et Y définis sur (Ω, \mathcal{A}, P) sont dits indépendants si les tribus $\sigma(X)$ et $\sigma(Y)$ sont indépendantes, ou de façon plus explicite, lorsque X est à n dimensions, Y à m dimensions, si pour tout $A \in \mathcal{B}(\mathbb{R}^n)$, $B \in \mathcal{B}(\mathbb{R}^m)$, on a :

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Définition 27 Soit $(X_i)_{i \in I}$ une suite de variables ou vecteurs aléatoires définis sur (Ω, \mathcal{A}, P) . Les variables ou vecteurs aléatoires X_i , $i \in I$ sont dits (mutuellement) indépendants si les tribus engendrées par ces variables ou vecteurs $(\sigma(X_i))_{i \in I}$ sont (mutuellement) indépendantes, ou de façon plus explicite, si pour toute suite finie extraite X_{i_1}, \dots, X_{i_k} , et toute suite finie B_{i_1}, \dots, B_{i_k} de boréliens,

$$P(X_{i_1} \in B_{i_1}, \dots, X_{i_k} \in B_{i_k}) = P(X_{i_1} \in B_{i_1}) \dots P(X_{i_k} \in B_{i_k}).$$

Proposition 27 Soit $(X_i)_{i \in I}$ une suite de vecteurs aléatoires à $(m_i)_{i \in I}$ dimensions respectivement, définis sur (Ω, \mathcal{A}, P) , indépendants, et soit pour tout $i \in I$, g_i une fonction mesurable de $(\mathbb{R}^{m_i}, \mathcal{B}(\mathbb{R}^{m_i}))$ dans $(\mathbb{R}^{p_i}, \mathcal{B}(\mathbb{R}^{p_i}))$. Alors les vecteurs aléatoires $(g_i(X_i))_{i \in I}$ à $(p_i)_{i \in I}$ dimensions respectivement sont indépendants.

Preuve. Basée sur l'idée que pour tout $i \in I$, $(g_i \circ X_i)^{-1}(B_i) = X_i^{-1}(g_i^{-1}(B_i))$.

On considère dans toute la suite un vecteur aléatoire à n dimensions $X = (X_1, \dots, X_n)$, de loi conjointe P_X , de fonction de répartition notée F_X . On note P_{X_1}, \dots, P_{X_n} les lois marginales et F_{X_1}, \dots, F_{X_n} les fonctions de répartition marginales de X_1, \dots, X_n .

Les théorèmes suivants donnent des conditions simples pour vérifier que les variables aléatoires réelles X_1, \dots, X_n sont indépendantes.

Théorème 9 Les variables marginales X_1, \dots, X_n sont (mutuellement) indépendantes si et seulement si pour tous réels x_1, \dots, x_n ,

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n),$$

ou encore si et seulement si

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

Preuve. Basée sur l'idée que les classes d'événements $(X_i \leq x_i)$ sont stables par intersection et que les tribus engendrées par ces événements sont les $\sigma(X_i)$.

Théorème 10 Les variables marginales X_1, \dots, X_n sont (mutuellement) indépendantes si et seulement si la loi conjointe P_X est le produit des lois marginales P_{X_1}, \dots, P_{X_n} . En conséquence (Fubini), si B_1, \dots, B_n sont des boréliens de \mathbb{R} , et φ une fonction réelle P_X intégrable, alors

$$\int \dots \int_{B_1 \times \dots \times B_n} \varphi(x_1, \dots, x_n) dP_X(x_1, \dots, x_n) = \int_{B_1} \dots \int_{B_n} \varphi(x_1, \dots, x_n) dP_{X_1}(x_1) \dots dP_{X_n}(x_n).$$

Dans le cas des vecteurs aléatoires de loi absolument continue, on peut montrer à l'aide du théorème de Fubini-Tonelli le résultat suivant sur les densités.

Proposition 28 *Si les variables marginales X_1, \dots, X_n sont indépendantes, de lois marginales absolument continues, si l'on note f_{X_1}, \dots, f_{X_n} les densités marginales respectives, alors la densité conjointe de X est de la forme :*

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n).$$

Réciproquement, si le vecteur aléatoire $X = (X_1, \dots, X_n)$ possède une densité conjointe de la forme

$$f_X(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n),$$

alors les variables marginales X_1, \dots, X_n sont indépendantes, et leurs densités respectives sont égales à une constante multiplicative près à f_1, \dots, f_n .

Le théorème (loi produit) ci-dessus permet de montrer les résultats, très utiles, suivants.

Proposition 29 *Soit X_1 et X_2 deux variables aléatoires réelles indépendantes admettant des espérances mathématiques. Alors le produit $X_1 X_2$ admet aussi une espérance mathématique et :*

$$E[X_1 X_2] = E[X_1]E[X_2].$$

Proposition 30 *Soit X_1 et X_2 deux variables aléatoires réelles indépendantes ayant des moments d'ordre deux finis. Alors*

$$\text{Cov}(X_1, X_2) = 0.$$

On peut retenir que deux variables aléatoires réelles indépendantes sont non corrélées, mais **attention** : la réciproque est fautive. On peut regarder ce contre-exemple : soit X la variable aléatoire de loi $P_X = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$, et $Y = X^2$. La loi du couple $Z = (X, Y)$ est donnée par $P_Z = \frac{1}{3}(\delta_{(-1,1)} + \delta_{(0,0)} + \delta_{(1,1)})$, on a $E[X] = 0$, $E[XY] = 0$, d'où $\text{Cov}(X, Y) = 0$ alors que $Y = X^2$.

On en déduit facilement que la variance d'une somme de variables aléatoires réelles indépendantes est égale à la somme des variances.

Proposition 31 *Si X_1, \dots, X_n sont des variables aléatoires réelles, ayant des moments d'ordre deux finis, (mutuellement) indépendantes ou même seulement deux à deux non corrélées, alors*

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

3.6 Calcul de lois

On considère dans ce paragraphe un espace de probabilité (Ω, \mathcal{A}, P) et un vecteur aléatoire à n dimensions $X = (X_1, \dots, X_n)$ défini sur (Ω, \mathcal{A}, P) , de loi conjointe P_X , de fonction de répartition conjointe F_X .

Il s'agit ici d'apprendre à déterminer pour une fonction mesurable $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^p$, la loi du vecteur aléatoire (ou si $p = 1$, de la variable aléatoire réelle) $Y = \varphi(X)$.

3.6.1 Première méthode : utilisation de la fonction de répartition conjointe

On cherche ici à exprimer la fonction de répartition F_Y de Y en fonction de F_X , en utilisant par exemple le théorème de transfert appliqué à une indicatrice :

$$P(Y_1 \leq y_1, \dots, Y_p \leq y_p) = P(\varphi(X) \in]-\infty, y_1] \times \dots \times]-\infty, y_p]) = \int \dots \int \mathbb{I}_{\varphi(x) \in]-\infty, y_1] \times \dots \times]-\infty, y_p]} dP_X(x_1, \dots, x_n).$$

Si la loi de $Y = \varphi(X)$ est discrète, on peut en déduire la fonction de masse. Si la loi de $Y = \varphi(X)$ est absolument continue, on peut dériver F_Y par rapport à ses p variables et en déduire la densité de Y .

3.6.2 Deuxième méthode : détermination directe de la fonction de masse dans le cas d'une loi discrète

3.6.3 Troisième méthode : détermination directe de la densité dans le cas d'une loi absolument continue

Supposons que X est à valeurs dans un ouvert $U \subset \mathbb{R}^n$.

Supposons tout d'abord également que φ est une bijection de U dans un ouvert V de \mathbb{R}^n , ayant des dérivées partielles continues. On note J_ψ le jacobien de la bijection inverse $\psi = \varphi^{-1}$, i.e. si

$$\begin{aligned} \varphi : (x_1, \dots, x_n) &\mapsto (y_1, \dots, y_n) = (\varphi_1(x_1, \dots, x_n), \dots, \varphi_n(x_1, \dots, x_n)), \\ \psi = \varphi^{-1} : (y_1, \dots, y_n) &\mapsto (x_1, \dots, x_n) = (\psi_1(y_1, \dots, y_n), \dots, \psi_n(y_1, \dots, y_n)), \end{aligned}$$

$$J_\psi = \det \begin{bmatrix} \frac{\partial \psi_1(y_1, \dots, y_n)}{\partial y_1} & \dots & \frac{\partial \psi_1(y_1, \dots, y_n)}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial \psi_n(y_1, \dots, y_n)}{\partial y_1} & \dots & \frac{\partial \psi_n(y_1, \dots, y_n)}{\partial y_n} \end{bmatrix}.$$

Si $J_\psi \neq 0$ (sauf éventuellement sur un ensemble négligeable I tel que $\psi(I)$ est aussi négligeable), alors $Y = (Y_1, \dots, Y_n) = (\varphi_1(X_1, \dots, X_n), \dots, \varphi_n(X_1, \dots, X_n))$ est un vecteur aléatoire de loi absolument continue, dont la densité conjointe est donnée par :

$$f_Y(y_1, \dots, y_n) = f_X(\psi_1(y_1, \dots, y_n), \dots, \psi_n(y_1, \dots, y_n)) |J_\psi| \mathbb{1}_V(y_1, \dots, y_n).$$

Ce résultat est basé sur un théorème d'identification, couplé à un changement de variables sur des intégrales multiples, comme pour les variables aléatoires réelles.

Supposons maintenant que $\varphi : (x_1, \dots, x_n) \in U \mapsto (\varphi_1(x_1, \dots, x_n), \dots, \varphi_p(x_1, \dots, x_n))$ est à valeurs dans \mathbb{R}^p , avec $p < n$. On introduit alors une fonction $\tilde{\varphi} : (x_1, \dots, x_n) \in U \mapsto (\tilde{\varphi}_1(x_1, \dots, x_n), \dots, \tilde{\varphi}_n(x_1, \dots, x_n))$ qui est une bijection de U dans un ouvert V de \mathbb{R}^n et telle que $\tilde{\varphi}_1 = \varphi_1, \dots, \tilde{\varphi}_p = \varphi_p$. On obtient par le résultat précédent, sous les conditions de régularité adéquates, l'expression de la densité conjointe de $\tilde{\varphi}(X) = (\varphi_1(X_1, \dots, X_n), \dots, \varphi_p(X_1, \dots, X_n), \tilde{\varphi}_{p+1}(X_1, \dots, X_n), \dots, \tilde{\varphi}_n(X_1, \dots, X_n))$, et en intégrant cette densité conjointe par rapport à ses $(n-p)$ dernières variables, on en déduit la densité conjointe de $(\varphi_1(X_1, \dots, X_n), \dots, \varphi_p(X_1, \dots, X_n))$.

Dans tous les autres cas, on peut revenir à la première méthode.

Exemple : cas d'un changement en polaires.

Autre exemple fondamental : la densité d'une somme de deux variables aléatoires réelles **indépendantes** X_1 et X_2 , de densités respectives f_1 et f_2 , est donnée par le **produit de convolution** $f_1 * f_2(u) = \int f_1(u-v)f_2(v)d\lambda(v)$.

4 Lois usuelles dans \mathbb{R}

4.1 Lois discrètes

Loi	Paramètre(s)	Fonction de masse	Espérance	Variance	Modélisation	Observations
Dirac δ_a	a	$\pi(x) = \mathbb{I}_a(x), x \in \mathbb{R}$	a	0	X est une variable aléatoire prenant la valeur a quel que soit le résultat de l'expérience.	Fonction caractéristique (F.c.) $\varphi(t) = e^{ita}$.
Uniforme sur $\{1, \dots, n\}$		$\pi(x) = \frac{1}{n}, x \in \{1, \dots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	X est une variable aléatoire désignant un nombre entier compris entre 1 et n choisi au hasard, de façon équiprobable.	
Bernoulli $\mathcal{B}(p)$	$p \in [0, 1]$	$\pi(x) = p^x(1-p)^{1-x}, x \in \{0, 1\}$	p	$p(1-p)$	Dans une expérience à deux issues - succès et échec - avec une probabilité de succès égale à p , X est la variable aléatoire qui vaut 1 si l'expérience conduit à un succès, 0 sinon.	Si $X_i \sim \mathcal{B}(p)$ pour $i = 1, \dots, n$, $\{X_i, i = 1, \dots, n\}$ indépendantes, $\sum_{i=1}^n X_i \sim \mathcal{B}(n, p)$. F.c. $\varphi(t) = pe^{it} + (1-p)$.
Binomiale $\mathcal{B}(n, p)$	$n \in \mathbb{N}^*, p \in [0, 1]$	$\pi(x) = \binom{n}{x} p^x(1-p)^{n-x}, x \in \{0, 1, \dots, n\}$	np	$np(1-p)$	On répète n fois l'expérience à deux issues de la loi de Bernoulli de façon indépendante. X est la variable aléatoire correspondant au nombre total de succès obtenus.	Si $X_1 \sim \mathcal{B}(n_1, p), X_2 \sim \mathcal{B}(n_2, p)$, X_1 et X_2 indépendantes, alors $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$. Si $X \sim \mathcal{B}(n, p)$, alors $n - X \sim \mathcal{B}(n, 1-p)$. F.c. $\varphi(t) = (pe^{it} + (1-p))^n$.
Poisson $\mathcal{P}(\lambda)$	$\lambda > 0$	$\pi(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x \in \mathbb{N}$	λ	λ	Loi limite d'une binomiale de paramètres (n, p) lorsque $n \rightarrow +\infty, p \rightarrow 0, np \rightarrow \lambda$.	Si $X_1 \sim \mathcal{P}(\lambda_1), X_2 \sim \mathcal{P}(\lambda_2)$, X_1 et X_2 indépendantes, alors $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$. F.c. $\varphi(t) = e^{\lambda(e^{it}-1)}$.
Géométrique $\mathcal{G}(p)$	$p \in [0, 1]$	$\pi(x) = (1-p)^{x-1}p, x \in \{1, 2, \dots\}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	On répète plusieurs fois l'expérience à deux issues de la loi de Bernoulli, de façon indépendante. X est la variable aléatoire correspondant au nombre de fois où il a fallu répéter l'expérience pour obtenir un succès.	F.c. $\varphi(t) = pe^{it}/(1-(1-p)e^{it})$.

Loi	Paramètre(s)	Fonction de masse	Espérance	Variance	Modélisation	Observations
Binomiale négative	$n \in \mathbb{N}^*, p \in [0, 1]$	$\pi(x) = \binom{x-1}{n-1} p^n (1-p)^{x-n},$ $x \in \{n, n+1, \dots\}$	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$	On répète plusieurs fois l'expérience à deux issues de la loi de Bernoulli, de façon indépendante. X est la variable aléatoire correspondant au nombre de fois où il a fallu répéter l'expérience pour obtenir n succès.	Si $n = 1$, on retrouve la loi géométrique. Si X_1 suit une loi binomiale négative de paramètres (n_1, p) , X_2 suit une loi binomiale négative de paramètres (n_2, p) , X_1 et X_2 indépendantes, alors $X_1 + X_2$ suit une loi binomiale négative de paramètres $(n_1 + n_2, p)$.
Hypergéométrique	$N \in \mathbb{N}^*,$ $N_1 \in \{1, \dots, N\},$ $n \in \{1, \dots, N\}$	$\pi(x) = \frac{\binom{N_1}{x} \binom{N-N_1}{n-x}}{\binom{N}{n}},$ $x \in \mathbb{N}, x \geq \max(0, n-N+N_1),$ $x \leq \min(n, N_1)$	$\frac{nN_1}{N}$	$\frac{nN_1(N-N_1)(N-n)}{N^2(N-1)}$	On tire n boules sans remise dans une urne contenant N_1 boules blanches, $N - N_1$ boules noires. X est la variable aléatoire correspondant au nombre total de boules blanches tirées.	On retrouve la loi binomiale en faisant tendre N_1 et N vers $+\infty$, $\frac{N_1}{N}$ vers p .

4.2 Loix absolument continues

Loi	Paramètre(s)	Densité	Espérance	Variance	Modélisation	Observations
Uniforme $\mathcal{U}([a, b])$	$a, b \in \mathbb{R}, a < b$	$f(x) = \frac{1}{b-a} \mathbb{I}_{[a, b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	X est la variable aléatoire représentant un nombre choisi au hasard entre a et b .	Si $X \sim \mathcal{U}([0, 1])$, si F^{-1} désigne l'inverse généralisée de la fonction de répartition d'une loi P , alors $F^{-1}(X) \sim P$. F.c. $\varphi(t) = \sin(at)/(at)$ pour $\mathcal{U}([-a, a])$.
Exponentielle $\mathcal{E}(\lambda)$	$\lambda > 0$	$f(x) = \lambda e^{-\lambda x} \mathbb{I}_{[0, +\infty[}(x)$	$1/\lambda$	$1/\lambda^2$		Si $X \sim \mathcal{E}(\lambda), X \sim \gamma(1, \lambda)$. F.c. $\varphi(t) = \lambda/(\lambda - it)$.
Gamma $\gamma(p, \lambda)$	$p > 0, \lambda > 0$	$f(x) = \frac{\lambda^p}{\Gamma(p)} e^{-\lambda x} (\lambda x)^{p-1} \mathbb{I}_{[0, +\infty[}(x)$	p/λ	p/λ^2		Si $X_1 \sim \gamma(p_1, \lambda), X_2 \sim \gamma(p_2, \lambda), X_1$ et X_2 indépendantes, alors $X_1 + X_2 \sim \gamma(p_1 + p_2, \lambda)$. Pour $n \in \mathbb{N}^*, \Gamma(n) = (n-1)!$
Beta I $\beta_1(a, b)$	$a > 0, b > 0$	$f(x) = \frac{1}{\beta(a, b)} (1-x)^{b-1} x^{a-1} \mathbb{I}_{[0, 1]}(x)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b+1)(a+b)^2}$	Si $X \sim \gamma(a, 1), Y \sim \gamma(b, 1), X$ et Y indépendantes, alors $\frac{X}{X+Y} \sim \beta_1(a, b)$.	$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Les statistiques d'ordre d'une loi $\mathcal{U}([0, 1])$ suivent des lois Beta I.
Beta II $\beta_2(a, b)$	$a > 0, b > 0$	$f(x) = \frac{1}{\beta(a, b)} \frac{x^{a-1}}{(1+x)^{a+b}} \mathbb{I}_{[0, +\infty[}(x)$	$\frac{a}{b-1}$ si $b > 1$	$\frac{a(a+b-1)}{(b-1)^2(b-2)}$ si $b > 2$	Si $X \sim \gamma(a, 1), Y \sim \gamma(b, 1), X$ et Y indépendantes, alors $\frac{X}{Y} \sim \beta_2(a, b)$.	Si $X \sim \beta_1(a, b), \frac{X}{1-X} \sim \beta_2(a, b)$ et réciproquement.
Weibull $W(a, \lambda)$	$a > 1, \lambda > 0$	$f(x) = a\lambda x^{a-1} e^{-\lambda x^a} \mathbb{I}_{[0, +\infty[}(x)$	$\frac{\Gamma(1+\frac{1}{a})}{\lambda^{\frac{1}{a}}}$	$\frac{\Gamma(1+\frac{2}{a}) - \Gamma^2(1+\frac{1}{a})}{\lambda^{\frac{2}{a}}}$	$X \sim W(a, \lambda)$ si $X^a \sim \mathcal{E}(\lambda)$	
Gaussienne ou normale $\mathcal{N}(m, \sigma^2)$	$m \in \mathbb{R}, \sigma^2 > 0$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	Loi limite du théorème central limite.	Si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$. Si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2), X_2 \sim \mathcal{N}(m_2, \sigma_2^2), X_1$ et X_2 indépendantes, alors $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. F.c. $\varphi(t) = e^{imt - \frac{1}{2}\sigma^2 t^2}$.

Lois issues des lois gaussiennes

Loi	Paramètre(s)	Densité	Espérance	Variance	Modélisation	Observations
Cauchy		$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$			Si X_1 et X_2 sont indépendantes, de même loi $\mathcal{N}(0, 1)$, si $U \sim \mathcal{U}(\frac{\pi}{2}, \frac{3\pi}{2})$, X_1/X_2 et $\tan U$ suivent la loi de Cauchy	F.c. $\varphi(t) = e^{- t }$.
Log-normale	$m \in \mathbb{R}, \sigma^2 > 0$	$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}} \mathbb{I}_{]0, +\infty[}(x)$	$e^{m+\sigma^2/2}$	$e^{2m+\sigma^2} (e^{\sigma^2} - 1)$	X suit la loi log-normale de paramètres (m, σ^2) si $\ln X \sim \mathcal{N}(m, \sigma^2)$.	
Chi-Deux $\chi^2(n)$	$n \in \mathbb{N}^*$	$f(x) = \frac{1}{2\Gamma(n/2)} e^{-x/2} \left(\frac{x}{2}\right)^{n/2-1} \mathbb{I}_{]0, +\infty[}(x)$	n	$2n$	Si X_1, \dots, X_n sont i.i.d. de loi $\mathcal{N}(0, 1)$, $X = X_1^2 + \dots + X_n^2 \sim \chi^2(n)$.	Si $X_1 \sim \chi^2(n_1), X_2 \sim \chi^2(n_2)$, X_1 et X_2 indépendantes, alors $X_1 + X_2 \sim \chi^2(n_1 + n_2)$. Si $X \sim \chi^2(n), X \sim \gamma(n/2, 1/2)$.
Student $\mathcal{T}(n)$	$n \in \mathbb{N}^*$	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$	0 si $n > 1$	$\frac{n}{n-2}$ si $n > 2$	Si $X \sim \mathcal{N}(0, 1), Y \sim \chi^2(n), X$ et Y indépendantes, alors $\frac{X}{\sqrt{Y/n}} \sim \mathcal{T}(n)$.	Si $X \sim \mathcal{T}(n), \frac{X^2}{n} \sim \beta_2\left(\frac{1}{2}, \frac{n}{2}\right)$.
Fisher-Snedecor $\mathcal{F}(m, n)$	$m \in \mathbb{N}^*, n \in \mathbb{N}^*$	$f(x) = \frac{m^{\frac{m}{2}} n^{\frac{n}{2}}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)} x^{\frac{m}{2}-1} (n + mx)^{-\frac{n+m}{2}}$	$\frac{n}{n-2}$ si $n > 2$	$\frac{2m^2(m+n-2)}{m(n-2)^2(n-4)}$ si $n > 4$	Si $X \sim \chi^2(m), Y \sim \chi^2(n), X$ et Y indépendantes, $\frac{X/m}{Y/n} \sim \mathcal{F}(m, n)$.	Si $X \sim \mathcal{F}(m, n)$, alors $1/X \sim \mathcal{F}(n, m)$.

5 Lois usuelles dans \mathbb{R}^n

5.1 La loi multinomiale

On répète n fois une expérience à k issues, de probabilités respectives p_1, \dots, p_k , de façon indépendante, et on considère pour tout $i = 1, \dots, k$, le nombre X_i de réalisations de l'issue i (parmi les n répétitions).

Définition 28 On dit que le vecteur aléatoire $X = (X_1, \dots, X_k)$ suit une loi multinomiale de paramètres (n, p_1, \dots, p_k) , et on note $X \sim \mathcal{M}(n, p_1, \dots, p_k)$. On a

$$P((X_1, \dots, X_k) = (x_1, \dots, x_k)) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}.$$

Propriétés.

Les variables marginales X_1, \dots, X_k sont linéairement dépendantes par construction : $\sum_{i=1}^k X_i = n$, et on a bien sûr $\sum_{i=1}^k p_i = 1$.

Chaque variable marginale X_i suit une loi binomiale $\mathcal{B}(n, p_i)$.

On a donc $E[X] = (np_1, \dots, np_k)$, et la matrice de variances-covariances de X est égale à

$$\Sigma_X = \begin{pmatrix} np_1(1-p_1) & -np_1p_2 & \dots & -np_1p_k \\ -np_1p_2 & np_2(1-p_2) & \dots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_1p_k & -np_2p_k & \dots & np_k(1-p_k) \end{pmatrix}.$$

Remarque : Cette matrice Σ_X n'est pas inversible.

La loi multinomiale a toute son importance en statistique. Elle est notamment à la base d'un test très utilisé en statistique : le test du Chi-deux.

5.2 Vecteurs gaussiens - loi multinormale

Retour sur la loi gaussienne dans \mathbb{R} : Densités, forme des densités, centrage et réduction d'une v.a.r. gaussienne, quantiles des lois gaussiennes, loi gaussienne dégénérée.

5.2.1 Définitions - Premières propriétés

Définition 29 Un vecteur aléatoire de dimension k $X = (X_1, \dots, X_k)$ est dit gaussien si toute combinaison linéaire de ses variables marginales $a_1X_1 + \dots + a_kX_k$ est une v.a.r. gaussienne.

Proposition 32 La loi d'un vecteur aléatoire $X = (X_1, \dots, X_k)$ gaussien est entièrement déterminée par la donnée de son espérance $m_X = E[X]$ et de sa matrice de variances-covariances Σ_X . On note alors $X \sim \mathcal{N}_k(m_X, \Sigma_X)$. La fonction caractéristique de la loi $\mathcal{N}_k(m_X, \Sigma_X)$ est donnée par $\varphi_X(t) = e^{it'm_X} e^{-\frac{1}{2}t'\Sigma t}$.

Proposition 33 Soit $X = (X_1, \dots, X_k)$ un vecteur gaussien. Par définition, les variables marginales X_1, \dots, X_k sont des v.a.r. gaussiennes. Mais la réciproque est fautive.

Preuve. On prend $X_1 \sim \mathcal{N}(0, 1)$, ε une variable aléatoire de Rademacher, c'est-à-dire qui prend les valeurs 1 et -1 avec probabilité $1/2$, et on pose $X_2 = \varepsilon X_1$. On peut voir que X_1 et X_2 sont des v.a.r. gaussiennes centrées réduites, mais que $X_1 + X_2$ n'est pas une variable gaussienne (notamment, on a $P(X_1 + X_2 = 0) \neq 0!$). Par conséquent (X_1, X_2) n'est pas un vecteur gaussien.

Proposition 34 Si X_1, \dots, X_k sont des v.a.r. gaussiennes **indépendantes**, alors le vecteur aléatoire $X = (X_1, \dots, X_k)$ est un vecteur gaussien.

Définition 30 Un vecteur aléatoire $X = (X_1, \dots, X_k)$ est dit gaussien centré réduit ou standard si $X \sim \mathcal{N}_k(0, I_k)$.

Proposition 35 $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(m_X, \Sigma_X)$ si et seulement si X peut s'écrire $X = AY + m_X$, avec $Y \sim \mathcal{N}_k(0, I_k)$, A étant de taille $k \times k$, satisfaisant $AA' = \Sigma_X$ et $\text{rang}(A) = \text{rang}(\Sigma_X)$.

Si $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(m_X, \Sigma_X)$, A une matrice de taille $l \times k$ et $B \in \mathbb{R}^l$, alors $AX + B \sim \mathcal{N}_l(Am_X + B, A\Sigma_X A')$. En particulier, le vecteur $\Sigma_X^{-1/2}(X - m_X)$ est gaussien standard.

5.2.2 Indépendance

Proposition 36 Soit $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(m_X, \Sigma_X)$. Les variables marginales X_1, \dots, X_k sont indépendantes si et seulement si Σ_X est diagonale, autrement dit si et seulement si elles sont non corrélées.

Soit $Z = (X_1, \dots, X_p, Y_1, \dots, Y_q)$ un vecteur gaussien. Les vecteurs (X_1, \dots, X_p) et (Y_1, \dots, Y_q) sont indépendants si et seulement s'ils sont non corrélés.

Preuve. Utilisation de la fonction caractéristique.

Attention : on rappelle que ces propriétés sont fausses dans le cadre général des vecteurs aléatoires ! Si on reprend l'exemple ci-dessus de $X_1 \sim \mathcal{N}(0, 1)$, et $X_2 = \varepsilon X_1$, alors les variables X_1 et X_2 sont non corrélées, mais dépendantes.

5.2.3 Densité

Proposition 37 Soit $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(m_X, \Sigma_X)$. X admet une densité f si et seulement si $\det \Sigma_X \neq 0$, et on a

$$f(x_1, \dots, x_k) = \frac{1}{(\sqrt{2\pi})^k \sqrt{\det \Sigma_X}} e^{-\frac{1}{2}(x - m_X)' \Sigma_X^{-1} (x - m_X)},$$

avec $x = (x_1, \dots, x_k)'$.

Preuve. On a $Y = \Sigma_X^{-1/2}(X - m_X) \sim \mathcal{N}_k(0, I_k)$, donc ses variables marginales Y_1, \dots, Y_k sont des v.a.r. gaussiennes centrées réduites indépendantes. Y a donc pour densité

$$g(y) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2} \sum_{i=1}^k y_i^2}.$$

Il suffit alors d'appliquer la formule du changement de variables. Le jacobien vaut $1/(\det \Sigma_X^{1/2}) = (\det \Sigma_X)^{-1/2}$, ce qui donne le résultat.

Exemple. Cas du vecteur gaussien de dimension 2 et lien avec le coefficient de corrélation linéaire.

Définition 31 Un vecteur gaussien X dont la matrice de variances-covariances a un déterminant nul est dit dégénéré.

5.2.4 Vecteurs gaussiens et loi du Chi-Deux

Rappels sur la loi du Chi-Deux.

Proposition 38 Si $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(m_X, \Sigma_X)$, avec Σ_X inversible, alors $(X - m_X)' \Sigma_X^{-1} (X - m_X)$ suit une loi du $\chi^2(k)$.

Preuve. Il suffit ici aussi de voir que $\Sigma_X^{-1/2}(X - m_X) \sim \mathcal{N}_k(0, I_k)$.

Proposition 39 Soit $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(0, I_k)$.

$X'AX$ suit une loi du χ^2 si et seulement si A est une matrice de projection orthogonale, i.e. $A^2 = A$. Le nombre de degrés de liberté du χ^2 est alors égal au rang de A .

Soit A, B deux matrices de projection orthogonale. $X'AX$ et $X'BX$ sont indépendantes si et seulement si $AB = 0$.

Théorème 11 (Théorème de Cochran) Soit $X = (X_1, \dots, X_k) \sim \mathcal{N}_k(0, I_k)$. Soit A_1, \dots, A_p p matrices symétriques de taille $k \times k$ telles que $\sum_{l=1}^p X' A_l X = X' X$ (décomposition de la norme de X au carré). Alors les trois conditions suivantes sont équivalentes :

- $\sum_{l=1}^p \text{rang}(A_l) = k$.
- Pour tout $l = 1, \dots, p$, $X' A_l X \sim \chi^2(\text{rang } A_l)$.
- Les $X' A_l X$ sont indépendantes.

Application aux projections sur des sous-espaces vectoriels de \mathbb{R}^k .

6 Conditionnement, espérance et variance conditionnelles

On considère un espace de probabilité (Ω, \mathcal{A}, P) .

Rappel de probabilités de base : probabilité conditionnelle d'un événement par rapport à un autre.

Soit A et B deux événements de (Ω, \mathcal{A}, P) avec $P(B) > 0$. On s'intéresse à la réalisation de l'événement A , sachant que l'événement B est réalisé.

Si A et B sont incompatibles, A ne se réalisera pas, mais si $A \cap B \neq \emptyset$, il est possible que A se réalise. Cependant, l'univers des possibles n'est plus Ω tout entier : il est restreint à B , ce qui justifie la définition suivante : on appelle probabilité conditionnelle de A sachant B le réel

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

L'application $A \mapsto P(A|B)$ est une mesure de probabilité sur (Ω, \mathcal{A}) .

Les événements A et B sont donc indépendants si et seulement si $P(A|B) = P(A)$ (autrement dit la connaissance de B n'influe pas sur la probabilité de $A \Rightarrow$ vision plus intuitive de l'indépendance), ou si $P(A) > 0$, $P(B|A) = P(B)$.

Formule de Bayes :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

Formule des probabilités totales : Soit $\{B_i, i \in \mathcal{I}\}$ un ensemble d'événements disjoints tels que $\Omega = \cup_{i \in \mathcal{I}} B_i$. Alors

$$P(A) = \sum_{i \in \mathcal{I}} P(A|B_i)P(B_i).$$

On va chercher à étendre cette notion de probabilité conditionnelle. Pour cela, on commence par étudier la notion d'espérance conditionnelle, qui s'avère être souvent beaucoup plus utile en statistique (et qui est plus facile à définir).

6.1 Espérance conditionnelle : le cas général

On considère une variable aléatoire réelle Y définie sur (Ω, \mathcal{A}, P) , telle que $E[|Y|] < +\infty$, et une sous-tribu \mathcal{S} de \mathcal{A} .

On suppose ici que l'on ne s'intéresse qu'aux événements de la sous-tribu \mathcal{S} , et on se demande si l'on peut trouver une version simplifiée de Y , qui résumerait l'information contenue dans Y , sachant que l'on se restreint aux événements de \mathcal{S} .

Définition 32 On appelle espérance conditionnelle de Y sachant \mathcal{S} la classe des variables aléatoires Z telles que :

- (i) Z est \mathcal{S} -mesurable.
 - (ii) Pour tout $A \in \mathcal{S}$, $\int Y \mathbb{I}_A dP = \int Z \mathbb{I}_A dP$ (ou $E[Y \mathbb{I}_A] = E[Z \mathbb{I}_A]$).
- On la note $E[Y|\mathcal{S}]$.

Une variable aléatoire Z qui vérifie (i) et (ii) est appelée une version de l'espérance conditionnelle de Y sachant \mathcal{S} .

La démonstration d'existence repose sur le théorème de Radon-Nikodym.

Deux versions de l'espérance conditionnelle de Y sachant \mathcal{S} sont égales P -p.s.

Conséquence importante : On confond souvent la classe avec l'une quelconque de ses versions : $E[Y|\mathcal{S}]$ est vue comme une v.a.r. "définie P -p.s.", et c'est le "résumé" cherché de Y .

Dans toute la suite (sauf lorsqu'on parlera de probabilité conditionnelle), on considèrera donc $E[Y|\mathcal{S}]$ comme une v.a.r. (définie P -p.s.), et les (in)égalités seront des (in)égalités P -p.s.

Remarque. Si $Y = (Y_1, \dots, Y_n)$ est un vecteur aléatoire de dimension n , on note $E[Y|\mathcal{S}]$ le vecteur aléatoire $(E[Y_1|\mathcal{S}], \dots, E[Y_n|\mathcal{S}])$.

Les propriétés de l'espérance conditionnelle dans le cas général sont données ici sans démonstration. Elles peuvent être démontrées en exercice (voir aussi le livre d'A. Monfort).

Proposition 40 (Propriétés générales de l'espérance conditionnelle)

1. *Rappel :* $E[Y|\mathcal{S}]$ est une v.a.r. \mathcal{S} -mesurable, et pour tout $A \in \mathcal{S}$, $\int E[Y|\mathcal{S}]\mathbb{I}_A dP = \int Y\mathbb{I}_A dP$.
2. L'espérance conditionnelle possède les propriétés de positivité et de linéarité : si $Y \geq 0$ alors $E[Y|\mathcal{S}] \geq 0$, et pour tout $(a, b) \in \mathbb{R}^2$, si Y_1 et Y_2 sont des v.a.r. intégrables, $E[aY_1 + bY_2|\mathcal{S}] = aE[Y_1|\mathcal{S}] + bE[Y_2|\mathcal{S}]$.
3. $E[E[Y|\mathcal{S}]] = E[Y]$.
4. Si Y est \mathcal{S} -mesurable, alors $E[Y|\mathcal{S}] = Y$.
5. Si Y est \mathcal{S} -mesurable, et si Z est une v.a.r. intégrable telle que $E[|YZ|] < +\infty$, on a $E[YZ|\mathcal{S}] = YE[Z|\mathcal{S}]$.
6. Si $\mathcal{S}' \subset \mathcal{S}$, $E[E[Y|\mathcal{S}]|\mathcal{S}'] = E[E[Y|\mathcal{S}']|\mathcal{S}] = E[Y|\mathcal{S}']$ (on retrouve la propriété 3. avec $\mathcal{S}' = \{\Omega, \emptyset\}$).
7. Si Y est \mathcal{S}' mesurable, et si \mathcal{S}' et \mathcal{S} sont indépendantes, alors $E[Y|\mathcal{S}] = E[Y]$.
8. Inégalité de Jensen. Si Y prend ses valeurs dans un intervalle I de \mathbb{R} , et si ϕ est une fonction convexe sur I telle que $E[|\phi(Y)|] < +\infty$, on a

$$\phi(E[Y|\mathcal{S}]) \leq E[\phi(Y)|\mathcal{S}].$$

Si de plus ϕ est strictement convexe, l'égalité a lieu si et seulement si Y est \mathcal{S} -mesurable.

On peut définir, à partir de l'espérance conditionnelle, la variance conditionnelle.

Définition 33 Si Y admet un moment d'ordre deux, on définit la variance conditionnelle de Y sachant \mathcal{S} par :

$$Var(Y|\mathcal{S}) = E[(Y - E[Y|\mathcal{S}])^2|\mathcal{S}].$$

On a alors le résultat fondamental suivant :

Théorème 12 (Variance totale) $Var(Y) = E[Var(Y|\mathcal{S})] + Var(E[Y|\mathcal{S}])$.

6.2 Probabilité conditionnelle - Version régulière

Définition 34 Si A est un événement de \mathcal{A} , on peut considérer $Y = \mathbb{I}_A$ et on définit ainsi la classe $P(A|\mathcal{S})$ comme $E[\mathbb{I}_A|\mathcal{S}]$, appelée probabilité conditionnelle de A sachant \mathcal{S} .

La question qui se pose alors : peut-on pour chaque $A \in \mathcal{A}$, choisir dans la classe $P(A|\mathcal{S})$ une version $P^*(A|\mathcal{S})$ telle que pour tout $\omega \in \Omega$, $P^*(\cdot|\mathcal{S})(\omega)$ définisse une probabilité sur (Ω, \mathcal{A}) et écrire pour toute v.a.r. intégrable Y :

$$E[Y|\mathcal{S}] = \int_{\Omega} Y(\omega') dP^*(\omega'|\mathcal{S}) \quad P - \text{p.s.}?$$

Définition 35 (Version régulière) On appelle version régulière de la probabilité conditionnelle sachant \mathcal{S} une application $P_{|\mathcal{S}}^* : \mathcal{A} \times \Omega \rightarrow [0, 1]$ telle que :

(i) Pour tout $A \in \mathcal{A}$, $P_{|\mathcal{S}}^*(A, \cdot)$ est une version de $P(A|\mathcal{S})$.

(ii) Pour tout $\omega \in \Omega$, $P_{|\mathcal{S}}^*(\cdot, \omega)$ est une mesure de probabilité sur (Ω, \mathcal{A}) .

On a alors

$$E[Y|\mathcal{S}] = \int_{\Omega} Y(\omega') dP_{|\mathcal{S}}^*(\omega', \cdot) \quad P - p.s..$$

Problème : il n'existe pas forcément de version régulière de la probabilité conditionnelle.

Cependant, on va voir qu'il existe toujours une loi conditionnelle de Y sachant \mathcal{S} , et cela suffira pour calculer $E[Y|\mathcal{S}]$, grâce au théorème de transfert.

Théorème 13 Y étant une v.a.r., il existe une application $P_{Y|\mathcal{S}}^* : \mathcal{B}(\mathbb{R}) \times \Omega \rightarrow [0, 1]$ telle que :

(i) Pour tout $B \in \mathcal{B}(\mathbb{R})$, $P_{Y|\mathcal{S}}^*(B, \cdot)$ est une version de $P(Y \in B|\mathcal{S})$.

(ii) Pour tout $\omega \in \Omega$, $P_{Y|\mathcal{S}}^*(\cdot, \omega)$ est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

On a alors pour toute fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ mesurable, telle que $E[|\phi(Y)|] < +\infty$,

$$E[\phi(Y)|\mathcal{S}] = \int_{\mathbb{R}} \phi(y) dP_{Y|\mathcal{S}}^*(y, \cdot) \quad P - p.s.$$

Application au calcul de la variance conditionnelle.

Preuve de l'existence : Théorème de Jirina (c.f. livre d'A. Monfort ou polycopié de M. Carbon).

Définition 36 (Loi conditionnelle) L'application $P_{Y|\mathcal{S}}^*$ est appelée loi conditionnelle de Y sachant \mathcal{S} . Pour tout $\omega \in \Omega$, si la mesure de probabilité $P_{Y|\mathcal{S}}^*(\cdot, \omega)$ est absolument continue par rapport à la mesure de Lebesgue, on peut introduire la densité de probabilité conditionnelle de Y sachant \mathcal{S} .

6.3 Conditionnement par une variable aléatoire réelle

Un cas particulier très important est celui où la sous-tribu \mathcal{S} est la tribu engendrée par une v.a.r. X définie sur $(\Omega, \mathcal{A}, P) : \mathcal{S} = \sigma(X)$.

On considère une v.a.r. Y définie sur (Ω, \mathcal{A}, P) , telle que $E[|Y|] < +\infty$, et on suppose que $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{X}, \mathcal{X})$.

Définition 37 On appelle espérance conditionnelle de Y sachant X l'espérance conditionnelle de Y sachant la sous-tribu $\sigma(X)$, et on la note $E[Y|X]$.

Proposition 41 (Lemme de Doob) Pour toute v.a.r. Z $\sigma(X)$ -mesurable, il existe une fonction $\varphi : (\mathbb{X}, \mathcal{X}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable telle que $Z = \varphi(X)$.

Conséquence directe : Il existe $\varphi : (\mathbb{X}, \mathcal{X}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable telle que $E[Y|X] = \varphi(X)$.

Définition 38 Pour tout $x \in \mathbb{X}$, l'espérance conditionnelle de Y sachant $X = x$, est alors définie par

$$E[Y|X = x] = \varphi(x).$$

La fonction φ est appelée fonction de régression de Y en X .

Proposition 42 Par le théorème de transfert, on montre que pour tout $A \in \mathcal{X}$, $\int Y \mathbb{I}_{X^{-1}(A)} dP = \int \varphi(x) \mathbb{I}_A dP_X(x)$.

On reprend les propriétés énoncées ci-dessus dans le cas général, auxquelles on ajoute des propriétés spécifiques.

Proposition 43

1. Si $Y \geq 0$ alors $E[Y|X] \geq 0$, et pour tout $(a, b) \in \mathbb{R}^2$, si Y_1 et Y_2 sont des v.a.r. intégrables, $E[aY_1 + bY_2|X] = aE[Y_1|X] + bE[Y_2|X]$.

2. $E[E[Y|X]] = E[Y]$.
3. Pour toute fonction mesurable $\phi : (\mathbb{X}, \mathcal{X}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $E[|\phi(X)|] < +\infty$, $E[\phi(X)|X] = \phi(X)$.
4. Pour toute fonction mesurable $\phi : (\mathbb{X}, \mathcal{X}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, si Z est une v.a.r. intégrable telle que $E[|\phi(X)Z|] < +\infty$, on a $E[\phi(X)Z|X] = \phi(X)E[Z|X]$.
5. Les v.a.r. X et Y sont indépendantes si et seulement si pour toute fonction borélienne bornée ϕ , $E[\phi(Y)|X] = E[\phi(Y)]$.
6. En particulier, si X et Y sont indépendantes, $E[Y|X] = E[Y]$.
7. Inégalité de Jensen. Si Y prend ses valeurs dans un intervalle I de \mathbb{R} , et si ϕ est une fonction convexe sur I telle que $E[|\phi(Y)|] < +\infty$, on a

$$\phi(E[Y|X]) \leq E[\phi(Y)|X].$$

Définition 39 Si A est un événement de \mathcal{A} , on peut considérer $Y = \mathbb{I}_A$ et on définit ainsi la classe $P(A|X)$ comme $E[\mathbb{I}_A|X]$, appelée probabilité conditionnelle de A sachant X . Cette classe est vue dans ce cas particulier comme une classe de fonctions définies sur \mathbb{X} .

On s'intéresse ici à l'existence d'une version régulière, puis d'une loi conditionnelle sachant X .

Définition 40 (Version régulière) On appelle version régulière de la probabilité conditionnelle sachant X une application $P_{|X}^* : \mathcal{A} \times \mathbb{X} \rightarrow [0, 1]$ telle que :

- (i) Pour tout $A \in \mathcal{A}$, $P_{|X}^*(A, \cdot)$ est une version de $P(A|X)$.
- (ii) Pour tout $x \in \mathbb{X}$, $P_{|X}^*(\cdot, x)$ est une mesure de probabilité sur (Ω, \mathcal{A}) .

La version régulière de la probabilité conditionnelle n'existant pas toujours, la proposition suivante montre que l'on pourra néanmoins calculer $E[Y|X = x]$, à l'aide de la loi conditionnelle de Y sachant $X = x$ (théorème de transfert).

Théorème 14 Y étant une v.a.r., il existe une application $P_{Y|X}^* : \mathcal{B}(\mathbb{R}) \times \mathbb{X} \rightarrow [0, 1]$ telle que :

- (i) Pour tout $B \in \mathcal{B}(\mathbb{R})$, $P_{Y|X}^*(B, \cdot)$ est une version de $P(Y \in B|X)$.
- (ii) Pour tout $x \in \mathbb{X}$, $P_{Y|X}^*(\cdot, x)$ est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Définition 41 (Lois conditionnelles) L'application $P_{Y|X}^*$ est appelée loi conditionnelle de Y sachant X , et l'application $P_{Y|X}^*(\cdot, x)$ est appelée loi conditionnelle de Y sachant $X = x$, et notée $P_{Y|X=x}^*$. Pour tout $x \in \mathbb{X}$, si la mesure de probabilité $P_{Y|X=x}^*$ est absolument continue par rapport à la mesure de Lebesgue, on peut introduire la densité de probabilité conditionnelle de Y sachant $X = x$.

On a alors le résultat suivant :

Proposition 44 Si $\phi : \mathbb{X} \times \mathbb{R} \rightarrow \mathbb{R}$ est mesurable, avec $E[|\phi(X, Y)|] < +\infty$,

$$E[\phi(X, Y)|X = x] = \int_{\mathbb{R}} \phi(x, y) dP_{Y|X=x}^*(y) \quad P_X - p.s.$$

Dans les paragraphes suivants, on considère deux exemples fondamentaux.

6.4 Conditionnement par une v.a.r. discrète

Supposons ici que X est une v.a.r. à valeurs dans $(\mathbb{X}, \mathcal{X})$, et que pour tout $x \in \mathbb{X}$, $P(X = x) > 0$. D'après la définition de la probabilité conditionnelle donnée en rappel, pour tout $A \in \mathcal{A}$,

$$P(A|X = x) = \frac{P(A \cap \{X = x\})}{P(X = x)}.$$

Heureusement, on peut vérifier que ceci est cohérent avec notre définition de l'espérance conditionnelle sachant $X = x$: $P(A|X = x) = E[\mathbb{I}_A|X = x]$.

De plus, dans ce cas, l'application $A \mapsto P(A|X = x)$ est une mesure de probabilité et on en déduit que :

$$P_{Y|X=x}^*(B) = P(Y \in B|X = x) = \frac{P(Y \in B, X = x)}{P(X = x)}.$$

On obtient ainsi un calcul effectif de l'espérance conditionnelle :

- Si Y est une variable discrète, pour $\phi : \mathbb{X} \times \mathbb{R} \rightarrow \mathbb{R}$ mesurable, avec $E[|\phi(X, Y)|] < +\infty$,

$$E[\phi(X, Y)|X = x] = \sum_y \phi(x, y) \frac{P(Y = y, X = x)}{P(X = x)}.$$

- Si Y est de loi absolument continue, on introduit $F_{Y|X=x}^*(y) = P(Y \leq y|X = x)$, puis si elle existe, $f_{Y|X=x}(y) = F_{Y|X=x}'(y)$. Alors la densité conditionnelle de Y sachant $X = x$ est égale à $f_{Y|X=x}$, et pour $\phi : \mathbb{X} \times \mathbb{R} \rightarrow \mathbb{R}$ mesurable, avec $E[|\phi(X, Y)|] < +\infty$,

$$E[\phi(X, Y)|X = x] = \int_{\mathbb{R}} \phi(x, y) f_{Y|X=x}(y) dy.$$

6.5 Conditionnement dans le cas d'un couple de loi absolument continue

Soit f la densité conjointe du couple (X, Y) , f_X la densité de la variable marginale X . La loi de probabilité conditionnelle de Y sachant $X = x$ est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} et sa densité est donnée par :

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}.$$

On convient qu'en tout point x tel que $f_X(x) = 0$, $f_{Y|X=x}(y) = g(y)$, où g est une densité quelconque sur \mathbb{R} . On a alors pour toute fonction mesurable $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $E[|\phi(X, Y)|] < +\infty$,

$$E[\phi(X, Y)|X = x] = \int_{\mathbb{R}} \phi(x, y) f_{Y|X=x}(y) dy.$$

6.6 L'espérance conditionnelle comme projecteur

Supposons que Y admette un moment d'ordre 2. On montre le résultat intéressant suivant :
Si Z_0 minimise $\int (Y - Z)^2 dP$ parmi les v.a. \mathcal{S} -mesurables Z , alors Z_0 est une version de $E[Y|\mathcal{S}]$.

On considère l'espace de Hilbert $L^2(\Omega, \mathcal{A}, P)$, muni du produit scalaire $\langle X, Y \rangle = E[XY]$, et $L_{\mathcal{S}}^2$ le sous-espace de $L^2(\Omega, \mathcal{A}, P)$ défini comme l'ensemble des classes de P -équivalence contenant un représentant \mathcal{S} mesurable. Alors $E[Y|\mathcal{S}]$ s'interprète comme la projection orthogonale de Y sur $L_{\mathcal{S}}^2$. On peut ainsi retrouver les propriétés de l'espérance conditionnelle de façon géométrique (en particulier, théorème des trois perpendiculaires).

\hookrightarrow Intérêt en statistique : $E[Y|X]$ vue comme la fonction de X $\phi(X)$ qui minimise $\int (Y - \phi(X))^2 dP$.

7 Fonctions génératrices, fonction caractéristique

Soit (Ω, \mathcal{A}, P) un espace de probabilité, et X une variable ou un vecteur aléatoire défini sur cet espace.

7.1 Variables discrètes à valeurs dans \mathbb{N} : fonction génératrice (des moments factoriels)

On suppose ici que X est une v.a.r. de loi discrète, définie sur (Ω, \mathcal{A}, P) , à valeurs dans \mathbb{N} , de loi P_X et de fonction de masse π_X .

Définition 42 On appelle fonction génératrice de la loi de X (ou par abus fonction génératrice de X) la fonction $G_X : s \mapsto E[s^X] = \sum_{n \in \mathbb{N}} s^n \pi_X(n)$.

Par une étude classique de séries, on obtient les propriétés de la fonction G_X dont on cite ici les plus importantes. En particulier, on va voir que la fonction génératrice caractérise la loi de X , et qu'elle constitue un outil pratique de calcul des moments (factoriels), d'étude de l'indépendance ou d'étude de convergence.

Proposition 45 (Premières propriétés)

1. La fonction G_X est définie en tout réel s tel que $E[s^X] < +\infty$.
2. La série $\sum s^n \pi_X(n)$ étant absolument convergente dans l'intervalle $[-1, 1]$, G_X est continue sur $[-1, 1]$.
3. G_X admet des dérivées de tous les ordres, que l'on obtient comme somme des séries dérivées, pour tout $s \in]-1, 1[$.

Théorème 15 La fonction génératrice d'une v.a.r. de loi discrète à valeurs dans \mathbb{N} détermine la loi de cette variable.

Idee de la preuve : On se sert du fait que l'on peut dériver G_X à tout ordre, et que pour tout $k \in \mathbb{N}$, $G_X^{(k)}(0) = k! \pi_X(k)$.

Proposition 46 La fonction G_X admet une dérivée à gauche $G_X^{(k)}(1)$ d'ordre k en 1 si et seulement si le moment factoriel d'ordre k de X défini par $E[X(X-1)\dots(X-k+1)]$ existe et est fini. On a alors

$$E[X(X-1)\dots(X-k+1)] = G_X^{(k)}(1).$$

En particulier, pour $k = 1$, $E[X] = G_X'(1)$, et pour $k = 2$, $E[X(X-1)] = G_X''(1)$, d'où $Var(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$.

La preuve utilise le lemme d'Abel pour les séries. Pour plus de détails, on renvoie à D. Foata, A. Fuchs *Calcul des probabilités*.

Proposition 47 Si X et Y sont des v.a.r. indépendantes, de lois discrètes à valeurs dans \mathbb{N} , alors pour tout s (pour lequel $G_{X+Y}(s)$, $G_X(s)$, et $G_Y(s)$ sont définis),

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

La preuve est évidente.

Théorème 16 Soit $(X_k)_{k \in \mathbb{N}}$ une suite de v.a.r. à valeurs dans \mathbb{N} de fonctions de masse respectives $(\pi_{X_k})_{k \in \mathbb{N}}$, et X une v.a.r. à valeurs dans \mathbb{N} de fonction de masse π_X . Alors les deux propriétés suivantes sont équivalentes :

- (i) Pour tout $n \in \mathbb{N}$, $\lim_{k \rightarrow +\infty} \pi_{X_k}(n) = \pi_X(n)$ (on dit que X_k tend en loi vers X).
- (ii) Pour tout $s \in]0, 1[$, on a $\lim_{k \rightarrow +\infty} G_{X_k}(s) = G_X(s)$.

7.2 Variables aléatoires réelles quelconques : fonction génératrice des moments

On suppose ici que X est une v.a.r. (quelconque), définie sur (Ω, \mathcal{A}, P) , de loi P_X .

Dans ce cadre plus général que le précédent, on modifie un peu la notion de fonction génératrice. La nouvelle fonction ainsi introduite est appelée fonction génératrice des moments. Elle possède, comme on va le voir, les mêmes propriétés que la fonction génératrice pour les variables de loi discrète à valeurs dans \mathbb{N} . Les preuves des résultats annoncés ici sont plus difficiles, c'est pourquoi on n'en parlera pas, mais on renvoie si nécessaire, à P. Billingsley, *Probability and Measure*.

On introduit la fonction $g_X : u \mapsto E[e^{uX}]$ de la variable réelle u , définie sur l'ensemble des valeurs u pour lesquelles $E[e^{uX}] < +\infty$.

Cette fonction est toujours définie en $u = 0$, et $g_X(0) = 1$.

Définition 43 Si la fonction $g_X : u \mapsto E[e^{uX}]$ est définie sur un voisinage ouvert de 0, alors elle est appelée fonction génératrice des moments de (la loi de) X .

On peut également introduire une autre fonction, qui aura les mêmes applications, et qui par exemple est continue et bornée sur $[0, +\infty[$ lorsque X est à valeurs positives (et qu'alors g_X est continue et bornée sur $] -\infty, 0]$) : la *transformée de Laplace*, définie par

$$L_X(u) = E[e^{-uX}].$$

Proposition 48 (Premières propriétés)

1. Pour $a, b \in \mathbb{R}$, $g_{aX+b}(u) = e^{bu}g_X(au)$.
2. Si la loi de X est symétrique (c'est-à-dire si X a la même loi que $-X$), la fonction g_X est paire.
3. La fonction g_X est convexe.

Théorème 17 La fonction génératrice des moments d'une v.a.r. détermine la loi de cette variable.

Proposition 49 Soit X une v.a.r. admettant une fonction génératrice des moments g_X , définie pour tout u dans un intervalle ouvert $]u_1, u_2[$ ($u_1 < 0 < u_2$). Alors :

- pour tout $k \in \mathbb{N}$, $E[|X|^k] < +\infty$,
- pour tout $u \in]-u_0, u_0[$, avec $0 < u_0 < \min(-u_1, u_2)$,

$$g_X(u) = 1 + E[X]u + E[X^2]\frac{u^2}{2!} + \dots + E[X^n]\frac{u^n}{n!} + \dots$$

- pour tout $k \in \mathbb{N}$, $E[X^k] = g_X^{(k)}(0)$.

Remarque : une v.a.r. qui admet une fonction génératrice des moments admet donc des moments de tous les ordres. Mais en revanche, une v.a.r. peut admettre des moments de tous les ordres sans admettre de fonction génératrice (exemple de la loi log-normale). Ainsi, la suite des moments d'une v.a.r. ne détermine pas sa loi! (\neq cas d'une v.a.r. de loi discrète à valeurs dans \mathbb{N}).

Proposition 50 Si X et Y sont des v.a.r. indépendantes admettant chacune une fonction génératrice des moments, alors pour tout u (pour lequel $g_{X+Y}(u)$, $g_X(u)$, et $g_Y(u)$ sont définis),

$$g_{X+Y}(u) = g_X(u)g_Y(u).$$

Attention : La réciproque est fautive. Contre-exemple : un couple aléatoire (X, Y) de densité conjointe définie par $f(x, y) = 2\mathbb{I}_{\{(x,y) \in [0,1]^2, x \leq y \leq 1/2+x \text{ ou } y \leq -1/2+x\}}$.

Remarque : Lorsque la v.a.r. X est à valeurs dans \mathbb{N} , on préfère utiliser généralement la fonction génératrice (des moments factoriels) G_X , qui, si elle est définie dans un voisinage ouvert de 1, est liée à la fonction génératrice g_X par la relation : $g_X(u) = G_X(e^u)$.

7.3 Fonction caractéristique d'une v.a.r.

Les deux inconvénients majeurs de l'utilisation de la fonction génératrice dans l'étude de la loi de la v.a.r. X sont les suivants. D'une part, la fonction génératrice g_X n'est pas définie pour tout réel u . D'autre part, on ne sait pas "inverser" cette fonction, en d'autres termes, on ne sait pas retrouver explicitement la loi de X à partir de la fonction g_X seule.

Pour remédier à cela, on utilise une autre fonction, la fonction caractéristique, qui elle sera définie pour tout réel t , et qui sera "inversible". Cependant, l'utilisation de cette nouvelle fonction peut sembler plus difficile, puisqu'elle fait appel à la théorie des fonctions d'une variable complexe.

On considère une v.a.r. X définie sur (Ω, \mathcal{A}, P) .

Définition 44 Soit X une v.a.r. On appelle fonction caractéristique de X la fonction φ_X définie comme la transformée de Fourier de sa loi de probabilité, c'est-à-dire la fonction φ_X définie pour tout réel t par

$$\varphi_X(t) = E[e^{itX}].$$

Si X est une v.a.r. de loi discrète à valeurs dans \mathbb{X} , de fonction de masse π_X , alors

$$\varphi_X(t) = \sum_{x \in \mathbb{X}} \pi_X(x) e^{itx}.$$

Si X est une v.a.r. de loi absolument continue de densité f_X , alors

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx.$$

Proposition 51 (Premières propriétés)

1. φ_X est définie et continue pour tout nombre réel t .
2. φ_X est bornée, et pour tout réel t , $|\varphi_X(t)| \leq 1$.
3. Pour tous réels a et b , $\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$.
4. Si la loi de X est symétrique, φ_X est une fonction réelle paire.
5. Toute combinaison convexe de fonctions caractéristiques est une fonction caractéristique.

Théorème 18 La fonction caractéristique φ_X détermine la loi de X (d'où le terme de fonction caractéristique !). En particulier, si $\int_{\mathbb{R}} |\varphi_X(t)| dt < +\infty$, alors X admet une densité f_X continue et

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) e^{-itx} dt.$$

Sinon, si F_X désigne la fonction de répartition de X , on a toujours :

$$F_X(a) - F_X(b) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{+T} \varphi_X(t) \frac{e^{-ita} - e^{-itb}}{it} dt.$$

Comme la fonction génératrice des moments, la fonction caractéristique permet un calcul aisé des moments de tous ordres d'une v.a.r.

Théorème 19 Soit X une v.a.r de fonction caractéristique φ_X . Supposons qu'il existe $n \in \mathbb{N}^*$ tel que $E[|X|^n] < +\infty$. Alors,

1. φ_X est continûment dérivable jusqu'à l'ordre n inclus.
2. Pour $k = 0, 1, \dots, n$, $\varphi_X^{(k)}(0) = i^k E[X^k]$ (en particulier $\varphi_X'(0) = iE[X]$).
3. Lorsque $t \rightarrow 0$, $\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!} E[X^k] + o(|t|^n)$.

Enfin, on a, pour un couple de v.a.r. indépendantes, le résultat suivant.

Proposition 52 Si X et Y sont des v.a.r. indépendantes, alors pour tout réel t ,

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t).$$

Attention : là encore, la réciproque est fautive. Un contre-exemple est donné par X qui suit une loi de Cauchy, et $Y = X$.

Fonctions caractéristiques de quelques lois usuelles

Loi	Fonction caractéristique $\varphi_X(t)$
Bernoulli $\mathcal{B}(p)$	$pe^{it} + (1 - p)$
Binomiale $\mathcal{B}(n, p)$	$(pe^{it} + (1 - p))^n$
Poisson $\mathcal{P}(\lambda)$	$e^{\lambda(e^{it} - 1)}$
Géométrique $\mathcal{G}(p)$	$pe^{it} / (1 - (1 - p)e^{it})$
Uniforme sur $[-a, a]$	$\sin(at) / (at)$
Exponentielle $\mathcal{E}(\lambda)$	$\lambda / (\lambda - it)$
Gaussienne $\mathcal{N}(m, \sigma^2)$	$e^{imt} e^{-\sigma^2 t^2 / 2}$

7.4 Fonction caractéristique d'un vecteur aléatoire

On considère maintenant le cas où X est un vecteur aléatoire de dimension n : $X = (X_1, \dots, X_n)$. On donne ici seulement la définition, et quelques propriétés utiles. Pour en savoir plus, on renvoie par exemple à D. Foata, A. Fuchs *Calcul des probabilités*.

Définition 45 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire défini sur (Ω, \mathcal{A}, P) . On appelle fonction caractéristique de X la fonction φ_X définie pour tout vecteur $t = (t_1, \dots, t_n) \in \mathbb{R}^n$, par

$$\varphi_X(t) = E[e^{i\langle t, X \rangle}] = E[e^{i(t_1 X_1 + \dots + t_n X_n)}].$$

Théorème 20 La fonction caractéristique φ_X détermine la loi de X .

Conséquence : Théorème de Cramer-Wold. La loi de X est entièrement déterminée par celles de toutes les combinaisons linéaires de ses variables marginales.

Théorème 21 Les v.a.r. X_1, \dots, X_n sont indépendantes si et seulement si la fonction caractéristique de $X = (X_1, \dots, X_n)$ est égale au produit des fonctions caractéristiques de ses variables marginales X_1, \dots, X_n .

Fonction caractéristique de la loi d'un vecteur gaussien $\mathcal{N}_n(m_X, \Sigma_X)$

$$\varphi_X(t) = e^{i\langle m_X, t \rangle} e^{-\frac{1}{2} t' \Sigma_X t}.$$

8 Convergences des suites de variables aléatoires

8.1 Les différents types de convergence

Une suite $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires réelles définies sur (Ω, \mathcal{A}, P) étant une suite de fonctions de Ω dans \mathbb{R} , il existe différentes façons de définir la convergence de $(X_n)_{n \in \mathbb{N}}$ vers une constante $a \in \mathbb{R}$ ou vers une autre variable aléatoire X définie sur (Ω, \mathcal{A}, P) .

8.1.1 Convergence en moyenne d'ordre $r > 0$

Définition 46 Si pour tout $n \in \mathbb{N}$, $E[|X_n|^r] < +\infty$ avec $r > 0$, on dit que :

(i) la suite $(X_n)_{n \in \mathbb{N}}$ converge en moyenne d'ordre r (ou dans $L^r(\Omega, \mathcal{A}, P)$) vers a si $E[|X_n - a|^r] \rightarrow 0$ lorsque n tend vers $+\infty$.

(ii) la suite $(X_n)_{n \in \mathbb{N}}$ converge en moyenne d'ordre r vers X si $E[|X_n - X|^r] \rightarrow 0$ lorsque n tend vers $+\infty$ ou si la suite $(X_n - X)_{n \in \mathbb{N}}$ converge vers 0 en moyenne d'ordre r .

Lorsque $r = 2$, on parle de convergence en moyenne quadratique.

Remarque : Si $(X_n)_{n \in \mathbb{N}}$ converge vers X en moyenne d'ordre r , cela n'implique pas que le moment $E[|X|^r]$ soit fini, mais s'il l'est, alors $E[|X_n|^r] \rightarrow E[|X|^r]$.

8.1.2 Convergence presque sûre ou convergence forte

Définition 47 On dit que :

(i) la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers a s'il existe un ensemble $S \subset \Omega$ de probabilité 1 ($P(S) = 1$) tel que pour tout $\omega \in S$, $X_n(\omega) \rightarrow a$ lorsque n tend vers $+\infty$. On note alors $X_n \xrightarrow{p.s.} a$.

(ii) la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X s'il existe un ensemble $S \subset \Omega$ de probabilité 1 tel que pour tout $\omega \in S$, $X_n(\omega) \rightarrow X(\omega)$ lorsque n tend vers $+\infty$, ou si $X_n - X \xrightarrow{p.s.} 0$. On note alors $X_n \xrightarrow{p.s.} X$.

On donne ci-dessous des critères de convergence presque sûre. Ces critères ne sont donnés que pour la convergence presque sûre vers 0 d'une suite de v.a.r., mais on en déduira directement les résultats correspondants pour la convergence presque sûre vers X .

Théorème 22 La suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers 0 si et seulement si pour tout $\varepsilon > 0$,

$$P(\sup_{m \geq n} |X_m| > \varepsilon) \rightarrow_{n \rightarrow +\infty} 0.$$

Proposition 53 Si pour tout $\varepsilon > 0$, la série de terme général $P(|X_n| > \varepsilon)$ est convergente, la suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers 0.

Proposition 54 S'il existe $r > 0$ tel que la série de terme général $E[|X_n|^r]$ soit convergente, alors la suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers 0.

Preuve. Inégalité de Bienaymé-Chebychev.

8.1.3 Convergence en probabilité

Définition 48 On dit que :

(i) la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers a si pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow +\infty} P(|X_n - a| > \varepsilon) = 0.$$

On note $X_n \xrightarrow{P} a$.

(ii) la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow +\infty} P(|X_n - X| > \varepsilon) = 0,$$

ou si $X_n - X \xrightarrow{P} 0$. On note $X_n \xrightarrow{P} X$.

Théorème 23 Si $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ sont des suites de v.a.r. qui convergent en probabilité respectivement vers X et Y , et si $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est une fonction continue en tout point (x, y) de \mathbb{R}^2 , alors la suite des v.a.r. $(h(X_n, Y_n))_{n \in \mathbb{N}}$ converge en probabilité vers $h(X, Y)$.

En particulier, si la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , et si g est une fonction réelle continue en tout point de \mathbb{R} , alors $g(X_n) \xrightarrow{P} g(X)$.

Théorème 24 Si la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X et si $P(X = 0) = 0$ alors $1/X_n \xrightarrow{P} 1/X$.

On déduit de ces deux théorèmes les propriétés opératoires de la convergence en probabilité.

Les différentes notions de convergence vues ci-dessus peuvent s'étendre aisément au cas vectoriel en considérant une norme $\|\cdot\|$ sur \mathbb{R}^n au lieu de la valeur absolue $|\cdot|$.

8.1.4 Convergence en loi ou convergence étroite

On considère ici toujours la suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$, et on note pour tout $n \in \mathbb{N}$, F_{X_n} la fonction de répartition de X_n , φ_{X_n} la fonction caractéristique de X_n , P_{X_n} la loi de X_n , F_X la fonction de répartition de X , φ_X la fonction caractéristique de X , P_X la loi de X .

Définition 49 On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi (ou étroitement) vers X si en tout point x de continuité de F_X , on a $F_{X_n}(x) \rightarrow F_X(x)$ lorsque n tend vers $+\infty$. On note $X_n \xrightarrow{\mathcal{L}} X$.

Remarque : En toute rigueur, on devrait dire que la suite des lois des X_n converge en loi ou étroitement vers la loi de X .

Exemples.

Cette définition de la convergence en loi ne se généralisera pas aisément au cas vectoriel. On donne ici une autre définition de la convergence en loi qui, elle se généralisera de façon directe.

Définition 50 (Définition équivalente) On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi (ou étroitement) vers X si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée, alors $\int f dP_{X_n} \rightarrow \int f dP_X$.

Théorème 25 Si pour tout $n \in \mathbb{N}$, la loi P_{X_n} de la v.a.r. X_n admet une densité f_{X_n} par rapport à la mesure de Lebesgue, si X admet une densité f_X , et si pour presque tout x de \mathbb{R} , $f_{X_n}(x) \rightarrow f_X(x)$, alors $X_n \rightarrow^{\mathcal{L}} X$.

Théorème 26 (Paul Lévy)

(i) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. qui converge en loi vers une v.a.r. X . Alors la suite des fonctions caractéristiques $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge vers φ_X uniformément dans tout intervalle fini $[-u, u]$.

(ii) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. telle que la suite des fonctions caractéristiques $(\varphi_{X_n})_{n \in \mathbb{N}}$ converge (simplement) vers une fonction φ continue en 0. Alors φ est la fonction caractéristique d'une v.a.r. X et $X_n \rightarrow^{\mathcal{L}} X$.

Les fonctions caractéristiques fournissent donc un outil très pratique d'étude de la convergence en loi d'une suite de v.a.r.

Exemple. Convergence de la loi binomiale $\mathcal{B}(n, p_n)$ vers la loi de Poisson $\mathcal{P}(\lambda)$ lorsque $np_n \rightarrow \lambda$.

Théorème 27 Si la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X , et si g est une fonction réelle continue en tout point de \mathbb{R} , alors $g(X_n) \rightarrow^{\mathcal{L}} g(X)$.

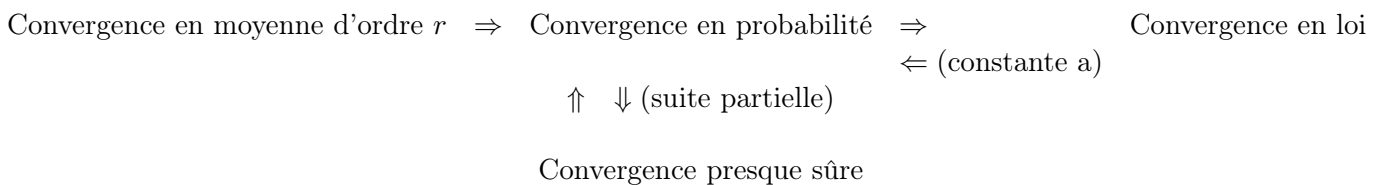
Ces quatre dernières propositions peuvent se généraliser directement au cas vectoriel.

Théorème 28 Si $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ sont des suites de v.a.r. qui convergent en loi respectivement vers X et a , alors la suite des vecteurs aléatoires $(X_n, Y_n)_{n \in \mathbb{N}}$ converge en loi vers (X, a) .

Conséquences : opérations algébriques si $Y_n \rightarrow^{\mathcal{L}} a$ (constante).

8.2 Liens entre les différentes convergences

On peut représenter les liens entre les différentes notions de convergence par le diagramme suivant.



L'étude de sommes de variables aléatoires indépendantes et de même loi joue un rôle capital en statistique. Les deux paragraphes suivants donnent des résultats de convergence pour de telles sommes.

8.3 Lois des grands nombres

Considérons le cas simple où l'on a des v.a.r. X_1, \dots, X_n de loi de Bernoulli de même paramètre p , indépendantes. Alors $\sum_{i=1}^n X_i \sim \mathcal{B}(n, p)$, et l'inégalité de Chebychev nous donne : pour tout $\varepsilon > 0$,

$$P \left(\left| \frac{\sum_{i=1}^n X_i}{n} - p \right| > \varepsilon \right) \leq \frac{p(1-p)}{n\varepsilon^2} \rightarrow_{n \rightarrow +\infty} 0.$$

On en déduit par définition que $\frac{\sum_{i=1}^n X_i}{n} \rightarrow^P p$.

En fait, ce résultat se généralise à d'autres types de lois, et à des v.a.r. indépendantes deux à deux et non corrélées deux à deux, ou à la convergence presque sûre.

8.3.1 Lois faibles des grands nombres

Théorème 29 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. de même loi telles que

- $E[|X_1|] < +\infty$ et les X_n sont indépendantes deux à deux,

OU

- $E[X_1^2] < +\infty$ et les X_n sont non corrélées deux à deux.

Alors $\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{P} E[X_1]$

Remarque : En fait, la convergence a lieu en moyenne d'ordre 1 dans le premier cas, en moyenne d'ordre 2 dans le deuxième cas.

8.3.2 Loi forte des grands nombres (Kolmogorov)

Théorème 30 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes, de même loi, telle que $E[|X_1|] < +\infty$. Alors

$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow{p.s.} E[X_1]$

8.4 Théorèmes centraux limites

On a vu que la somme de n v.a.r. indépendantes de même loi de Bernoulli de paramètre p_n tel que $np_n \rightarrow \lambda$ converge en loi vers la loi de Poisson de paramètre λ .

Dans le même ordre d'idée, on peut montrer que la somme de n v.a.r. indépendantes de même loi de Bernoulli de paramètre $p \in]0, 1[$, après centrage et réduction, converge en loi vers une loi normale centrée réduite. C'est le théorème de De Moivre - Laplace. En fait, ce résultat se généralise à d'autres lois.

Théorème 31 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes, de même loi, telles que $E[X_1^2] < +\infty$.

Alors

$$\sqrt{n} \frac{\frac{\sum_{i=1}^n X_i}{n} - E[X_1]}{\sqrt{\text{Var}(X_1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Preuve. Utilisation d'un développement limité de la fonction caractéristique.

Remarque : on peut alléger un peu les hypothèses \leftrightarrow théorème de Lindeberg.

Ce théorème a son pendant vectoriel.

Théorème 32 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de vecteurs aléatoires du second ordre indépendants, de même loi, d'espérance m , de matrice de variances-covariances Σ .

Alors

$$\sqrt{n} \left(\frac{\sum_{i=1}^n X_i}{n} - m \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma).$$

Ces deux théorèmes disent toute l'importance du rôle de la loi normale en statistique.