Corrélation

Nous avons déjà défini l'indépendance des v.a. Il reste alors à définir une mesure de dépendance entre les v.a. Étant données deux v.a. X et Y, sont-elles positivement ou négativement liées, lorsqu'elles ne sont pas indépendantes ? Pour formuler mathématiquement ces questions, nous introduisons dans la suite les notions de covariance et de coefficient de corrélation. Elles auront pour but de fournir une mesure quantitative de "liaison" éventuelle entre les v.a. Ces définitions et les propriétés sous-jacentes sont vraies aussi bien dans le cas des espaces discrets que continus.

Pour étudier les conséquences de ces définitions, nous aurons besoin de l'inégalité suivante :

Proposition (Inégalité de Cauchy-Schwarz). Soient X et Y deux v.a. admettant chacune une variance. Nous avons :

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

Définitions. Soient X et Y deux v.a. admettant chacune une variance. La *covariance entre X et Y* est définie par :

$$cov(X,Y) \ = \ \mathbb{E}\,[(X-\mathbb{E}X)(Y-\mathbb{E}Y)].$$

Si les variances de X et de Y sont non nulles, on définit le *coefficient de corrélation entre X et Y* par :

$$corr(X,Y) \ = \ \frac{cov(X,Y)}{\sqrt{\mathbb{V}X \ \mathbb{V}Y}}.$$

L'inégalité de Cauchy-Schwarz implique que ce coefficient est situé dans l'intervalle $[-1,+1]$.

La covariance et le coefficient de corrélation mesurent en quelque sorte le lien entre les v.a. Si $corr(X,Y)$ est positif les deux v.a. sont positivement corrélées et, donc, l'augmentation de l'une "favorise" l'augmentation en moyenne de l'autre. S'il est négatif, la situation s'inverse.

On démontre aussi facilement :

Proposition. Si X et Y sont deux v.a. indépendantes admettant chacune une espérance, alors :

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

Si, de plus, elles admettent chacune une variance, alors :

$$corr(X,Y) = 0.$$

La réciproque est fausse :

Exemple. Soit X une v.a. prenant une des valeurs $-1, 0, 1$ avec la même probabilité $\frac{1}{3}$. Posons $Y = X^2$. Il est évident que X et Y sont fortement liées et en particulier :

$$Pr(X=1, Y=0) \ = \ 0 \neq Pr(X=1)Pr(Y=0).$$

Néanmoins, on a $cov(X,Y) = corr(X,Y) = 0$.

Opérations Simples sur les Variables Aléatoires

Soit X une v.a. réelle et soit h une application de $\mathbb{R}$ dans $\mathbb{R}$ *mesurable* (i.e. pour tout borélien B, $h^{-1}(B) \in \mathcal{B}$). Alors $Y = h(X)$ est une nouvelle v.a. On peut calculer sa loi de probabilité. Dans certains cas, il y a des formules simples qui donnent la f.r. (ou la densité de Y, si celle-ci existe) à partir de celle de X.

Changement de Variable

Soit la F la f.r. de la v.a. X. On suppose que l'intervalle I est l'intervalle de variations de F (i.e. F est strictement croissante sur I et constante en dehors de I). Soit h une application réelle qui est strictement croissante et continue sur I. Nous avons alors :

Proposition. Soit $Y = h(X)$, où X, h satisfont les conditions ci-dessus. Alors la f.r. G de Y vaut :

$$G(y) = F(h^{-1}(y)).$$

Corollaire. Si U est une v.a. uniforme tirée dans $[0,1]$ et la f.r. F satisfait la condition ci-dessus, alors la v.a. $X = F^{-1}(U)$ admet la f.r. F.

Une application en simulation. Pour simuler des événements qui arrivent de façon aléatoire, nous avons besoin de générer des v.a. suivant une distribution donnée. Il arrive souvent qu'une source de génération uniforme soit disponible. Ce corollaire nous permet de construire une v.a. X à partir d'une v.a. uniformément tirée dans $[0,1]$.

Sous certaines hypothèses supplémentaires, on pourra calculer directement la densité de $Y = h(X)$:

Proposition. Reconsidérons les hypothèses de la proposition précédente et supposons que, de plus, X admette une densité $f(x)$ sur I et que h soit continûment dérivable sur I. Alors $Y = h(X)$ admet la densité :

$$g(y) = \left(h^{-1}\right)'(y) f\left(h^{-1}(y)\right).$$

Exemple 1. Soit X une v.a. uniforme sur $[0,1]$. Trouver la densité de $Y = Arcsin(X)$.

Solution. On vérifie facilement que l'on est dans les conditions de l'application de la proposition précédente, pour $I = [0,1]$.

On démontre alors que Y admet la densité :

$$g(y) = cos(y), \quad y \in [0, \pi/2].$$

Exemple 2. Disposant d'une source de génération uniforme dans l'intervalle $[0,1]$, comment peut-on générer des v.a. exponentielles de paramètre $\lambda > 0$?

Solution. Soit U une v.a. uniforme sur $[0,1]$. Le corollaire précédent confirme que $X = \frac{\ln(1-U)}{-\lambda}$ est une v.a. de f.r.

$$F(x) = Pr(X \leq x) = 1 - e^{-\lambda x}, \quad \forall x \geq 0.$$

Somme des Variables Aléatoires Indépendantes

Nous avons vu que la f.c. d'une somme de v.a. indépendantes est le produit des f.c. des v.a. Cette propriété permet de transformer le calcul sur la somme en un calcul sur les f.c. A titre d'exemple, nous allons démontrer que la somme de deux v.a. normales indépendantes est encore une v.a. normale.

Théorème. Soit X une v.a. normale de type $\mathcal{N}(m_1, \sigma_1)$ et Y une autre v.a. normale indépendante de X de type $\mathcal{N}(m_2, \sigma_2)$. Alors la v.a. $Z = X + Y$ est une v.a. normale de type :

$$\mathcal{N}\left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right).$$

Il arrive toutefois que le calcul des f.c. ou l'inversion de la transformation soit difficile et qu'un calcul direct sur les f.r. (ou sur les densités) soit plus simple. La proposition suivante peut dans certains cas simplifier le calcul de la distribution de la somme de deux v.a. indépendantes.

Proposition. Soient X et Y deux v.a. indépendantes admettant les densités f_X et f_Y respectivement. Alors la v.a. $Z = X + Y$ admet la densité :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx.$$

Exemple. Trouver la densité de la somme de deux v.a. indépendantes et uniformément tirées dans $[0, 1]$.

Solution. Si les v.a. X et Y sont uniformes dans l'intervalle unitaire, elles ont toutes deux la densité $\mathbf{1}_{[0,1]}$. Etant par ailleurs indépendantes, la proposition ci-dessus s'applique à la leur somme Z qui sera donc de densité :

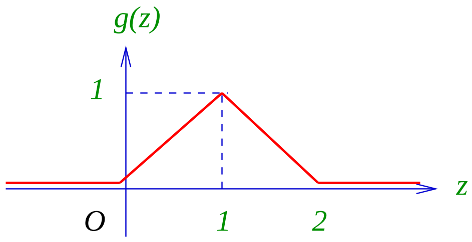
$$g(z) = \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(x) \mathbf{1}_{[0,1]}(z - x) dx.$$

L'expression à intégrer vaut 1 si l'on a simultanément $0 \le x \le 1$ et $0 \le z - x \le 1$; 0 sinon.

Nous sommes ramenés à considérer deux cas pour z :

- Si $0 \le z \le 1$, ces deux conditions sont satisfaites, ssi $0 \le x \le z$, et $g(z)$ sera égale à $\int_0^z dx = z$.
- Si $1 \le z \le 2$, ces deux conditions sont satisfaites, ssi $z - 1 \le x \le 1$, et $g(z)$ sera égale à $\int_{z-1}^1 dx = 2 - z$.

Densité de la somme de deux v.a. uniformes et indépendantes



Quelques Lois Utiles en Statistique

Nous présentons dans la suite deux lois qui interviennent couramment en tests statistiques.

Loi du Khi-deux

Définition. La somme des carrés $\sum_{i=1}^{k} X_i^2$ de k v.a. normales indépendantes, centrées-réduites est une v.a. qui suit *une loi du* χ_2 *à* k *degrés de liberté*. Cette v.a. est désignée par χ_k^2.
Dans le cas où $k = 1$, la v.a. suit une loi du χ_2 à 1 degré de liberté et elle est le carré d'une v.a. de type $\mathcal{N}(0,1)$. Elle admet la densité :

$$f(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}, \quad x \geq 0.$$

Nous admettons la proposition suivante sur la loi du χ_2 :

Proposition. Soit une v.a. χ_k^2. Alors :

- La densité de χ_k^2 vaut :

$$f(x) \; = \; \frac{1}{2^{k/2}\Gamma(k/2)} e^{-x/2} x^{k/2-1}, \quad x \geq 0.$$

- Son f.c. est donnée par :

$$\varphi(t) \; = \; \mathbb{E}(e^{it\chi_k^2}) \; = \; \frac{1}{(1-2it)^{k/2}}.$$

- Son espérance vaut k et sa variance $2k$.

Loi de Student

Définition. On appelle *loi de Student à* k *degré de liberté* la loi du quotient $\frac{X}{\sqrt{Z_k/k}}$, où X et Z_k sont deux v.a. indépendantes, la première suivant une loi normale centrée-réduite et la seconde une loi du χ_2 à k degrés de liberté. Cette loi est notée par t_k.
Nous admettons la proposition suivante à propos de la loi de Student :

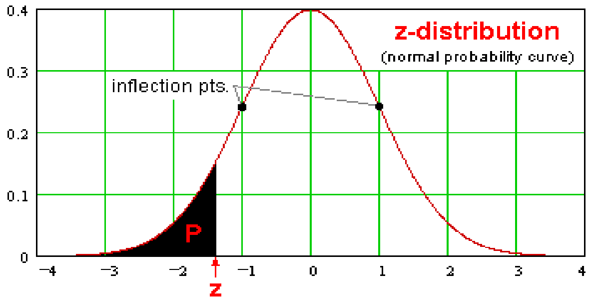
Proposition. Pour une loi t_k de Student, nous avons :

- La densité de la loi t_k vaut :

$$f(x) \; = \; \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma(\frac{k}{2})\sqrt{k\pi}} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}.$$

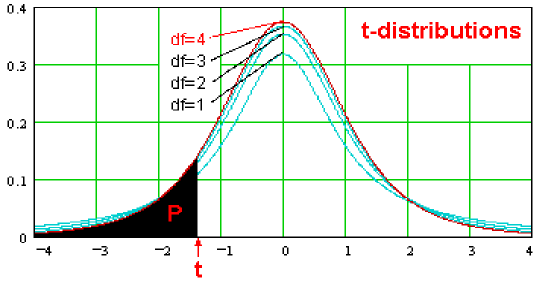
- Son espérance est 0 et sa variance vaut $\frac{k}{k-2}$.

Loi Normale Centrée-Réduite



$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

Loi de Student à n Degrés de Liberté



$$f(t) = \frac{\Gamma(\frac{n+1}{2})(1+\frac{x^2}{n})^{-\frac{n}{2}-\frac{1}{2}}}{\Gamma(\frac{n}{2})\sqrt{\pi n}}$$

Loi du Khi-2 à n Degrés de Liberté

$$f(\chi^2) = (\chi^2)^{\frac{n}{2}-1}\,\mathrm{exp}(-\frac{1}{2}\chi^2)2^{-\frac{n}{2}}\frac{1}{\Gamma(\frac{n}{2})}$$