

Chapitre 2

Vecteurs aléatoires

2.1 Paire de variables aléatoires

Nous étendons les définitions du module précédent concernant une variable aléatoire au cas d'une paire de variables aléatoires.

2.1.1 Fonction de répartition jointe et marginale

Une paire de v.a. est caractérisée par sa *fonction de répartition jointe* $F_{XY}(x, y)$ qui est la probabilité que X prenne une valeur inférieure ou égale au réel x et que Y prenne une valeur inférieure ou égale au réel y :

$$F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x, Y \leq y) \quad (2.1)$$

Chaque v.a. X et Y est individuellement caractérisée par sa *fonction de répartition marginale* $F_X(x)$ et $F_Y(y)$:

$$F_X(x) = P(X \leq x) \quad (2.2)$$

$$F_Y(y) = P(Y \leq y) \quad (2.3)$$

$$(2.4)$$

Elles possèdent les propriétés suivantes :

P1. $0 \leq F_{XY}(x, y) \leq 1$.

P2. $\lim_{x \rightarrow -\infty} F_{XY}(x, y) = \lim_{y \rightarrow -\infty} F_{XY}(x, y) = 0$ et $\lim_{x \rightarrow \infty, y \rightarrow \infty} F_{XY}(x, y) = 1$.

P3. Les fonctions de répartition marginales de X et Y peuvent être obtenues à partir de la fonction de répartition jointe en constatant que

$$\lim_{y \rightarrow \infty} F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq \infty\}) = P(X \leq x) = F_X(x)$$

$$\lim_{x \rightarrow \infty} F_{XY}(x, y) = P(\{X \leq \infty\} \cap \{Y \leq y\}) = P(Y \leq y) = F_Y(y)$$

2.1.2 Densité de probabilité jointe et marginale

La fonction de *densité de probabilité jointe* est définie par

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}. \quad (2.5)$$

Elle possède les propriétés suivantes :

P4. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1.$

P5. $P(X \leq a, Y \leq b) = F_{XY}(a, b) = \int_{-\infty}^a \int_{-\infty}^b f_{XY}(x, y) dy dx.$

P6. $P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{XY}(a_2, b_2) - F_{XY}(a_1, b_2) - F_{XY}(a_2, b_1) + F_{XY}(a_1, b_1) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{XY}(x, y) dx dy.$ Cette propriété entraîne que $P(x < X \leq x + \Delta x, y < Y \leq y + \Delta y) = f_{XY}(x, y) \Delta x \Delta y$ lorsque $\Delta x \rightarrow 0$ et $\Delta y \rightarrow 0$.

Les fonctions de *densité de probabilité marginale* sont définies par

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (2.6)$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} \quad (2.7)$$

Elle peuvent être obtenues à partir de la densité jointe de la manière suivante :

P7. A partir de (2.6) et de la propriété P3, on a

$$\begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} = \frac{dF_{XY}(x, +\infty)}{dx} = \frac{d}{dx} \left[\int_{-\infty}^x \int_{-\infty}^{\infty} f_{XY}(\xi, \eta) d\xi d\eta \right] \\ &= \frac{d}{dx} \int_{-\infty}^x \left[\int_{-\infty}^{\infty} f_{XY}(\xi, \eta) d\eta \right] d\xi = \int_{-\infty}^{\infty} f_{XY}(x, \eta) d\eta \end{aligned}$$

et de manière similaire, on a

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \int_{-\infty}^{\infty} f_{XY}(\xi, y) d\xi.$$

Dans le cas discret, l'utilisation des impulsions de Dirac $\delta(\cdot)$ permet d'écrire (2.5) sous la forme

$$f_{XY}(x, y) = \sum_{i,j} p_{ij} \delta(x - x_i) \delta(y - y_j)$$

où $p_{ij} = P(X = x_i, Y = y_j)$, et (2.6) et (2.7) sous la forme

$$\begin{aligned} f_X(x) &= \sum_{i,j} p_{ij} \delta(x - x_i) \\ f_Y(y) &= \sum_{i,j} p_{ij} \delta(y - y_j). \end{aligned}$$

2.1.3 Variables aléatoires indépendantes

Deux v.a. X et Y sont indépendantes si

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

d'où

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

Dans le cas discret, l'indépendance de X et Y entraîne que $P(X = a, Y = b) = P(X = a)P(Y = b)$.

Notons que si X et Y sont deux v.a. indépendantes, alors $Z = g(X)$ et $W = h(Y)$ sont également deux v.a. indépendantes.

2.1.4 Fonctions de deux variables aléatoires

Soit $g_1(\cdot)$ et $g_2(\cdot)$ deux fonctions continues. De manière similaire au cas d'une variable, on peut calculer la densité de probabilité jointe des variables

$$\begin{aligned} Z &= g_1(X, Y) \\ W &= g_2(X, Y) \end{aligned}$$

en fonction de la densité jointe des v.a. X et Y .

Soit (z, w) une paire de valeurs particulières que peuvent prendre la paire de v.a. (Z, W) , et soient $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$ les m racines respectives du système d'équations

$$\begin{aligned} z &= g_1(x, y) \\ w &= g_2(x, y). \end{aligned}$$

Le jacobien de la transformation $(g_1(\cdot), g_2(\cdot))$ est le déterminant

$$J(x, y) = \det \begin{bmatrix} \frac{\partial g_1(x, y)}{\partial x} & \frac{\partial g_1(x, y)}{\partial y} \\ \frac{\partial g_2(x, y)}{\partial x} & \frac{\partial g_2(x, y)}{\partial y} \end{bmatrix}.$$

Alors on peut montrer que

$$f_{ZW}(z, w) = \sum_i \frac{f_{XY}(x_i, y_i)}{|J(x_i, y_i)|}. \quad (2.8)$$

Un cas particulier très important est celui d'une transformation linéaire

$$\begin{aligned} z &= g_1(x, y) = a_{11}x + a_{12}y + b_1 \\ w &= g_2(x, y) = a_{21}x + a_{22}y + b_2 \end{aligned}$$

qui s'écrit sous forme vectorielle

$$\mathbf{z} = A\mathbf{x} + \mathbf{b}.$$

La solution de cette équation existe si $\det(A) \neq 0$, elle est unique et est donnée par

$$\mathbf{x} = A^{-1}(\mathbf{z} - \mathbf{b}).$$

Son jacobien est le déterminant de la matrice A pour tous les points (x, y) :

$$J(x, y) = \det(A) = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Par conséquent, si $\det(A) \neq 0$, la densité jointe de $\mathbf{Z} = (Z, W)$ peut s'exprimer en fonction de celle de $\mathbf{X} = (X, Y)$ par

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{f_{\mathbf{X}}(A^{-1}(\mathbf{z} - \mathbf{b}))}{|\det(A)|}. \quad (2.9)$$

i.e.,

$$f_{ZW}(z, w) = \frac{f_{XY}\left(\frac{a_{22}(z-b_1)-a_{12}(w-b_2)}{a_{11}a_{22}-a_{12}a_{21}}, \frac{a_{11}(w-b_2)-a_{12}(z-b_1)}{a_{11}a_{22}-a_{12}a_{21}}\right)}{|a_{11}a_{22} - a_{12}a_{21}|}.$$

Prenons l'exemple de la somme de deux v.a. X et Y indépendantes :

$$Z = X + Y.$$

Pour utiliser (2.9), il faut deux équations. On peut ajouter par exemple l'équation

$$W = Y$$

qui ne nous intéresse pas mais permet d'obtenir un système de deux équations à deux inconnues (et de plein rang)

$$\begin{bmatrix} Z \\ W \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Pour ce système linéaire particulier, (2.9) devient

$$f_{ZW}(z, w) = f_{XY}(z - w, w)$$

ou encore, puisque X et Y sont indépendantes,

$$f_{ZW}(z, w) = f_X(z - w)f_Y(w).$$

La fonction souhaitée est la densité marginale de Z et non la densité jointe de Z et W , ce qu'on calcule grâce à la propriété P7 :

$$f_Z(z) = \int_{-\infty}^{\infty} f_{ZW}(z, w)dw = \int_{-\infty}^{\infty} f_X(z - w)f_Y(w)dw$$

où on reconnaît la formule du produit de convolution, si bien qu'on a montré que la densité de probabilité de la somme de deux v.a. indépendantes est le produit de convolution des densités de probabilité de chaque variable :

$$f_Z = f_X \star f_Y \quad (2.10)$$

2.1.5 Espérance d'une fonction de deux variables aléatoires

L'espérance de la fonction $g(X, Y) = Z$, où X et Y sont deux v.a. de densité de probabilité jointe $f_{XY}(x, y)$, est définie par

$$E[Z] = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy. \quad (2.11)$$

Dans le cas discret, cette grandeur devient, avec $p_{ij} = P(X = x_i, Y = y_j)$,

$$E[Z] = E[g(X, Y)] = \sum_{i,j} g(x_i, y_j) p_{ij}.$$

Comme premier exemple, reprenons celui de la somme de deux v.a. $Z = g(X, Y) = X + Y$, mais cette fois sans que X et Y soient nécessairement indépendantes. On calcule

$$\begin{aligned} E[Z] = E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{XY}(x, y) dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{XY}(x, y) dx \right) dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E[X] + E[Y]. \end{aligned}$$

Cette relation, combinée avec le fait que pour toute constante α , $E[\alpha X] = \alpha E[X]$, montre que l'opérateur d'espérance est linéaire.

Dans un second exemple, considérons le produit de v.a. X et Y , qui cette fois sont indépendantes : $Z = g(X, Y) = XY$. La définition (2.11) devient alors successivement

$$\begin{aligned} E[Z] = E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E[X] E[Y]. \end{aligned}$$

Les *moments joints* de X et Y sont les espérances de $g(X, Y) = X^n Y^m$, i.e.

$$E[X^n Y^m] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^n y^m f_{XY}(x, y) dx dy.$$

Une quantité très importante est la fonction de *covariance* définie par

$$COV[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] \quad (2.12)$$

et qui peut après quelques manipulations encore être mise sous la forme

$$COV[X, Y] = E[XY] - \mu_X \mu_Y = E[XY] - E[X]E[Y]. \quad (2.13)$$

Dans le cas où $COV[X, Y] = 0$, les deux v.a. X et Y sont dites *décorrélées* et dans celui où $E[XY] = 0$ elles sont dites *orthogonales*.

Remarquons que $COV[X, X] = E[(X - \mu_X)^2] = \sigma_X^2$.

On peut normaliser la covariance (cfr propriété P8 ci-dessous) et définir ainsi le *coefficient de corrélation* par

$$\rho = \frac{COV[X, Y]}{\sigma_X \sigma_Y} \quad (2.14)$$

La fonction de covariance possède deux propriétés importantes :

P8. La fonction de covariance est bornée : $|COV[X, Y]| \leq \sigma_X \sigma_Y$, comme on le montrera à l'exercice 7, et par conséquent la valeur absolue du coefficient de corrélation est inférieure à l'unité : $|\rho| \leq 1$.

P9. Si X et Y sont deux v.a. indépendantes, on a vu plus haut que $E[XY] = E[X]E[Y]$ et dès lors $COV[X, Y] = 0$. De même, dans ce cas, $\rho = 0$, et donc des v.a. indépendantes sont décorréliées. Par contre, deux v.a. décorréliées ne sont nécessairement indépendantes, comme on le verra à l'exercice 8.

2.1.6 Transformées

On peut également étendre la définition de fonction caractéristique d'une v.a. du module précédent à celle de *fonction caractéristique jointe* de deux v.a. X et Y par

$$\Phi_{XY}(\omega_1, \omega_2) = E[e^{j(\omega_1 X + \omega_2 Y)}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\omega_1 x + \omega_2 y)} f_{XY}(x, y) dx dy. \quad (2.15)$$

Cette fonction est le complexe conjugué de la transformée de Fourier bidimensionnelle de la fonction $f_{XY}(x, y)$. La transformée inverse permet d'obtenir la densité de probabilité à partir de la fonction caractéristique par le fait que

$$f_{XY}(x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j(\omega_1 x + \omega_2 y)} \Phi_{XY}(\omega_1, \omega_2) d\omega_1 d\omega_2. \quad (2.16)$$

Les fonctions caractéristiques marginales de X et Y sont directement obtenues à partir de leur fonction caractéristique jointe par

$$\begin{aligned} \Phi_X(\omega) &= \Phi_{XY}(\omega, 0) \\ \Phi_Y(\omega) &= \Phi_{XY}(0, \omega). \end{aligned}$$

Les moments joints peuvent être calculés par

$$E[X^m Y^n] = \frac{1}{j^{m+n}} \left. \frac{\partial^{m+n} \Phi_{XY}(\omega_1, \omega_2)}{\partial \omega_1^m \partial \omega_2^n} \right|_{(\omega_1, \omega_2) = (0, 0)}$$

Si X et Y sont indépendantes, on a vu à la section 1.3, que $Z = g(X)$ et $W = h(Y)$ sont également deux v.a. indépendantes. En particulier, $Z = g(X) = \exp(j\omega_1 X)$ et $W = h(Y) = \exp(j\omega_2 Y)$ sont deux v.a. indépendantes, ce qui entraîne que $E[ZW] = E[Z]E[W]$, c'est-à-dire que

$$E[e^{j\omega_1 X + j\omega_2 Y}] = E[e^{j\omega_1 X}]E[e^{j\omega_2 Y}]$$

ce qui montre que la fonction caractéristique jointe de deux v.a. indépendantes est le produit des deux fonctions caractéristiques marginales :

$$\Phi_{XY}(\omega_1, \omega_2) = \Phi_X(\omega_1)\Phi_Y(\omega_2).$$

Enfin, comme la densité de probabilité de la v.a. $Z = X + Y$ est le produit de convolution (2.10) des densités de probabilité des deux v.a. X et Y si celles-ci sont indépendantes, de la même manière que la transformée de Fourier d'un produit de convolution est le produit des transformées de Fourier, on a que

$$\Phi_Z(\omega) = \Phi_X(\omega)\Phi_Y(\omega).$$

Finalement, on étend la définition de fonction caractéristique jointe à celle de *fonction génératrice de moment*

$$\hat{\Phi}_{XY}(s_1, s_2) = E[e^{s_1 X + s_2 Y}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{s_1 x + s_2 y} f_{XY}(x, y) dx dy. \quad (2.17)$$

2.1.7 Densité de probabilité conditionnelle

Considérons d'abord deux v.a. discrètes X et Y . En appliquant les définitions de probabilité conditionnelle $P(A|B)$ et $P(B|A)$ aux événements $A = \{\zeta | X(\zeta) = x_i\}$ et $B = \{\zeta | Y(\zeta) = y_j\}$, on trouve

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \quad \text{si } P(Y = y_j) \neq 0 \quad (2.18)$$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} \quad \text{si } P(X = x_i) \neq 0 \quad (2.19)$$

$$P(X = x_i | Y = y_j) = 0 \quad \text{si } P(Y = y_j) = 0 \quad (2.20)$$

$$P(Y = y_j | X = x_i) = 0 \quad \text{si } P(X = x_i) = 0. \quad (2.21)$$

Si X est une v.a. continue tandis que Y est une v.a. discrète, on peut définir la *fonction de répartition conditionnelle* par

$$P(X \leq x | Y = y_j) = \frac{P(X \leq x, Y = y_j)}{P(Y = y_j)} = F_{X|Y}(x|y_j) \quad \text{si } P(Y = y_j) \neq 0 \quad (2.22)$$

$$P(X \leq x | Y = y_j) = 0 = F_{X|Y}(x|y_j) \quad \text{si } P(Y = y_j) = 0 \quad (2.23)$$

et la *densité de probabilité conditionnelle* de X étant donné que $Y = y$ est

$$f_{X|Y}(x|y_j) = \frac{dF_{X|Y}(x|y_j)}{dx}.$$

Dans le cas continu, comme $P(Y = y) = 0$, on ne peut pas utiliser (2.22). On définit la *fonction de répartition conditionnelle* par la limite

$$F_{X|Y}(x|y) = \lim_{\Delta y \rightarrow 0} P(X \leq x | y \leq Y < y + \Delta y)$$

qui vaut

$$\begin{aligned} F_{X|Y}(x|y) &= \lim_{\Delta y \rightarrow 0} P(X \leq x | y \leq Y < y + \Delta y) = \lim_{\Delta y \rightarrow 0} \frac{P(X \leq x, y \leq Y < y + \Delta y)}{P(y \leq Y < y + \Delta y)} \\ &= \lim_{\Delta y \rightarrow 0} \frac{\int_{-\infty}^x \int_y^{y+\Delta y} f_{XY}(\xi, \eta) d\xi d\eta}{\int_y^{y+\Delta y} f_Y(\eta) d\eta} \\ &= \lim_{\Delta y \rightarrow 0} \frac{\int_{-\infty}^x f_{XY}(\xi, y) \Delta y d\xi}{f_Y(y) \Delta y} = \frac{\int_{-\infty}^x f_{XY}(\xi, y) d\xi}{f_Y(y)}. \end{aligned}$$

En dérivant cette dernière expression par rapport à x , on obtient la *densité de probabilité conditionnelle* de X si $Y = y$

$$f_{X|Y}(x|y) = \frac{dF_{X|Y}(x|y)}{dx} = \frac{f_{XY}(x, y)}{f_Y(y)}. \quad (2.24)$$

De manière similaire, la *densité de probabilité conditionnelle* de Y si $X = x$ est

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}. \quad (2.25)$$

2.1.8 Espérance conditionnelle

L'*espérance conditionnelle* de X si $Y = y$ est définie par

$$E[X|y_j] = \sum_i x_i P(X = x_i | Y = y_j) \quad (2.26)$$

dans le cas discret et par

$$E[X|y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \quad (2.27)$$

dans le cas continu.

Il faut souligner que l'espérance conditionnelle de X étant donnée une valeur particulière y de Y est une fonction déterministe de y : $E[X|y] = g(y)$. Comme cette valeur y est le résultat d'une expérience aléatoire, on peut s'intéresser à la v.a. $g(Y)$, c'est-à-dire $E[X|Y]$. En particulier, on peut calculer la moyenne $E[E[X|Y]]$ de la variable aléatoire $E[X|Y]$:

$$\begin{aligned} E[E[X|Y]] &= \int_{-\infty}^{\infty} E[X|y] f_Y(y) dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{XY}(x, y) dy \right) dx = \int_{-\infty}^{\infty} x f_X(x) dx = E[X]. \end{aligned}$$

2.1.9 Variables aléatoires complexes

Une variable aléatoire complexe Z est une v.a. de la forme

$$Z = X + jY$$

où X et Y sont une paire de v.a. réelles. On peut étendre les définitions précédentes au cas complexe. Par exemple, la moyenne est

$$\mu_Z = E[Z] = E[X] + jE[Y] = \mu_X + j\mu_Y,$$

la variance est

$$\sigma_Z^2 = E[|Z - \mu_Z|^2] = E[(Z^* - \mu_Z^*)(Z - \mu_Z)] = \dots = \sigma_X^2 + \sigma_Y^2$$

où l'astérisque $\star$ note le complexe conjugué, et la fonction de covariance de deux v.a. complexes Z et W est

$$COV[Z, W] = E[(Z^* - \mu_Z^*)(W - \mu_W)] = COV^*[Z, W].$$

2.2 Vecteurs aléatoires de dimension n

Les différentes notions introduites pour une paire de v.a. s'étendent très facilement à un vecteur aléatoire

$$\mathbf{X} = (X_1, X_2, \dots, X_n)$$

de dimension n quelconque. Par exemple, la fonction de répartition jointe (2.1) devient, avec $\mathbf{x} = (x_1, x_2, \dots, x_n)$,

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

tandis que la densité jointe de probabilité (2.5) s'écrit

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \frac{\partial^n F_{\mathbf{X}}(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}.$$

Les densités marginales de probabilités s'obtiennent en intégrant la densité jointe par rapport aux autres variables, par exemple

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_2 \dots dx_n.$$

On trouve les densités conditionnelles de manière similaire :

$$f_{X_1|X_2 \dots X_n}(x_1|x_2, \dots, x_n) = \frac{f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n)}{f_{X_2 \dots X_n}(x_2, \dots, x_n)}.$$

On peut étendre les types particuliers de variables aléatoires (uniformes, etc) présentés à la section précédente aux vecteurs aléatoires. Le cas de v.a. *gaussiennes* jointes mérite une attention particulière, et fait l'objet de la section suivante.

2.2.1 Variables aléatoires gaussiennes multivariées

Les v.a. $X_1, X_2, \dots, X_n$ sont des *variables aléatoires gaussiennes jointes* (on dit encore *multivariées*) ssi leur densité jointe de probabilité est

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}. \quad (2.28)$$

où

$$\begin{aligned} \boldsymbol{\mu} &= [E[X_1] \dots E[X_n]]^T \\ \Sigma &= \begin{bmatrix} \sigma_{X_1}^2 & COV[X_1, X_2] & \dots & COV[X_1, X_n] \\ COV[X_2, X_1] & \sigma_{X_2}^2 & \dots & COV[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ COV[X_n, X_1] & COV[X_n, X_2] & \dots & \sigma_{X_n}^2 \end{bmatrix} \end{aligned}$$

La matrice Σ est appelée *matrice de covariance*, on remarque qu'elle est symétrique, et on peut montrer qu'elle est semi-définie positive.

La fonction (2.28) est centrée au point $\boldsymbol{\mu}_{\mathbf{X}}$, et est constante pour les valeurs du vecteur $\mathbf{x}$ telles que $(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$ est constant. On définit ainsi des courbes de niveau qui peuvent être visualisées dans le cas bidimensionnel ($n = 2$) : ce sont des ellipses, dont l'orientation des axes dépend de σ_{X_1} , σ_{X_2} et ρ .

Les v.a. gaussiennes possèdent donc la propriété très intéressante d'être entièrement déterminées par les moments d'ordre 1 et 2.

La fonction caractéristique jointe de v.a. gaussiennes multivariées est une généralisation immédiate du cas uni-dimensionnel :

$$\Phi_{X_1 X_2 \dots X_n}(\omega_1, \omega_2, \dots, \omega_n) = \Phi_{\mathbf{X}}(\boldsymbol{\omega}) = e^{j\boldsymbol{\mu}^T \boldsymbol{\omega} - \frac{1}{2}\boldsymbol{\omega}^T \Sigma \boldsymbol{\omega}}. \quad (2.29)$$

Comme nous l'avons vu plus haut, en dérivant cette fonction caractéristique à l'origine un nombre déterminé de fois, on peut calculer tous les moments. Dans le cas gaussien, ces opérations donnent les résultats suivants, si $X_1, X_2, \dots, X_n$ sont des v.a. gaussiennes jointes et si X est une v.a. gaussienne de moyenne nulle,

$$E[X_1 X_2 X_3 X_4] = E[X_1 X_2]E[X_3 X_4] + E[X_1 X_3]E[X_2 X_4] + E[X_1 X_4]E[X_2 X_3] \quad (2.30)$$

$$E[X_1 X_2 \dots X_{2k+1}] = 0 \quad (2.31)$$

$$E[X^{2k+1}] = 0 \quad (2.32)$$

$$E[X^{2k}] = 1.3.5.7 \dots (2k-1) \sigma_X^{2k}. \quad (2.33)$$

Enfin, un changement de variables linéaire peut s'avérer particulièrement intéressant. La densité jointe de

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

est, à partir de (2.9)

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y})}{|\det(\mathbf{A})|} = \frac{1}{\sqrt{(2\pi)^n \det \Sigma} |\det(\mathbf{A})|} e^{-\frac{1}{2}(\mathbf{A}^{-1}\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{A}^{-1}\mathbf{y} - \boldsymbol{\mu})}. \quad (2.34)$$

Comme

$$\begin{aligned}(A^{-1}\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (A^{-1}\mathbf{y} - \boldsymbol{\mu}) &= (\mathbf{y} - A\boldsymbol{\mu})^T (A\Sigma A^T)^{-1} (\mathbf{y} - A\boldsymbol{\mu}) \\ &= (\mathbf{y} - \mathbf{m})^T S^{-1} (\mathbf{y} - \mathbf{m})\end{aligned}$$

avec

$$\begin{aligned}S &= A\Sigma A^T \\ \mathbf{m} &= A\boldsymbol{\mu},\end{aligned}$$

la relation (2.34) peut s'écrire

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^n |\det S|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{m})^T S^{-1}(\mathbf{y}-\mathbf{m})}$$

ce qui montre que $(Y_1, \dots, Y_n)$ sont également des v.a. gaussiennes jointes. Une transformation linéaire de v.a. gaussiennes jointes produit donc d'autres v.a. gaussiennes jointes. En particulier, on cherchera une transformation A qui diagonalise la matrice de covariance, de manière à décorréler les différentes composantes du vecteur aléatoire.

2.3 Convergence d'une suite de variables aléatoires

L'étude de la convergence d'une suite de variables aléatoires $\{X_n\}_{n \geq 1} = \{X_1, X_2, X_3, \dots, X_n, \dots\}$ lorsque $n \rightarrow \infty$ présente la même importance en probabilité que l'étude de la convergence d'une suite de réels $\{x_n\} = \{x_1, x_2, x_3, \dots, x_n, \dots\}$ lorsque $n \rightarrow \infty$ en analyse.

Nous allons sans entrer dans les détails rappeler les principaux différents types de convergence de suite de v.a.

2.3.1 Convergence presque sûre

Définition

Comme une variable aléatoire est une fonction de l'espace des résultats possibles d'une expérience, en fixant un résultat particulier ζ , la suite

$$\{X_1(\zeta), X_2(\zeta), X_3(\zeta), \dots, X_n(\zeta), \dots\}$$

est une suite de réels. Au contraire, en ne fixant pas un élément ζ particulier, on a une suite de variables aléatoires

$$\{X_1, X_2, X_3, \dots, X_n, \dots\}$$

qui représente donc un ensemble de suites de nombres réels.

Supposons que chacune de ces suites de réels converge vers un réel, c'est-à-dire que quel que soit le résultat ζ d'une expérience, la suite $\{X_n(\zeta)\}$ converge vers un certain $X(\zeta)$. Dans ce cas, on dit que la suite de v.a. $\{X_n\}$ converge *partout* ou encore *sûrement* vers la v.a. X .

Si maintenant il y a certains résultats ζ' pour lesquels la suite $\{X_n(\zeta')\}$ ne converge pas, mais que ces résultats ont une probabilité nulle d'être produits par l'expérience, ou de manière équivalente, qu'avec une probabilité 1 on obtienne un résultat ζ telle que la suite $\{X_n(\zeta)\}$ converge :

$$P(\{\zeta \mid \lim_{n \rightarrow \infty} X_n(\zeta) = X(\zeta)\}) = 1,$$

alors on dit que la suite de v.a. $\{X_n\}$ converge *presque partout* ou *presque sûrement* ou encore *avec probabilité 1* vers la v.a. X , ce qu'on note

$$P\left(\lim_{n \rightarrow \infty} X_n \rightarrow X\right) = 1 \quad (2.35)$$

ou encore

$$X_n \xrightarrow{\text{w.p.1}} X, \quad \text{ou} \quad X_n \xrightarrow{\text{a.s.}} X.$$

Pour établir la convergence presque sûre, nous avons besoin des lemmes suivants.

Continuité séquentielle des mesures de probabilités

Ce premier lemme technique montre que $P(\cdot)$ est une fonction continue des évènements A_n d'une suite infinie (dé)-croissante $\{A_1, A_2, A_3, \dots, A_n, \dots\}$.

Lemme 1 (Continuité des probabilités) (i) Soit une suite croissante d'évènements $\{A_n\}_{n \geq 1}$, i.e., tels que $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ et soit A leur limite :

$$A = \bigcup_{n=1}^{\infty} A_n = \lim_{n \rightarrow \infty} A_n$$

Alors $P(A) = \lim_{n \rightarrow \infty} P(A_n)$.

(ii) Soit une suite décroissante d'évènements $\{B_n\}_{n \geq 1}$, i.e., tels que $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$ et soit B leur limite :

$$B = \bigcap_{n=1}^{\infty} B_n = \lim_{n \rightarrow \infty} B_n$$

Alors $P(B) = \lim_{n \rightarrow \infty} P(B_n)$.

Pour démontrer la partie (i) de ce lemme, on réécrit $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$ comme l'union d'une suite d'évènements disjoints, de sorte que le troisième axiome A3' donne

$$\begin{aligned} P(A) &= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus A_2) + \dots \\ &= P(A_1) + \lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} P(A_{k+1} \setminus A_k) \\ &= P(A_1) + \lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} (P(A_{k+1}) - P(A_k)) \\ &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

On démontre la partie (ii) de manière similaire, en prenant les compléments des évènements de la suite $\{B_n\}_{n \geq 1}$

Lemmes de Borel-Cantelli

Considérons une suite d'événements indépendants $\{A_1, A_2, A_3, \dots, A_n, \dots\}$, et l'événement qu'on note $\{A_n \text{ i.o.}\}$ (i.o. pour infinitely often) de trouver une infinité d'indices n pour lesquels les événements A_n dans cette suite se produisent :

$$\{A_n \text{ i.o.}\} = \{\zeta \in \Omega \mid \zeta \in A_n \text{ pour une infinité d'indices } n \geq 1\}.$$

Les deux lemmes de Borel-Cantelli permettent de lier la probabilité de cet événement $P(A_n \text{ i.o.})$ à la somme des probabilités $P(A_n)$, pour tout $n \geq 1$. Le premier lemme s'applique à toute suite d'événements dont la somme des probabilités converge.

Lemme 2 (Premier Lemme de Borel-Cantelli) *Pour toute suite d'événements $\{A_n\}_{n \geq 1}$,*

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \implies \quad P(A_n \text{ i.o.}) = 0.$$

Pour démontrer ce lemme, souvenons-nous que les événements $A_1, A_2, A_3, \dots, A_n, \dots$ et $\{A_n \text{ i.o.}\}$ sont des sous-ensembles de Ω . Prenons un élément $\zeta \in \Omega$. On a alors les équivalences suivantes :

$$\begin{aligned} \zeta \in \{A_n \text{ i.o.}\} &\iff \text{pour tout } n \geq 1, \\ &\quad \zeta \text{ appartient à au moins un des sous-ensembles } A_n, A_{n+1}, \dots \\ &\iff \text{pour tout } n \geq 1, \zeta \in \bigcup_{k \geq n} A_k \\ &\iff \zeta \in \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k \end{aligned}$$

et donc

$$\{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k. \quad (2.36)$$

Comme la suite d'événements $\{\bigcup_{k \geq n} A_k\}_{n \geq 1}$ est décroissante, la partie (ii) du lemme 1 et (2.36) impliquent que

$$P(A_n \text{ i.o.}) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} A_k\right) \leq \lim_{n \rightarrow \infty} \sum_{k \geq n} P(A_k).$$

Enfin, comme $\sum_{k=1}^{\infty} P(A_k) < \infty$, la convergence de cette série implique que pour tout $\varepsilon > 0$ il existe un entier n suffisamment grand pour que $\sum_{k \geq n} P(A_k) < \varepsilon$, et par conséquent

$$\lim_{n \rightarrow \infty} \sum_{k \geq n} P(A_k) = 0,$$

ce qui établit le résultat.

Le second lemme fait presque office de résultat inverse, car il s'applique à toute suite d'événements *indépendants* dont la somme des probabilités diverge.

Lemme 3 (Second Lemme de Borel-Cantelli) *Pour toute suite d'évènements $\{A_n\}_{n \geq 1}$ indépendants,*

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad \implies \quad P(A_n \text{ i.o.}) = 1.$$

Pour démontrer ce lemme, on prend le complément de (2.36) :

$$\overline{\{A_n \text{ i.o.}\}} = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_k. \quad (2.37)$$

Comme la suite d'évènements $\{\bigcap_{k \geq n} \bar{A}_k\}_{n \geq 1}$ est croissante, la partie (i) du théorème 1 et (2.37) impliquent que

$$P(\overline{\{A_n \text{ i.o.}\}}) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_k\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{k \geq n} \bar{A}_k\right).$$

Comme les évènements $\bar{A}_k$ sont indépendants, et comme $1 - x \leq \exp(-x)$ si $x \geq 0$, pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned} P\left(\bigcap_{k \geq n} \bar{A}_k\right) &= \prod_{k \geq n} P(\bar{A}_k) = \prod_{k \geq n} (1 - P(A_k)) \\ &\leq \prod_{k \geq n} \exp(-P(A_k)) = \exp\left(-\sum_{k \geq n} P(A_k)\right) \\ &= \exp\left(-\sum_{k=1}^{\infty} P(A_k)\right) \cdot \exp\left(\sum_{k=1}^{n-1} P(A_k)\right) = 0, \end{aligned}$$

où la dernière égalité découle de l'hypothèse $\sum_{k=1}^{\infty} P(A_k) = \infty$ tandis que $\sum_{k=1}^{n-1} P(A_k) < \infty$ pour tout $n \in \mathbb{N}^*$. Dès lors,

$$P(A_n \text{ i.o.}) = 1 - P(\overline{\{A_n \text{ i.o.}\}}) = 1 - \lim_{n \rightarrow \infty} P\left(\bigcap_{k \geq n} \bar{A}_k\right) = 1.$$

Critère de convergence presque sûre

Le premier lemme de Borel-Cantelli nous permet d'établir une condition suffisante de convergence presque sûre.

Théorème 1 (Convergence presque sûre, condition suffisante) *Toute suite de v.a. $\{X_n\}_{n \geq 1}$ converge presque sûrement (i.e., avec probabilité 1) vers une v.a. X , si pour tout $\varepsilon > 0$*

$$\sum_{n=1}^{\infty} P(|X_n - X| \geq \varepsilon) < \infty. \quad (2.38)$$

Pour tout $\varepsilon > 0$, soit $\{a_n\}_{n \geq 1}$ une suite de réels non négatifs tels que $a_n \geq \varepsilon$ pour seulement un nombre fini d'indices $n \in \mathbb{N}^*$. Alors il existe un indice m suffisant grand mais fini tel que $a_n < \varepsilon$ pour tout $n > m$. Comme $a_n < \varepsilon$ pour tout $n > m$ et pour tout $\varepsilon > 0$, $\lim_{n \rightarrow \infty} a_n = 0$.

Soit $\zeta \in \Omega$ et $a_n = |X_n(\zeta) - X(\zeta)|$. Pour ζ fixé, on a donc que $\lim_{n \rightarrow \infty} a_n = 0$ si $|X_n(\zeta) - X(\zeta)| \geq \varepsilon$ pour seulement un nombre fini d'indices $n \in \mathbb{N}^*$. A cause de (2.38), le lemme 2 implique que la probabilité qu'il y ait un nombre fini d'indices $n \in \mathbb{N}^*$ tels que $|X_n(\zeta) - X(\zeta)| \geq \varepsilon$ vaut

$$1 - P(|X_n - X| \geq \varepsilon \text{ i.o.}) = 1.$$

Dès lors, pour tout $\zeta \in \Omega$, $P(\lim_{n \rightarrow \infty} a_n = 0)$, et donc $P(\lim_{n \rightarrow \infty} X_n = X) = 1$.

Dans le cas où les v.a. X_n sont indépendantes, le second lemme de Borel-Cantelli nous permet d'établir une condition nécessaire et suffisante de convergence presque sûre.

Théorème 2 (Convergence presque sûre, v.a. indépendantes) *Toute suite de v.a. $\{X_n\}_{n \geq 1}$ indépendantes converge presque sûrement (i.e., avec probabilité 1) vers une v.a. X , si et seulement si pour tout $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} P(|X_n - X| \geq \varepsilon) < \infty. \quad (2.39)$$

La condition suffisante est la même que dans le théorème précédent. Pour démontrer la condition nécessaire, supposons que (2.39) ne soit pas vérifiée. Dans ce cas, le lemme 3 entraîne que pour tout $\varepsilon > 0$,

$$P(|X_n - X| \geq \varepsilon \text{ i.o.}) = 1,$$

ce qui implique qu'avec probabilité 1, $a_n = |X_n(\zeta) - X(\zeta)| \geq \varepsilon$ pour un nombre infini de valeurs de l'indice $n \in \mathbb{N}$. De ce fait, $\lim_{n \rightarrow \infty} a_n \neq 0$, et donc $P(\lim_{n \rightarrow \infty} X_n = X) \neq 1$, ce qui montre que la condition (2.39) est nécessaire.

Enfin, le dernier critère est moins pratique que le précédent, mais permet d'établir une condition nécessaire et suffisante de convergence presque sûre sans devoir supposer les v.a. X_n indépendantes.

Théorème 3 (Convergence presque sûre) *Toute suite de v.a. $\{X_n\}_{n \geq 1}$ converge presque sûrement (i.e., avec probabilité 1) vers une v.a. X si et seulement si pour tout $\varepsilon > 0$,*

$$P(|X_n - X| \geq \varepsilon \text{ i.o.}) = 0. \quad (2.40)$$

($\Rightarrow$) Soit $\varepsilon > 0$. Observons que l'évènement $\{|X_n - X| \geq \varepsilon \text{ i.o.}\}$ est inclus dans l'ensemble $\Omega \setminus \{\zeta \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\zeta) = X(\zeta)\}$.

En effet, si pour un certain $\zeta \in \Omega$, $\lim_{n \rightarrow \infty} X_n(\zeta) = X(\zeta)$, alors la définition de limite entraîne que pour tout $\varepsilon > 0$, il existe $m \in \mathbb{N}^*$ suffisamment grand pour que $|X_n(\zeta) - X(\zeta)| < \varepsilon$ pour tout $n \geq m$. Par conséquent, $\zeta \notin \{|X_n - X| \geq \varepsilon \text{ i.o.}\}$ et donc

$$P(|X_n - X| \geq \varepsilon \text{ i.o.}) \leq 1 - P(\zeta \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\zeta) = X(\zeta)) = 0.$$

($\Leftarrow$) Soit $k \in \mathbb{N}^*$ un entier strictement positif, et soit N_k la v.a qui est le dernier indice n tel que $|X_n - X| \geq 1/k$. Si $|X_n - X| \geq 1/k$ pour une infinité d'indices $n \in \mathbb{N}^*$, on pose $N_k = \infty$. Le domaine de la v.a. N_k est donc $S_{N_k} = \mathbb{N}^* \cup \{\infty\}$. En prenant $\varepsilon = 1/k$ dans (2.40), on a $P(N_k = \infty) = 0$ pour tout $k \in \mathbb{N}^*$. Par conséquent, pour tout $k \in \mathbb{N}^*$, $P(N_k < \infty) = 1$, ce qui implique que $P(\lim_{n \rightarrow \infty} X_n = X) = 1$.

Exemples

Les exemples qui suivent s'appliquent tous à des suites de v.a. $\{X_n\}_{n \geq 1}$ indépendantes.

- **Suite 1.** Dans ce premier exemple, chaque v.a. X_n suit une distribution de Bernoulli dans $\{0, 1\}$, de paramètre $p = 1/n^2$:

$$\begin{aligned} X_n &= 1 && \text{avec probabilité } 1/n^2 \\ &= 0 && \text{avec probabilité } 1 - 1/n^2. \end{aligned} \quad (2.41)$$

Le théorème 1 montre que cette suite converge presque sûrement vers 0, car (2.38) est vérifiée avec $X = 0$.

- **Suite 2.** Dans ce second exemple, chaque v.a. X_n suit une distribution de Bernoulli dans $\{0, 1\}$, de paramètre $p = 1/n$:

$$\begin{aligned} X_n &= 1 && \text{avec probabilité } 1/n \\ &= 0 && \text{avec probabilité } 1 - 1/n. \end{aligned} \quad (2.42)$$

Le théorème 2 montre que cette suite ne converge pas presque sûrement vers 0, car (2.39) est vérifié avec $X = 0$.

- **Suite 3.** Dans ce troisième exemple,

$$\begin{aligned} X_n &= n && \text{avec probabilité } 1/n^2 \\ &= 0 && \text{avec probabilité } 1 - 1/n^2. \end{aligned} \quad (2.43)$$

Le théorème 1 montre que cette suite converge presque sûrement vers 0, car (2.38) est vérifiée avec $X = 0$.

- **Suite 4.** Dans ce dernier exemple,

$$\begin{aligned} X_n &= \sqrt{n} && \text{avec probabilité } 1/n \\ &= 0 && \text{avec probabilité } 1 - 1/n. \end{aligned} \quad (2.44)$$

Le théorème 2 montre que cette suite ne converge pas presque sûrement vers 0, car (2.39) est vérifiée avec $X = 0$.

2.3.2 Convergence en distribution et théorème central limite

Définition

Un autre type de convergence, beaucoup plus faible que le précédent, concerne les fonctions de répartition des v.a. X_n , que nous notons $F_n(\cdot)$. S'il existe une fonction continue $F(\cdot)$ qui est la fonction de répartition d'une v.a. X telle que pour tout x

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad (2.45)$$

alors on dit que la suite de v.a. $\{X_n\}$ converge *en distribution* vers la v.a. X , ce que nous noterons

$$X_n \xrightarrow{\text{dis}} X$$

Théorème central limite

Un exemple très important de convergence en distribution est fourni par le théorème central limite :

Soit $\{X_n\}$ une suite de v.a. indépendantes et identiquement distribuées¹ (i.i.d.), de moyenne μ et de variance σ^2 . A partir de cette suite de v.a., on construit une autre suite de v.a. $\{Z_n\}$ données par

$$Z_n = \sum_{i=1}^n \frac{X_i - \mu}{\sqrt{n\sigma^2}}. \quad (2.46)$$

Alors le théorème central limite énonce que la suite de v.a. $\{Z_n\}$ converge en distribution vers une v.a. normale Z :

$$Z_n \xrightarrow{\text{dis}} Z \sim N(0, 1)$$

c'est-à-dire

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx.$$

La démonstration de ce théorème illustre bien l'utilité de la fonction caractéristique. En effet, en

1. qui ont toutes la même loi de probabilité.

utilisant le fait que les v.a. X_n sont indépendantes et identiquement distribuées, on trouve

$$\begin{aligned}
\Phi_{Z_n}(\omega) &= E[\exp(j\omega Z_n)] = E\left[\exp\left(\sum_{i=1}^n \frac{j\omega(X_i - \mu)}{\sigma\sqrt{n}}\right)\right] \\
&= E\left[\exp\left(\frac{j\omega(X_1 - \mu)}{\sigma\sqrt{n}}\right) \exp\left(\frac{j\omega(X_2 - \mu)}{\sigma\sqrt{n}}\right) \dots \exp\left(\frac{j\omega(X_n - \mu)}{\sigma\sqrt{n}}\right)\right] \\
&= E\left[\exp\left(\frac{j\omega(X_1 - \mu)}{\sigma\sqrt{n}}\right)\right] E\left[\exp\left(\frac{j\omega(X_2 - \mu)}{\sigma\sqrt{n}}\right)\right] \dots E\left[\exp\left(\frac{j\omega(X_n - \mu)}{\sigma\sqrt{n}}\right)\right] \\
&= \left(E\left[\exp\left(\frac{j\omega(X - \mu)}{\sigma\sqrt{n}}\right)\right]\right)^n \\
&= \left(1 + \frac{j\omega}{\sigma\sqrt{n}}E[(X - \mu)] + \frac{(j\omega)^2}{2\sigma^2 n}E[(X - \mu)^2] + \dots\right)^n \\
&= \left(1 - \frac{\omega^2}{2n} + \dots\right)^n \\
&\approx \left(1 - \frac{\omega^2}{2n}\right)^n
\end{aligned}$$

Pour être tout à fait rigoureux, il faudrait considérer le reste de ce développement en série. Comme on peut montrer qu'il tend vers 0 lorsque $n \rightarrow \infty$, nous nous contentons de l'approximation au 2ème ordre faite à la dernière étape du développement donné ci-dessus, et on calcule que

$$\lim_{n \rightarrow \infty} \Phi_{Z_n}(\omega) = \lim_{n \rightarrow \infty} \left(1 - \frac{\omega^2}{2n}\right)^n = e^{-\omega^2/2} = \Phi_Z(\omega)$$

qui n'est autre que la fonction caractéristique d'une v.a. gaussienne Z de moyenne nulle et de variance unité, c'est-à-dire une v.a. normale. Comme la fonction caractéristique détermine de manière unique la densité de probabilité, nous avons montré que Z_n converge en distribution vers une v.a. normale Z .

Ce théorème est également valable sous des hypothèses plus faibles, et peut par exemple être étendu au cas où les v.a. X_n sont indépendantes, de même moyenne μ et de même variance σ^2 , mais pas nécessairement identiquement distribuées. Le théorème central limite justifie l'extraordinaire importance des v.a. gaussiennes : souvent, dans la nature, un phénomène aléatoire est le résultat d'une somme de nombreuses petites contributions indépendantes. Une force de ce théorème est de pouvoir déterminer la loi de probabilité de la v.a. résultante sans nécessiter une connaissance complète des lois de probabilité des v.a. X_n .

2.3.3 Convergence en probabilité et loi faible des grands nombres

Définition

A présent, on examine les probabilités $P(|X - X_n| > \varepsilon)$, où $\varepsilon > 0$. C'est une suite de réels, compris entre 0 et 1, et dépendant de n et de ε . Si cette suite converge vers 0 pour tout $\varepsilon > 0$,

i.e. si

$$\lim_{n \rightarrow \infty} P(|X - X_n| \geq \varepsilon) = 0 \quad \text{pour tout } \varepsilon > 0 \quad (2.47)$$

alors on dit que la suite de v.a. $\{X_n\}$ converge *en probabilité* vers la v.a. X (qui peut éventuellement être une constante).

Relation avec convergence presque sûre

Si la suite de v.a. $\{X_n\}$ converge presque sûrement vers la v.a. X , alors elle converge en probabilité vers la v.a. X .

En effet, le théorème 3 entraîne que

$$P(|X - X_n| \geq \varepsilon \text{ i.o.}) = 0,$$

ou encore, à cause de (2.37) avec $A_n = \{|X - X_n| > \varepsilon\}$,

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{|X - X_k| \geq \varepsilon\}\right) = 0,$$

qu'on peut encore être écrire, à cause la partie (ii) du lemme 1 et vu que la suite d'évènements $\{\bigcup_{k \geq n} |X - X_k| \geq \varepsilon\}_{n \geq 1}$ est décroissante,

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{k \geq n} \{|X - X_k| \geq \varepsilon\}\right) = 0,$$

ce qui implique à son tour que

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

et prouve que la suite de v.a. $\{X_n\}$ converge *en probabilité* vers X .

Loi faible des grands nombres

Un exemple particulier de convergence en probabilité est fourni par la loi faible des grands nombres :

Soit $\{X_n\}$ une suite de v.a. indépendantes, de même moyenne μ et de même variance σ^2 . A partir de cette suite de v.a., on construit une autre suite de v.a. $\{\bar{X}_n\}$ données par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (2.48)$$

Alors la loi faible des grands nombres énonce que la suite de v.a. $\{\bar{X}_n\}$ converge en probabilité vers μ :

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad \text{pour tout } \varepsilon > 0.$$

La démonstration de ce théorème est basée sur l'inégalité de Tchebycheff.

On calcule d'abord la moyenne de chaque v.a. $\bar{X}_n$

$$E[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} n\mu = \mu$$

et sa variance

$$\begin{aligned} VAR[\bar{X}_n] &= E[(\bar{X}_n - \mu)^2] = E\left[\frac{1}{n^2} \left(\sum_{i=1}^n X_i - n\mu\right)^2\right] = \frac{1}{n^2} E\left[\left(\sum_{i=1}^n (X_i - \mu)\right)^2\right] \\ &= \frac{1}{n^2} E\left[\sum_{i=1}^n (X_i - \mu)^2\right] + \frac{2}{n^2} E\left[\sum_{i \neq j} (X_i - \mu)(X_j - \mu)\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n E[(X_i - \mu)^2] + \frac{2}{n^2} \sum_{i \neq j} E[(X_i - \mu)(X_j - \mu)] \\ &= \frac{1}{n^2} n\sigma^2 + 0 = \frac{\sigma^2}{n}. \end{aligned}$$

Par conséquent, l'inégalité de Tchebycheff entraîne que

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{VAR[\bar{X}_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2},$$

ce qui tend bien vers 0 quand $n \rightarrow \infty$.

Exemples

On vérifie que les 4 exemples de suites données à la sous-section 2.3.1 sont tous des suites convergeant en probabilité vers 0, car $P(|X_n| > \varepsilon) \rightarrow 0$ pour $n \rightarrow \infty$ pour chacune d'elles.

2.3.4 Convergence en moyenne quadratique

Définition

Une suite de v.a. $\{X_n\}$ converge *en moyenne quadratique* vers la v.a. X si

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0, \quad (2.49)$$

ce qu'on note parfois

$$\text{l.i.m. } X_n = X$$

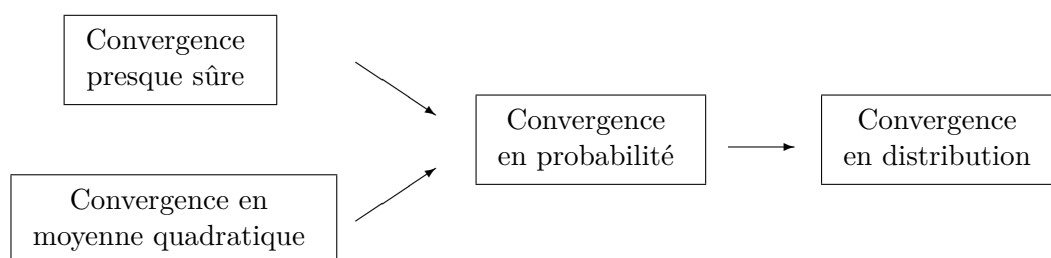
l'abréviation l.i.m. signifiant limit in mean (square).

Exemples

On vérifie que parmi les 4 exemples de suites données à la sous-section 2.3.1, seuls les deux premières convergent en moyenne quadratique vers 0, car $E[(X_n - 0)^2] \rightarrow 0$ pour $n \rightarrow \infty$ pour chacune des deux premières, alors que pour les deux dernières, $E[(X_n - 0)^2] = 1$ pour tout $n \geq 1$.

2.3.5 Relations entre les types de convergence

On peut relier ces différents types de convergence de la manière suivante, les flèches représentant une implication :



Les seuls types de convergence n'ayant pas de relation directe entre eux sont la convergence presque sûre et la convergence quadratique, et le type de convergence le plus faible est la convergence en distribution.

2.4 Exercices

- Des données sont enregistrées sur un disque à n pistes concentriques de telle manière qu'on puisse faire l'hypothèse que les mouvements successifs du bras de lecture soient indépendants les uns des autres. La probabilité que le bras aille chercher l'information sur la piste i est donnée par p_i ($1 \leq i \leq n$). Soit X la variable aléatoire donnant le nombre de pistes que le bras traverse entre deux lectures successives. X prend donc une des valeurs $0, 1, 2, \dots, n-1$.
 - Que vaut la probabilité $P(X = 0)$ que le bras n'ait pas bougé entre deux opérations de lecture successives, en fonction des p_i ($1 \leq i \leq n$) ? (Hint : il peut être commode d'utiliser deux variables aléatoires Y et Z donnant le numéro de la piste où se trouve le bras de lecture respectivement avant et après son déplacement).
 - Que valent les probabilités $P(X = j)$, en fonction des p_i ($1 \leq i \leq n$), pour $1 \leq j \leq n-1$?
 - Si $p_i = 1/n$ pour $1 \leq i \leq n$, que vaut le déplacement moyen $E[X]$ du bras ?
 - Si $p_i = 1/n$ pour $1 \leq i \leq n$, que vaut la variance de ce déplacement du bras ?

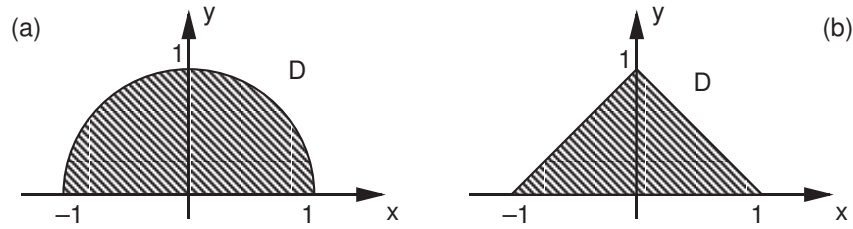


FIGURE 2.1 – Domaines définis aux exercices 2 (a) et 3 (b).

2. La densité de probabilité jointe de X et Y est uniforme dans le domaine D représenté à la figure 2.1 (a) :

$$f_{XY}(x, y) = k \quad \text{pour } x, y \in D. \quad (2.50)$$

- (a) Que vaut k pour que (2.50) soit effectivement une densité jointe de probabilité ?
 - (b) Quelles sont les densités marginales de X et de Y ?
 - (c) Quelle est la densité de probabilité conditionnelle $f_{Y|X}(y|x)$?
3. Répéter l'exercice 2 où le domaine D est à présent celui représenté à la figure 2.1 (b).
4. Soient X et Y deux variables aléatoires normales indépendantes (i.e., gaussiennes de moyennes nulles et de variances unités).

- (a) Quelle est la densité jointe de probabilité de R et Θ où

$$\begin{aligned} R &= \sqrt{X^2 + Y^2} \\ \Theta &= \begin{cases} \text{Arctg}(Y/X) + \frac{\pi}{2} & \text{si } x > 0 \\ \text{Arctg}(Y/X) + \frac{3\pi}{2} & \text{si } x \leq 0 \end{cases} \end{aligned}$$

avec $R \geq 0$ et $0 \leq \Theta < 2\pi$? Remarque : si le jacobien de cette transformation de variables est désigné par

$$J(x, y) = \det \begin{bmatrix} \frac{\partial r}{\partial x} & \frac{\partial \theta}{\partial x} \\ \frac{\partial r}{\partial y} & \frac{\partial \theta}{\partial y} \end{bmatrix}$$

et si le jacobien de la transformation inverse est

$$J^{-1}(r, \theta) = \det \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix},$$

on peut montrer qu'ils ont liés par la relation

$$|J(x, y)| = 1/|J^{-1}(r, \theta)|.$$

- (b) Les deux variables R et Θ sont-elles indépendantes ?

- (c) Quelle est la densité marginale de R ? Cette variable aléatoire est appelée *variable aléatoire de Rayleigh*.
- (d) Quelle est la valeur de a pour laquelle $P(R \leq a) = 1/2$?
5. Soient $X_1, X_2, X_3, \dots$ une suite d'échantillons de parole de moyennes nulles et de variances σ^2 . Supposons de plus que la covariance entre échantillons soit donnée par une fonction

$$COV(X_i, X_j) = \rho_{|i-j|} \sigma^2.$$

Des valeurs typiques de ρ_1 et de ρ_2 pour des signaux de parole sont $\rho_1 = 0.825$ et $\rho_2 = 0.562$. Un prédicteur linéaire du second ordre de ce signal de parole est $\hat{X}_i = aX_{i-1} + bX_{i-2}$ où les coefficients de prédiction a et b (en anglais : LPC coefficients) minimisent l'erreur quadratique moyenne entre la valeur réelle de l'échantillon X_i et sa valeur prédite $\hat{X}_i$, à savoir

$$\varepsilon = E[(X_i - \hat{X}_i)^2].$$

Calculer ce prédicteur (c'est-à-dire les coefficients de prédiction a et b . N. B. : on reviendra sur ce type de problème de manière plus approfondie).

6. Soient $X_1, X_2, \dots, X_n$ des variables aléatoires de même moyenne μ et de covariance

$$COV(X_i, X_j) = \begin{cases} \sigma^2 & \text{si } i = j \\ \rho\sigma^2 & \text{si } |i - j| = 1 \\ 0 & \text{sinon,} \end{cases}$$

avec $|\rho| < 1$. Calculer la moyenne et la variance de la variable aléatoire

$$S_n = \sum_{i=1}^n X_i.$$

7. Démontrer la propriété P8. Hint : $E[(X - \lambda Y)^2] \geq 0$ pour tout réel λ .
8. Soit Θ une v.a. uniforme dans l'intervalle $[0, 2\pi]$, et soient les deux v.a. $X = \cos \Theta$ et $Y = \sin \Theta$. Les v.a. X et Y sont-elles orthogonales? corrélées? indépendantes?
9. Soit $\{a_n\}$ une suite de réels qui converge vers un réel a : $\lim_{n \rightarrow \infty} a_n = a$, et soit une suite de v.a. $\{X_n\}$ telle que

$$\lim_{n \rightarrow \infty} E[(X_n - a_n)^2] = 0.$$

Démontrer que la suite de v.a. $\{X_n\}$ converge en moyenne quadratique vers a .

10. Soit $Z = X + Y$ où X et Y sont deux v.a. continues indépendantes, et où Y est une v.a. normale $N(0, 1)$. Que vaut la densité de probabilité conditionnelle

$$f_{Z|X}(z|x) \quad ?$$

11. Soient $X_1, X_2, \dots, X_n$ des variables aléatoires indépendantes discrètes dont la loi de probabilité est

$$P(X_i = k) = 1/N$$

pour $k = 1, 2, \dots, N$, et soit

$$\Theta_n = \min\{X_1, X_2, \dots, X_n\}.$$

- (a) Calculer $P(\Theta_n \geq k)$, avec $k = 1, 2, \dots, N$.
 (b) Calculer la loi de probabilité $P(\Theta_n = k)$ de la v.a. Θ_n . Hint : mettez le problème sous une forme où il est commode d'utiliser le résultat trouvé en a).
 12. Soient X et Y des variables aléatoires gaussiennes jointes dont le vecteur des moyennes et la matrice de covariance sont respectivement

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$$

avec $\sigma_X, \sigma_Y > 0$.

- (a) Quelle est la densité marginale $f_Y(y)$ de Y ?
 (b) Calculer la densité conditionnelle $f_{X|Y}(x|y)$.
 (c) Calculer la moyenne conditionnelle $E[X|Y = y]$. Pour quelle valeur de ρ cette moyenne conditionnelle vaut-elle μ_X ?
 (d) Calculer la variance conditionnelle $VAR[X|Y = y]$.
 13. Soient $X_1, X_2, \dots, X_n, \dots$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.), et soit $S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$. Calculez, pour $1 \leq m \leq n$,

$$E \left[\frac{S_m}{S_n} \right].$$

Hint : commencez par le cas (trivial) où $m = n$, puis par celui (non trivial) où $m = 1$, et utilisez le fait que les X_i sont i.i.d.

Remarque importante : Si X est une v.a. discrète et si Y est une v.a. continue, de densité de probabilité $f_Y(y)$, le théorème des probabilités totales s'écrit

$$P(X = x_i) = \int_{-\infty}^{+\infty} P(X = x_i|Y = y)f_Y(y)dy.$$

14. Soit X une variable aléatoire de Poisson, dont le paramètre Λ est lui-même une variable aléatoire exponentielle d'intensité 1, au lieu d'une quantité constante déterministe λT . Ici donc $f_\Lambda(\lambda) = e^{-\lambda}$ avec $S_\Lambda = [0, +\infty[$.
 (a) Calculer la probabilité $P(X = n)$. Hint : la solution est $P(X = n) = 1/2^{n+1}$.
 (b) Calculer la moyenne $E[X]$.

(c) Calculer la variance $VAR[X]$.

15. Soient $X_1, X_2, \dots, X_n, \dots$ une suite de variables aléatoires continues indépendantes et identiquement distribuées (i.i.d.). On dit qu'un "record" se produit au temps n si $X_n > \max\{X_1, X_2, \dots, X_{n-1}\}$. Autrement dit, X_n est un "record" s'il est plus grand que chaque X_i , $1 \leq i \leq n-1$, ou encore si $X_n = \max\{X_1, X_2, \dots, X_{n-1}, X_n\}$.

(a) Calculer la probabilité qu'il y ait un record au temps n .

(b) Calculer le nombre moyen de records qu'il y a eu au temps n . Hint : le nombre de records au temps n peut s'écrire comme une somme de variables de Bernoulli, ou encore de fonctions indicatrices $I(\text{"Record au temps } i\text{"})$ de l'événement "un record se passe au temps i " :

$$I(\text{"Record au temps } i\text{"}) = \begin{cases} 1 & \text{si } X_i > \max\{X_1, X_2, \dots, X_{i-1}\} \\ 0 & \text{sinon.} \end{cases}$$

En effet, on a alors

$$E[\text{"Nombre de records au temps } n"] = E\left[\sum_{i=0}^n I(\text{"Record au temps } i\text{"})\right] = \dots$$

16. Soit X une variable aléatoire, et $Y = aX + b$ où a et b sont deux constantes réelles. Que vaut le coefficient de corrélation $\rho(X, Y)$ entre X et Y ?
17. Soient X_1, X_2 et X_3 trois v.a. normales indépendantes. Calculer la densité de probabilité jointe de Y_1, Y_2 et Y_3 où

$$\begin{aligned} Y_1 &= X_1 + X_2 + X_3 \\ Y_2 &= X_1 - X_2 \\ Y_3 &= X_1 - X_3. \end{aligned}$$

18. Soient X et Y deux variables aléatoires indépendantes suivant des lois Gamma de paramètres respectifs (λ, α) et (λ, β) (i.e. $f_X(x) = \lambda(\lambda x)^{\alpha-1} e^{-\lambda x} / \Gamma(\alpha)$, etc).

(a) Quelle est la densité de probabilité jointe de $U = X + Y$ et $V = X/(X + Y)$?

(b) Les deux variables U et V sont-elles indépendantes ?

(c) Montrer que U est une v.a. Gamma. Quels sont ses paramètres ? N.B. La loi de la v.a. V est appelée loi Bêta.

19. Démontrer que la convergence en moyenne quadratique entraîne la convergence en probabilité, i.e. que si $\{X_n\}_{n \geq 1}$ est une suite de v.a. qui converge en moyenne quadratique vers une v.a. X , alors cette suite converge également en probabilité vers la v.a. X . Le contraire est-il vrai ?

20. Démontrer que $\{X_n\}_{n \geq 1}$ est une suite de v.a. qui converge en probabilité vers 0 si et seulement si

$$\lim_{n \rightarrow \infty} E \left[\frac{|X_n|}{1 + |X_n|} \right] = 0.$$

Hint : observez que (i) $|X_n|/(1 + |X_n|) \leq 1$ et que (ii) $|X_n| \geq \varepsilon$ si et seulement si

$$\frac{|X_n|}{1 + |X_n|} \geq \frac{\varepsilon}{1 + \varepsilon}.$$

Enfin, pour tout v.a. X et toute partition $[A_1, A_2, \dots, A_m]$ de Ω , le théorème des probabilités totales stipule que pour tout $x \in S_X$,

$$P(X \leq x) = \sum_{i=1}^m P(X \leq x \mid A_i)P(A_i)$$

et donc en prenant les densités

$$f_X(x) = \sum_{i=1}^m f_X(x \mid A_i)P(A_i)$$

et les espérances d'une fonction $g(X)$

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x)f_X(x)dx = \int_{-\infty}^{\infty} g(x) \left(\sum_{i=1}^m f_X(x \mid A_i)P(A_i) \right) dx \\ &= \sum_{i=1}^m \left(\int_{-\infty}^{\infty} (g(x)f_X(x \mid A_i))dx \right) P(A_i) = \sum_{i=1}^m E[g(X) \mid A_i]P(A_i) \end{aligned}$$

ce permet d'écrire le le théorème des probabilités totales en termes d'espérance de la v.a. $g(X)$ comme

$$E[g(X)] = \sum_{i=1}^m E[g(X) \mid A_i]P(A_i)$$