Probabilités II

STEP, MINES ParisTech*

7 octobre 2019 (#24161b4)

Table des matières

*Ce document est un des produits du projet boisgera/CDIS, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS).

Variables aléatoires réelles

En théorie moderne des probabilités, on préfère prendre un point de vue fonctionnel plutôt qu'ensembliste, et utiliser les variables aléatoires plutôt que les événements. Une variable aléatoire est une grandeur qui dépend du résultat de l'expérience. Par exemple,

— le nombre de 6 obtenus dans un lancer de 3 dés,
— le nombre d'appels dans un central téléphonique pendant une heure,
— la distance du point d'atteinte d'une flèche au centre de la cible,
— la valeur maximale d'un prix d'actif sur un intervalle de temps donné,

sont des variables aléatoires.

La définition formelle d'une variable aléatoire fait intervenir des éléments de la théorie de la mesure qui nous font pour l'instant défaut. On s'interessera dans un premier temps au cas d'une variable réelle dont on donne une définition partielle :

Soit Ω l'espace fondamental muni de sa tribu $\mathcal{A}$. Une *variable aléatoire* X est une application de $(\Omega, \mathcal{A})$ dans un ensemble E,

$$\omega \in \Omega \mapsto X(\omega) \in E$$

En pratique, l'ensemble E pourra être un ensemble fini ou dénombrable ou $\mathbb{R}$ ou $\mathbb{R}^d$ ou encore un espace plus sophistiqué tel que l'ensemble $C(\mathbb{R}_+, \mathbb{R}^d)$ des fonctions continues de $\mathbb{R}_+$ dans $\mathbb{R}^d$.

Remarque

La terminologie, consacrée par l'usage, peut être trompeuse. Une variable aléatoire n'est pas une variable (au sens de l'analyse) mais une fonction. Cette terminologie est apparentée à la notion de variable en physique ou en sciences humaines où on désigne volontiers par "variable" la valeur prise par une fonction de l'état du système étudié.

L'intérêt principal de travailler avec des variables aléatoires est de pouvoir substituer à l'espace abstrait Ω des résultats de l'expérience l'espace E, mieux connu dans la pratique. Ainsi, grâce à une variable aléatoire X, nous pouvons transporter la structure abstraite du modèle probabiliste $(\Omega, \mathcal{A}, \mathbb{P})$ sur l'espace d'arrivée E, en posant pour $B \subset E$

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega, X(\omega) \in B\}) \tag{1}$$

Cette formule défini une nouvelle probabilité, notée $\mathbb{P}_X$ et définie sur E, qui s'appelle la *loi de la variable* X.

NOTATION

Il est usuel de noter l'ensemble $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ par $\{X \in B\}$, ce qui allège les écritures. On se rappelera néanmoins que cette notation désigne un sous-ensemble de Ω.

Comme $\mathbb{P}(A)$ n'est définie que pour les A de la tribu $\mathcal{A}$, la formule (1) ne permet de définir $\mathbb{P}_X(B)$ que pour les ensembles B tels que $X^{-1}(B) \in \mathcal{A}$, d'où l'importance de la proposition suivante :

Proposition

a) La famille $\mathcal{E}$ des parties B de E telles que $X^{-1}(B) \in \mathcal{A}$ est une tribu de E.

b) L'application $\mathbb{P}_X$ définie pour $B \in \mathcal{E}$ par

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$$

définit une probabilité sur le couple $(E, \mathcal{E})$.

Démonstration Les 3 propriétés de la définition d'une tribu pour $\mathcal{E}$ ainsi que les deux propriétés de la définition de la probabilité pour $\mathbb{P}_X$ découlent immédiatement des mêmes propriétés pour $\mathcal{A}$ et $\mathbb{P}$, une fois remarquées les propriétés élémentaires suivantes :

$$X^{-1}(\varnothing) = \varnothing, X^{-1}(E) = \Omega, X^{-1}(B^c) = X^{-1}(B)^c$$
$$X^{-1}(\cap_i A_i) = \cap_i X^{-1}(A_i), X^{-1}(\cup_i A_i) = \cup_i X^{-1}(A_i)$$

■

$\mathbb{P}_X$ sera plus facile à caractériser que $\mathbb{P}$ puisque E est un ensemble connu (on pourra en particulier utiliser ses propriétés topologiques) alors que Ω est un espace abstrait. Les variables que nous rencontrerons dans ce cours seront soit à valeurs dans un ensemble dénombrable, soit à valeurs dans $\mathbb{R}$ ou dans $\mathbb{R}^d$. Nous les appellerons respectivement des variables aléatoires discrètes, réelles ou des vecteurs aléatoires. Leurs lois seront alors des probabilités respectivement sur un ensemble dénombrable, sur $\mathbb{R}$ ou sur $\mathbb{R}^d$. Le cas discret est considéré connu.

La proposition ci-dessus implique que l'ensemble $X^{-1}(B)$ soit un évènement, pour tout B dans $\mathcal{E}$. Dans le cas où $E = \mathbb{R}$, on notera $\mathcal{E}_R$ la tribu associée [1]. Cela nous conduit à poser :

Définition – variable aléatoire réelle

Soit l'espace d'état Ω munit de la tribu $\mathcal{A}$ des évènements. Une application X de Ω dans $\mathbb{R}$ est une *variable aléatoire réelle* si $X^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{E}_{\mathbb{R}}$.

Définition – loi d'une variable aléatoire réelle

La probabilité $\mathbb{P}_X$, définie sur $(\mathbb{R}, \mathcal{E}_{\mathbb{R}})$ par $\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$ pour $B \in \mathcal{E}_{\mathbb{R}}$ est appelée *loi de la variable* X, ou *distribution* de X.

On a alors le résultat très utile suivant que nous admettrons dans un premier temps.

1. Nous n'avons pas les outils permettant de caractériser cette tribu pour le moment. On verra par la suite qu'elle est très similaire à la tribu des ensembles mesurables de $\mathbb{R}$, à une collection d'ensembles négligeables près.

Proposition

Si $X_1, \dots, X_n$ sont des variables aléatoires réelles et si g est une fonction continue par morceaux de $\mathbb{R}^n$ dans $\mathbb{R}$, alors $Y = g(X_1, \dots, X_n)$ est une variable aléatoire réelle [2].

Comme application de ce résultat, on a les propriétés suivantes :

Proposition

Soient X, Y et $(X_n)_{n\in\mathbb{N}^\star}$ des variables aléatoires réelles. On a

1. $X+Y$, XY, $\frac{X}{Y}$ si $Y \neq O$, sont des variables aléatoires.
2. $\sup_{1\leq p} X_n$, $\inf_{1\leq p} X_n$, sont des variables aléatoires.
3. $\sup_{n\geq 1} X_n$, $\inf_{n\geq 1} X_n$, sont des variables aléatoires.
4. Si $X_n(\omega) \to n \to \infty Z(\omega)$, $\forall\omega$, alors la limite Z est une variable aléatoire.
5. $Z = 1_A$ est une variable aléatoire $\Leftrightarrow A \in \mathcal{A}$.

Définition – variable aléatoire réelle à densité

Soit X une variable aléatoire. On dit que X a une *loi de densité* f (ou par abus de language "est de densité f"), si $\mathbb{P}_X$ admet la densité f et donc si pour tout réel x,

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy.$$

Exemple

La durée de fonctionnement d'un ordinateur avant sa première panne est une variable aléatoire positive de loi exponentielle de paramètre $1/100$, dont la densité est donnée par

$$f(x) = \begin{cases} \frac{1}{100}\exp\left(-\frac{x}{100}\right) & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Calculons la probabilité que cette durée de fonctionnement X soit comprise entre 50 et 150 heures, elle vaut

$$\mathbb{P}(X \in [50,150]) = \int_{50}^{150} \frac{1}{100}\exp\left(-\frac{x}{100}\right)dx = \exp(-1/2) - \exp(-3/2) \approx 0,38.$$

2. Ce résultat est en fait valable dans un cadre plus général que nous détaillerons dans la suite.

Calculons la probabilité que l'ordinateur fonctionne moins de 100 heures :

$$\mathbb{P}(X \leq 100) = \int_0^{100} \frac{1}{100} \exp\left(-\frac{x}{100}\right) dx = 1 - e^{-1} \approx 0,63.$$

Moments d'une variable aléatoire à densité

La densité de probabilité d'une variable aléatoire va nous permettre de calculer aisément des grandeurs caractéristiques telles que sa valeur moyenne et sa variance définies ci-dessous :

Définition

La variable aléatoire $X : \Omega \to \mathbb{R}$ de densité f est dite *intégrable* si l'intégrale $|x|f(x)$ est définie, autrement dit si le produit $xf(x)$ est absolument intégrable [3]. On définit alors son *espérance* par

$$\mathbb{E}(X) = \int_{\mathbb{R}} xf(x)dx$$

Remarque

$\mathbb{E}(X)$ est un nombre réel qui donne une valeur moyenne résumant la variable aléatoire X.

L'espérance mathématique d'une variable aléatoire est un concept fondamental de la théorie des probabilités. La dénomination d'espérance pour cette quantité fait référence aux problèmes de jeux et d'espérance de gain. Cette terminologie imagée a été introduite par Pascal.

On note $\mathcal{L}^1$ l'ensemble de toutes les variables réelles X à densité intégrables. Les propriétés suivantes découlent directement des propriétés de l'intégrale.

Proposition

— $\mathcal{L}^1$ est un espace vectoriel et $\forall X, Y \in \mathcal{L}^1$, $\forall a, b \in \mathbb{R}$

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

3. Comme f est positive, on peut en fait se convaincre que $xf(x)$ intégrable équivaut à $xf(x)$ absolument intégrable : si $xf(x)$ est intégrable, ses "restrictions" $g(x) = xf(x)1_{\mathbb{R}_-}(x)$ et $h(x) = xf(x)1_{\mathbb{R}_+}(x)$ sont intégrables (passer par la restriction aux intervalles $\mathbb{R}_-$ et $\mathbb{R}_+$ puis par le critère qui étend par 0 à $\mathbb{R}$; les deux opération préservent l'intégrabilité). Or $|x|f(x) = g(x) - h(x)$ (sauf en 0), donc elle est intégrable.

— $X \in \mathcal{L}^1 \leftrightarrow |X| \in \mathcal{L}^1$, et dans ce cas

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$

— Si $X \geq 0$ et $X \in \mathcal{L}^1$, alors $\mathbb{E}(x) \geq 0$.
— Si $X, Y \in \mathcal{L}^1$ sont telles que $X \leq Y$, alors

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

— L'espérance d'une variable presque-sûrement constante est égale à cette constante :

$$\text{Si } \mathbb{P}(X(\omega) = a) = 1, \text{ alors } \mathbb{E}(X) = a.$$

— Si $\exists b \in \mathbb{R}$ tel que $|X| \leq b$, alors $X \in \mathcal{L}^1$ et $\mathbb{E}(X) \leq b$.

Rappel : cas discret

Dans le cas d'une variable aléatoire discrète Y à valeurs dans $\mathbb{N}$, son espérance est définie par la quantité $\mathbb{E}(Y) = \sum_{i \in \mathbb{N}} i\mathbb{P}(Y = i)$, pourvu que celle-ci soit finie. On voit immédiatement que les propriétés ci-dessus sont également vérifiées.

Outre l'espace $\mathcal{L}^1$, nous pouvons définir l'espace $\mathcal{L}^2$ des variables aléatoires réelles dont le carré X^2 est dans $\mathcal{L}^1$.

Définition

La variable aléatoire $X : \Omega \to \mathbb{R}$ de densité f est dite *de carré intégrable* si $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f(x)dx < +\infty$, autrement dit si le produit $x^2 f(x)$ est intégrable. Sa *variance* est définie par

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

Proposition

$\mathcal{L}^2$ est un sous-espace vectoriel de $\mathcal{L}^1$, et si $X \in \mathcal{L}^2$,

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}$$

Démonstration Soient X et Y deux variables aléatoires réelles de $\mathcal{L}^2$ et $a \in \mathbb{R}$. Comme $(aX + Y)^2 \leq 2a^2X^2 + 2Y^2$, alors $aX + Y \in \mathcal{L}^2$. Ainsi, $\mathcal{L}^2$ est un espace vectoriel.

L'inclusion $\mathcal{L}^2 \subset \mathcal{L}^1$ découle de $|X| \leq 1 + |X^2|$ et de la proposition précédente.

La première inégalité a déjà été vue ci-dessus. Pour la seconde, nous pouvons nous limiter au cas où X est positive. Soit alors $a = \mathbb{E}(X)$ et $Y = X - a$. Par linéarité, on a

$$\mathbb{E}(Y^2) = \mathbb{E}(X^2) - 2a\mathbb{E}(X) + a^2 = \mathbb{E}(X^2) - a^2.$$

Et $\mathbb{E}(Y^2) \geq 0$ par le troisième point de la proposition ci-dessus. Par conséquent, $\mathbb{E}(X)^2 = a \leq \mathbb{E}(X^2)$ ce qui est le résultat recherché. ■

Remarque

En vertu de cette proposition, $\mathbb{V}(X)$ est **positive** et sa racine carrée σ_X s'appelle l'*écart-type* de X. L'écart-type est une grandeur qui mesure la moyenne (en un certain sens) de l'écart des valeurs de X à sa moyenne, d'où son nom.

On a également

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

que l'on obtient en développant $(X - \mathbb{E}(X))^2$. Cette manipulation anodine est fort utile dans la pratique. On retiendra que "La variance est égale à la moyenne des carrés moins le carré de la moyenne". On désigne le terme $\mathbb{E}(X^2)$ par l'expression *moment d'ordre deux* tandis que la variance est parfois appelée *moment centré d'ordre deux*.

Exemple : TODO illustration écart type

On peut remarquer que si X et Y sont dans $\mathcal{L}^2$, la variable aléatoire XY est dans $\mathcal{L}^1$. En effet, on a $|XY| \leq \frac{1}{2}(X^2 + Y^2)$. On peut ainsi définir la *covariance* de deux variables aléatoires :

Définition

Si X et Y sont dans $\mathcal{L}^2$, la variable aléatoire $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$ est intégrable. On appelle la *covariance* de X et Y l'espérance de cette variable aléatoire et on la note :

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

Le *coefficient de corrélation* des variables aléatoires X et Y est le nombre

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}$$

qui est bien défini lorsque $\mathbb{V}(X) > 0$ et $\mathbb{V}(Y) > 0$.

Du fait de la linéarité de l'espérance, on a

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

et d'ailleurs, on voit que la formule de calcul de la variance donnée plus haut est un cas particulier de cette formulation car $\mathbb{V}(X) = \mathrm{Cov}(X, X)$. La linéarité de l'espérance nous donne encore pour $a, a', b, b' \in \mathbb{R}$

$$\mathbb{E}((aX + b)(a'Y + b)) = aa'\mathbb{E}(XY) + ab'\mathbb{E}(X) + a'b\mathbb{E}(Y) + bb'$$
$$\mathbb{E}(aX + b)\mathbb{E}(a'Y + b) = aa'\mathbb{E}(X)\mathbb{E}(Y) + ab'\mathbb{E}(X) + a'b\mathbb{E}(Y) + bb'$$

On en déduit que la covariance est une forme bilinéaire sur l'espace vectoriel des variables aléatoires de carré intégrable, et nous avons

$$\mathrm{Cov}(aX + b, a'Y + b') = aa'Cov(X, Y)$$

En particulier, on a

$$\mathbb{V}(aX) = a^2\mathbb{V}(X)$$

On en déduit que les coefficients de corrélation de X et Y et de $aX+b$ et $a'Y+b'$ sont égaux lorsque $aa' > 0$.

Remarque

D'après ce qui précède, si X est une variable aléatoire de carré intégrable, d'espérance $\mathbb{E}(X)$ et d'écart-type $\sigma_X > 0$, alors la variable aléatoire

$$\frac{X - \mathbb{E}(X)}{\sigma_X}$$

est d'espérance nulle et de variance 1. On dira qu'une telle variable aléatoire est *centrée et réduite*.

Inégalité de Cauchy-Schwartz

Soit X et Y deux variables aléatoires de carré intégrable, alors on a *l'inégalité de Cauchy-Schwartz* :

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

avec égalité si et seulement si X et Y sont presque-sûrement proportionnelles.

Démonstration La première inégalité est évidente. Pour la seconde, on a $\forall x \in \mathbb{R}$ d'après la linéarité de l'espérance :

$$x^2\mathbb{E}(X^2) + 2x\mathbb{E}(XY) + \mathbb{E}(Y^2) = \mathbb{E}((xX + Y)^2) \geq 0.$$

Mais ceci n'est possible que si ce trinôme en x n'a au plus qu'une seul racine réelle ; son discriminant

$$4\left((\mathbb{E}(XY))^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2)\right)$$

doit donc être négatif ou nul ce qui donne le résultat.

Le discriminant est nul si et seulement si il admet une racine double x_0 et dans ce cas, $Y(\omega) = -x_0X(\omega)$ pour presque tout ω. ■

Remarque

On déduit de l'inégalité de Cauchy-Schwartz que le coefficient de corrélation de X et Y vérifie

$$-1 \leq \rho(X, Y) \leq 1.$$

Enfin, il peut être intéressant de pouvoir calculer l'espérance d'une fonction d'une variable aléatoire réelle à densité qui est une variable aléatoire en vertu de la proposition.

Proposition

Soit X une variable aléatoire réelle admettant la densité f, et g une fonction continue par morceaux de $\mathbb{R}$ dans $\mathbb{R}$. Alors $g(X)$ est intégrable si et seulement si l'intégrale

$$\int_{\mathbb{R}} |g(x)| f(x) dx,$$

est définie et dans ce cas

$$\mathbb{E}(g(X)) = \int_{R} g(x) f(x) dx.$$

Nous n'avons pas les éléments permettant de démontrer ce résultat, mais l'argument heuristique suivant permet de comprendre pourquoi il est vrai : supposons f et g continues. Posons $X_n = i/n$ si $i/n \leq X < (i + 1/n)$, pour $i \in \mathbb{Z}$. Ainsi, pour tout ω, $X_n(\omega) \to n \to \infty X(\omega)$ ### **TODO** développer dans le cas de l'intégrale HK

Exemples

Nous donnons ici quelques exemples de densités de probabilité. Nous reprenons en particulier les trois exemples de densités donnés au premier cours :

loi uniforme

sur $[a, b]$, où $a < b$ et on note $X \sim \mathcal{U}_{[a,b]}$ si X est de densité

$$\frac{1}{b-a} 1_{[a,b]}(x).$$

Son espérance vaut

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

et puisque

$$\mathbb{E}(X^2) = \int_a^b \frac{x^2}{b-a}dx = \frac{a^2+ab+b^2}{3},$$

alors sa variance vaut

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{(b-a)^2}{12}.$$

loi exponentielle

de paramètre $\theta > 0$ et on note $X \sim \mathcal{E}(\theta)$ si X est de densité

$$\theta e^{-\theta x} 1_{\{x>0\}}.$$

Son espérance et sa variance se calculent aisément (exercice) et valent

$$\mathbb{E}(X) = \frac{1}{\theta} \text{ et } \mathbb{V}(X) = \frac{1}{\theta^2}$$

loi gamma

loi normale

de paramètres μ et σ^2 et on note $X \sim \mathcal{N}(\mu, \sigma^2)$ i X est de densité [4]

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Son espérance et sa variance valent

$$\mathbb{E}(X) = \mu \text{ et } \mathbb{V}(X) = \sigma^2$$

Pour le voir, on fait d'abord le calcul dans le cas centré réduit ($\mu = 0$ et $\sigma^2 = 1$) puis on s'y ramène par le changement de variable $x \mapsto \frac{x-\mu}{\sigma}$.

4. Pour s'assurer qu'il s'agit bien d'une densité, on remarque que $I = \int_{-\infty}^{+\infty} f(x)dx$ vérifie

$$I^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x)f(y)dxdy = \int_0^{2\pi} d\theta \int_0^{+\infty} \frac{1}{2\pi}\rho e^{-\rho^2/2}d\rho$$

(en passant en coordonnées polaires dans l'intégrale double). Le calcul est ensuite aisé et on obtient $I = 1$.

Remarque

Dans les exemples ci-dessus, on peut remarquer que les densités sont paramétrées par des nombres réels (un ou deux dans nos exemples) qui sont liés directement aux valeurs de l'espérance et de la variance de la variable. C'est très important en statistique. En effet, si nous savons a priori que la loi de X appartient à une certaine classe de lois (lois exponentielles, lois normales), nous pourrons trouver laquelle en estimant ses paramètres en fonction des observations de X. Dans le cas de la loi normale, la moyenne et la variance (empiriques) des échantillons fourniront ainsi directement des estimateurs des paramètres.

Il existe des variables aléatoires qui n'ont pas d'espérance, comme le montre l'exemple suivant.

loi de Cauchy

Vecteurs aléatoires à densité

Nous allons généraliser ici la notion de variable aléatoire en considérant qu'elle peut prendre ses valeurs dans $\mathbb{R}^n$. Les vecteurs aléatoires se rencontrent naturellement lorsqu'on s'intéresse à plusieurs quantités conjointement, par exemple dans le cas de la météo, la température, la pluviométrie et la vitesse et la direction du vent.

Définitions

Une variable aléatoire X à valeurs dans $\mathbb{R}^n$ (ou vecteur aléatoire) est simplement une collection de n variables réelles définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, qui sont les *composantes* de X : on écrit $X = (X_1, \ldots, X_n)$.

De même qu'en dimension 1, la loi de X est caractérisée par la fonction de répartition multi-dimensionnelle $F : \mathbb{R}^n \to \mathbb{R}$ définie par

$$F(x_1, \ldots, x_n) = \mathbb{P}_X(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

Mais caractériser les fonctions de répartition sur $\mathbb{R}^n$ est assez délicat, de sorte que cette notion est rarement utilisée. Nous allons plus particulièrement nous intéresser aux vecteurs aléatoires à densité.

Définition

On dit que X admet la densité f si la fonction réelle f sur $\mathbb{R}^n$ est positive, intégrable et vérifie

$$\int_{\mathbb{R}^n} f(x)dx = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n)dx_1 \dots dx_n = 1$$

et si

$$\mathbb{P}_X(X_1 \leq x_1, \dots, X_n \leq x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(x_1, \dots, x_n)dx_1 \dots dx_n$$

De la même manière que dans la proposition, on a :

Proposition

Soit X un vecteur aléatoire de densité f, et soit g une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, continue par morceaux (i.e. continue sauf sur une “bonne” surface de dimension au plus $n-1$). On a alors $g(X) \in \mathcal{L}^1$ si et seulement si

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n)dx_1 \dots dx_n < \infty,$$

et dans ce cas, on a

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n)dx_1 \dots dx_n.$$

Moments d’un vecteur aléatoire

Définition

Si les composantes X_i du vecteur aléatoire $X = (X_1, \dots, X_n)$ sont intégrables, nos pouvons définir le *vecteur espérance*

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))$$

Si les composantes X_i du vecteur aléatoire $X = (X_1, \dots, X_n)$ sont de carré intégrable, la *matrice de covariance* de X est la matrice $C_X = (c_{i,j})_{1 \leq i \leq n, 1 \leq j \leq n}$ de taille $n \times n$ et dont les éléments valent

$$c_{i,j} = \mathrm{Cov}(X_i, X_j)$$

Proposition

La matrice de covariance est symétrique non-négative (ou encore semi-définie positive).

Démonstration La symétrie est évidente. Non-négative signifie que pour tous réels $a_1, \ldots, a_n$, on a $\sum_{i=1}^n \sum_{j=1}^n a_i a_j c_{i,j} \geq 0$. Un calcul simple montre que

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j c_{i,j} = \mathbb{V}(\sum_{i=1}^n a_i X_i).$$

■

Exemple : Vecteur Gaussien n-dimensionel

Un exemple de vecteurs aléatoires est celui des vecteurs gaussiens, que nous étudierons en détail au cours suivant. Soient $m \in R^n$ et C une matrice symétrique définie positive (c'est-à-dire telle que pour tout $x \in \mathbb{R}^n$ non identiquement nul $x^t C x > 0$ où ^{t} désigne la transposée). Le vecteur $X \in R^n$ est un vecteur aléatoire gaussien d'espérance m et de matrice de covariance C si sa densité s'écrit

$$f(x) = \frac{1}{(2\pi^{n/2})\sqrt{\det(C)}} \exp(-\frac{1}{2}(x-m)^t C^{-1})(x-m)$$

On a alors $\mathbb{E}(X) = m$ et $C_X = C$.

TODO figure densité Gaussienne

Variables aléatoires indépendantes

Dans ce paragraphe, on considère un couple (X, Y) de vecteurs aléatoires respectivement à valeurs dans $\mathbb{R}^m$ et $\mathbb{R}^n$. Les résultats s'étendent sans peine à une famille finie quelconque.

On peut se ramener aux évènements pour caractériser l'indépendance de deux variables aléatoires. En effet, considérons le vecteur aléatoire $Z = (X, Y)$, A et B deux ensembles dans $\mathcal{E}_{\mathbb{R}^m}$ et $\mathcal{E}_{\mathbb{R}^n}$. On a vu que les évènements $X \in A$ et $Y \in B$ sont aléatoires si et seulement si $\mathbb{P}_Z(X \in A, Y \in B) = \mathbb{P}(X^{-1}(A) \cap Y^{-1}(B)) = \mathbb{P}(X^{-1}(A))\mathbb{P}(Y^{-1}(B)) = \mathbb{P}_X(X \in A)\mathbb{P}_Y(Y \in B)$. Pour que deux vecteurs aléatoires soient indépendants, on va donc demander que ceci soit valable quelques soient A et B.

Définition

Les vecteurs aléatoires X et Y sont *indépendants* si pour tous ensembles A et B des tribus correspondantes,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

Cette définition se traduit en termes de densités dans la proposition suivante que l'on énonce pour un couple de variables aléatoires pour simplifier

Proposition

Soit X et Y deux variables aléatoires réelles de densités f_x et f_Y. X et Y sont indépendantes si et seulement si le couple $Z = (X, Y)$ a pour densité (sur $\mathbb{R}^2$) :

$$f_Z(x, y) = f_X(x)f_Y(y).$$

Démonstration S'il y a indépendance, la définition implique

$$P(X \leq x, Y \leq y) = P(X \leq x)\mathbb{P}(Y \leq y) = \int_{-\infty}^{x} f_X(u)du \int_{-\infty}^{y} f_Y(v)dv$$

ce qui montre que $\mathbb{P}_Z$ vérifie la définition d'un vecteur aléatoire à densité avec $f_Z = f_X f_Y$.

Inversement, si $f_Z = f_X f_Y$, on a pour tous A, B de $\mathcal{E}_{\mathbb{R}}$

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_X(x)f_Y(y)dxdy = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

■

Considérons maintenant deux fonctions g et h définies sur $\mathbb{R}^m$ et $\mathbb{R}^n$ telles que $g(X)$ et $h(Y)$ soient aussi des variables aléatoires (par exemple continues par morceaux).

Proposition

Avec les notations précédentes, si X et Y sont indépendantes de densités respectives f_X et f_Y, les variables aléatoires $g(X)$ et $h(Y)$ sont aussi indépendantes. Si de plus $g(X)$ et $h(Y)$ sont intégrables, alors le produit $g(X)h(Y)$ est aussi intégrable, et on a

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$$

Démonstration La première assertion est évidente par définition de l'indépendance. Par ailleurs, si $g(X)$ et $h(Y)$ sont intégrables, en notant $f_{(X,Y)}$ la densité du couple (X, Y), et en utilisant le théorème de Fubini, on a

$$\begin{aligned}
\mathbb{E}(g(X)h(Y)) &= \int_{\mathbb{R}^{m+n}} g(x)h(y)f_{(X,Y)}(x,y)dxdy \\
&= \int_{\mathbb{R}^m}\int_{\mathbb{R}^n} g(x)h(y)f_X(x)f_Y(y)dxdy \\
&= \left(\int_{\mathbb{R}^m} g(x)f_X(x)dx\right)\left(\int_{\mathbb{R}^n} h(y)f_Y(y)dy\right) \\
&= \mathbb{E}(g(X))\mathbb{E}(h(Y))
\end{aligned}$$

■

Remarque

Ce résultat est encore valable si X et Y n'admettent pas de densité mais nous ne disposons pas encore des outils de théorie de la mesure nécessaires à sa démonstration.

On déduit de ce résultat et de la définition de la covariance que :

Corollaire

Si les variables aléatoires réelles X et Y sont indépendantes et de carré intégrable, alors $\mathrm{Cov}(X, Y) = 0$ et $\rho(X, Y) = 0$.

Remarque

Attention, la réciproque est fausse. Par exemple, si $X \sim \mathcal{U}_{[-1,1]}$ et $Y = X^2$. X et Y ne sont clairement pas indépendante mais on a

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,X^2) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2) = 0$$

Identification de densité

Un problème important est le suivant. Soit X une variable aléatoire réelle, admettant la densité f_X. Soit g une fonction continue par morceaux, de sorte que $Y = g(X)$ soit aussi une variable aléatoire. Est-ce que Y admet une densité, et si oui, comment la calculer ?

On peut déjà remarquer que cette densité n'existe pas toujours. Si par exemple $g(x) = a$ pour tout x, la loi de Y est la masse de Dirac en a, qui n'a pas de densité.

Pour résoudre ce problème, l'idée consiste à essayer de mettre $E(h(Y)) = E(h \circ g(X))$ sous la forme $\int h(y) f_Y(y) dy$ pour une fonction convenable f_Y, et une classe de fonctions h suffisamment grande. La fonction f_Y sera alors la densité cherchée.

La proposition implique

$$\mathbb{E}(h(Y)) = \mathbb{E}(h \circ g(X)) = \int_{\mathbb{R}} h \circ g(x) f_X(x) dx$$

et on fait le changement de variable $y = g(x)$ dans cette intégrale. Cela nécessite que g soit dérivable et bijective "par morceaux", et il faut faire très attention aux domaines où g est croissante ou décroissante. Puisque la fonction h est arbitraire appelle couramment cette technique la *méthode de la fonction muette*. Cette approche résulte en fait de la proposition suivante que nous ne démontrerons pas :

Proposition

Si il existe une fonction f telle que pour toute fonction continue bornée h,

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x) f(x) dx$$

alors la loi de X admet la densité f.

L'idée de la preuve repose sur le fait que les fonctions continues bornées peuvent approcher une fonction $h = 1_{]-\infty,y]}$, pour laquelle la formule précédente donne la fonction de répartition de f.

Nous donnons ici quelques exemples d'application de cette méthode :

— Soit $Y = aX + b$, où a et b sont des constantes. Si $a = 0$, alors $Y = b$ et la loi de Y est la masse de Dirac en b (sans densité). Si $a \neq 0$, on fait le changement de variable $y = ax + b$, ce qui donne

$$\mathbb{E}(h(Y)) = \int_{\mathbb{R}} h(ax+b) f_X(x) dx = \int_{\mathbb{R}} h(y) f_X(\frac{y-b}{a}) \frac{1}{|a|} dy$$

Donc

$$f_Y(y) = f_X(\frac{y-b}{a}) \frac{1}{|a|}$$

— Soit $Y = X^2$. La fonction g est décroissante sur $\mathbb{R}_-$ et croissante sur $\mathbb{R}_+$. Le changement de variable $y = x^2$ donne alors

$$\mathbb{E}(h(Y)) = \int_{-\infty}^{0} h(x^2) f_X(x) dx + \int_{-\infty}^{0} h(x^2) f_X(x) dx$$

Dans le cas des vecteurs aléatoires, l'idée est la même. Soit $X = (X_1, \dots, X_n)$, un vecteur aléatoire de densité f_X sur $\mathbb{R}^n$, g une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^m$ et $Y = g(X)$. Plusieurs cas sont à considérer :

1. $m > n$, le vecteur Y n'admet pas de densité
2. $m = n$, on utilise comme dans le cas unidimensionel le changement de variable $y = g(x)$ dans

$$\mathbb{E}(h(Y)) = \mathbb{E}(h \circ g(X)) = \int_{\mathbb{R}^n} h \circ g(x) f_X(x) dx$$

   Supposons d'abord que g soit une bijection continûment différentiable de A dans B, ouverts de $\mathbb{R}^n$. Le théorème de changement de variable nous assure :

$$\int_A h \circ g f_X(x) dx = \int_B h(y) f_X \circ g^{-1}(y) \frac{1}{|\det Dg(y)|}, dy$$

   où Dg désigne la matrice de Jacobi associée à la différentielle de g. Dans le cas où $f_X(x) = 0$ en dehors de A, on obtient que Y admet la densité

$$f_Y(y) = 1_B(y) f_X \circ g^{-1}(y) \frac{1}{|\det Dg(y)|}.$$

   Lorsque g est simplement continûment différentiable, il existe souvent une partition finie $(A_i)_{1 \leq i \leq n}$ de l'ensemble $\{x; f(x) > 0\}$, telle que g soit injective sur chaque A_i. On note alors $B_i = g(A_i)$ l'image de A_i par g. On découpe alors l'intégrale selon les A_i, on applique la formule précédente à chaque morceau et on somme pour obtenir :

$$f_Y(y) = \sum_{i=1}^{n} 1_{B_i}(y) f_X \circ g^{-1}(y) \frac{1}{|\det Dg(y)|},$$

   où g^{-1} est bien définie sur chaque B_i comme image réciproque de la restriction de g à A_i.
3. $m < n$, on commence par "compléter" Y, en essayant de construire une application g' de $\mathbb{R}^n$ dans $\mathbb{R}^n$ dont les m premières composantes coïncident avec les composantes de g et pour laquelle on peut appliquer l'une des deux formules précédentes. On obtient ainsi la densité f'_Y de $Y' = g'(X)$ puis on obtient la densité de Y en calculant sa *loi marginale* :

$$f_Y(y_1, \dots, y_m) = \int_{\mathbb{R}^{n-m}} f_{Y'}(y_1, \dots, y_m, y_{m+1}, \dots y_n) dy_{m+1} \dots dy_n$$

TODO Exemples ou exercices ?

1. **Coordonnées polaires**
2. **Loi Beta**
3. **Somme de deux variables aléatoires indépendantes**

Exercices

Propriété de non vieillissement de la loi exponentielle

Références