

2

Variables aléatoires

Dans ce chapitre, on étudiera uniquement les variables aléatoires réelles. Les variables qualitatives ou ordinales (à valeurs dans un ensemble quelconque ou muni d'une structure d'ordre) ne feront pas l'objet d'une étude théorique ; on les trouvera évoquées dans les chapitres consacrés à la statistique.

2.1 LOI DE PROBABILITÉ ET MOMENTS D'UNE VARIABLE ALÉATOIRE RÉELLE

2.1.1 Définition et fonction de répartition

2.1.1.1 Généralités

Le concept de variable aléatoire formalise la notion de grandeur variant selon le résultat d'une expérience aléatoire.

Considérons le lancer de deux dés parfaitement équilibrés : cette expérience se traduit par l'ensemble Ω de tous les couples de chiffres de 1 à 6 :

$$\Omega = \{(1, 1) ; (1, 2) ; \dots ; (6, 6)\}$$

muni de la loi de probabilité P telle que $P(\omega) = \frac{1}{36}$, $\forall \omega \in \Omega$.

Intéressons-nous à la somme des points marqués par les deux dés. On définit ainsi une application S de Ω dans l'ensemble $E = \{2, 3, \dots, 12\}$ (fig. 2.1).

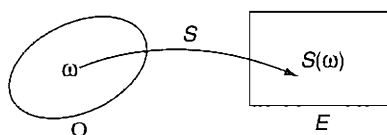


FIGURE 2.1

Pour obtenir la probabilité d'une valeur quelconque de S , il suffit de dénombrer les ω qui réalisent cette valeur. Ainsi :

$$P(S = 5) = P(\{(1, 4)(2, 3)(3, 2)(4, 1)\}) = \frac{4}{36}$$

et généralement $P(S = s) = P(\{S^{-1}(s)\})$.

On voit que, pour définir la loi de probabilité sur S , on transporte la loi de probabilité de Ω sur E par l'application S .

Si X est une application d'un ensemble probabilisé (Ω, \mathcal{C}, P) dans E , il faut donc que E soit probabilisable, c'est-à-dire muni d'un tribu \mathcal{T} et que l'image réciproque de tout élément de \mathcal{T} soit un événement, c'est-à-dire un élément de \mathcal{C} . On reconnaît ici la définition mathématique de la mesurabilité d'une fonction.

Une variable aléatoire X est donc une application mesurable de (Ω, \mathcal{C}, P) dans (E, \mathcal{T}) .

Lorsque $E = \mathbb{R}$, on utilise comme tribu la σ -algèbre engendrée par les intervalles de \mathbb{R} ; c'est la plus petite σ -algèbre (autrement dit l'intersection de toutes les σ -algèbres) contenant les intervalles. Cette tribu est appelée tribu borélienne et est notée \mathcal{B} .

DÉFINITION 1

L Une variable aléatoire réelle est une application mesurable de (Ω, \mathcal{C}, P) dans \mathbb{R} muni de sa tribu borélienne $(\mathbb{R}, \mathcal{B})$.

Pour tout borélien B , on définit $P_X(B)$ par :

$$\begin{aligned} P_X(B) &= P(\{\omega | X(\omega) \in B\}) \\ &= P([X^{-1}(B)]) \end{aligned}$$

ceci définit une probabilité sur $(\mathbb{R}, \mathcal{B})$ d'où la :

DÉFINITION 2

L On appelle loi de probabilité de X la mesure image de P par X et on la note P_X .

Pour une variable discrète, c'est-à-dire une variable ne pouvant prendre qu'un nombre fini (ou dénombrable) de valeurs x_1, x_2, \dots, x_n , la loi P_X est constituée de masses ponctuelles. P_X peut alors être représentée par un diagramme en bâtons.

Ainsi, pour l'exemple du lancer de deux dés, on a la figure 2.2.

2.1.1.2 Fonction de répartition

La fonction de répartition d'une variable aléatoire X est l'application F de \mathbb{R} dans $[0, 1]$ définie par :

$$F(x) = P(X < x)$$

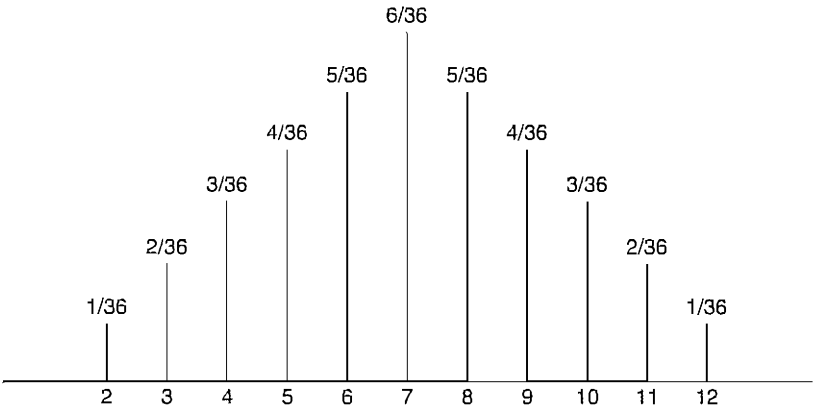


FIGURE 2.2

PROPRIÉTÉS (sans démonstration)

L F est une fonction monotone croissante continue à gauche. En tant que fonction monotone, elle admet un nombre de points de discontinuité au plus dénombrable. Réciproquement, toute fonction monotone croissante continue à gauche telle que $F(-\infty) = 0$ et $F(+\infty) = 1$ définit une loi de probabilité unique sur \mathbb{R} .

Un exemple de fonction de répartition correspondant à une variable discrète (celle de S définie précédemment) est donné par la figure 2.3.

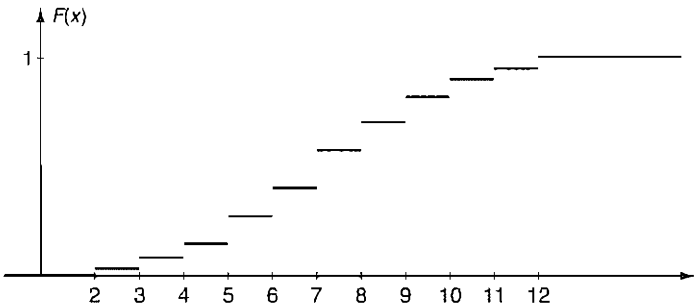


FIGURE 2.3

La figure 2.4 est un exemple de fonction de répartition correspondant à une variable continue (voir plus loin).

L'importance pratique de la fonction de répartition est qu'elle permet de calculer la probabilité de tout intervalle de \mathbb{R} :

$$P(a \leq X < b) = F(b) - F(a)$$

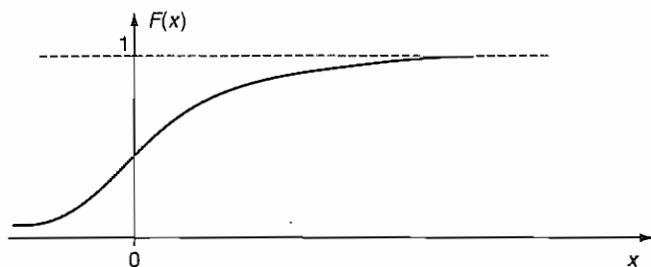


FIGURE 2.4

2.1.1.3 Variables continues

La notion de variable continue, ou plus exactement absolument continue, se confond avec celle de variable admettant une densité de probabilité.

DÉFINITION

Une loi de probabilité P_X admet une densité f si, pour tout intervalle I de \mathbb{R} , on a :

$$P_X(I) = \int_I f(x) \, dx = \int_{\mathbb{R}} \mathbb{1}_I(x) f(x) \, dx$$

($\mathbb{1}_I$ est la fonction indicatrice de I).

F est alors dérivable et admet f pour dérivée. On a donc :

$$P(a < X < b) = \int_a^b f(x) \, dx = F(b) - F(a) \quad (\text{fig. 2.5})$$

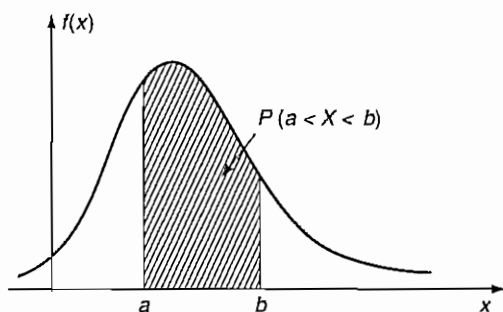


FIGURE 2.5

Une densité f est donc une fonction positive d'intégrale égale à 1 :

$$\int_{\mathbb{R}} f(x) \, dx = 1$$

On remarque que pour une variable à densité :

$$P(X = x) = 0 \quad \forall x$$

et on peut écrire :

$$P(x < X < x + dx) = f(x) dx$$

Exemple : La variable X , dont la loi est définie par $P(X > x) = \exp(-\lambda x)$ pour tout x positif, admet pour densité :

$$\begin{aligned} f(x) &= \lambda \exp(-\lambda x) & \text{si } x \geq 0 \\ f(x) &= 0 & \text{si } x < 0 \end{aligned} \quad (\text{fig. 2.6})$$

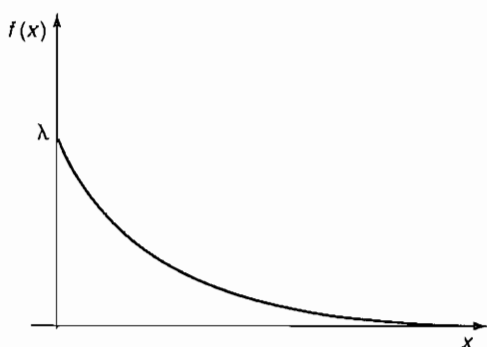


FIGURE 2.6

Elle est utilisée couramment pour représenter la durée de vie de phénomènes sans vieillissement (comme les composants électroniques).

2.1.1.4 Taux instantané de défaillance

Si X est une variable continue positive représentant une durée, on définit la fonction suivante :

$$h(x) = \frac{f(x)}{1 - F(x)}$$

appelées selon les domaines d'application : « taux instantané de défaillance », « fonction de hasard » ou encore « quotient de mortalités ». Pour une durée de vie X , $h(x)$ s'interprète comme la probabilité de décès immédiatement après x , sachant que l'on a vécu jusqu'à x .

En effet, pour dx infiniment petit :

$$P(x < X < x + dx / X > x) = \frac{f(x) dx}{1 - F(x)} = h(x) dx.$$

$1 - F(x)$ est appelée fonction de survie.

$h(x)$ caractérise la loi de X car on peut retrouver $F(x)$ à partir de $h(x)$:

$$h(x) = -\frac{d}{dx} \ln(1 - F(x))$$

$$F(x) = 1 - \exp\left(-\int_0^x h(t)dt\right)$$

Une fonction $h(x)$ croissante est caractéristique d'un phénomène de vieillissement.

Si $h(x) = c$, il y a absence de vieillissement, le décès est dû à des causes aléatoires externes : X suit alors la loi exponentielle $F(x) = 1 - \exp(-cx)$, qui sera étudiée plus loin.

2.1.2 Loi d'une fonction d'une variable aléatoire $Y = \varphi(X)$

On suppose X continue avec une densité f et une fonction de répartition F . φ sera supposé dérivable. On recherche g et G densité et fonction de répartition de Y .

2.1.2.1 φ bijective

φ est donc monotone. Si φ est croissante, on a $F(x) = G(\varphi(x))$ car $X < x \Leftrightarrow Y < \varphi(x)$ d'où :

$$\boxed{G(y) = F(\varphi^{-1}(y))} \quad (\text{fig. 2.7a})$$

En dérivant : $f(x) = g(\varphi(x))\varphi'(x)$ soit :

$$g(y) = \frac{f(x)}{\varphi'(x)}$$

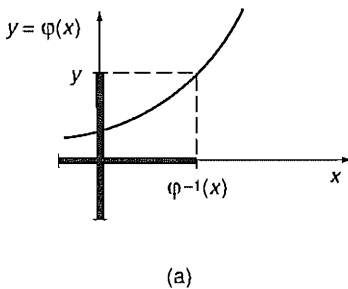


FIGURE 2.7a

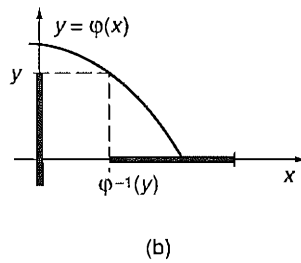


FIGURE 2.7b

ou encore :

$$g(y) = \frac{f[\varphi^{-1}(y)]}{\varphi'[\varphi^{-1}(y)]}$$

Si φ est décroissante $X < x \Leftrightarrow Y > \varphi(x)$, d'où :

$$\boxed{G(y) = 1 - F(\varphi^{-1}(y))} \quad (\text{fig. 2.7b})$$

et en dérivant :

$$g(y) = -\frac{f(x)}{\varphi'(x)}$$

Puisque φ est décroissante, $\varphi' < 0$, et on a la formule générale pour une application bijective φ quelconque :

$$g(y) = \frac{f(x)}{|\varphi'(x)|}$$

$$g(y) = \frac{f[\varphi^{-1}(y)]}{|\varphi'[\varphi^{-1}(y)]|}$$

Exemple :

$$Y = \exp(X) \quad \text{et} \quad X = \ln Y$$

$$g(y) = \frac{f(x)}{\exp(x)} = \frac{f(\ln y)}{y}$$

2.1.2.2 φ quelconque

Le principe consiste toujours à identifier la fonction de répartition $G(y)$ en recherchant l'antécédent pour X de l'événement $Y < y = \varphi(x)$.

Par exemple, si $Y = X^2$ avec X défini sur \mathbb{R} : $P(Y < y) = P(-\sqrt{y} < X < +\sqrt{y})$:

$$G(y) = F(\sqrt{y}) - F(-\sqrt{y})$$

$$g(y) = f(\sqrt{y}) \frac{1}{2\sqrt{y}} + f(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

$$g(y) = \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y}))$$

en particulier $g(y) = \frac{f(\sqrt{y})}{\sqrt{y}}$ si f est une fonction paire.

2.1.3 Indépendance de deux variables aléatoires

Soient X et Y deux variables aléatoires réelles définies sur le même espace probabilisé. Le couple (X, Y) est donc une application mesurable de (Ω, \mathcal{C}, P) dans \mathbb{R}^2 muni de sa tribu borélienne.

DÉFINITION

L X et Y sont indépendantes si, pour tout couple de boréliens B_i et B_j , on a :

$$P((X \in B_i) \cap (Y \in B_j)) = P(X \in B_i)P(Y \in B_j)$$

En d'autres termes, la loi de probabilité P_{XY} du couple (X, Y) n'est autre que la loi produit que l'on note :

$$P_{XY} = P_X \otimes P_Y$$

COROLLAIRE

X et Y sont indépendantes si et seulement si la fonction de répartition du couple (X, Y) définie par $H(x, y) = P(X < x \cap Y < y)$ est égale au produit des fonctions de répartition respectives de X et de Y , appelées fonctions de répartition marginales :

$$H(x, y) = F(x)G(y)$$

Si X et Y admettent des densités $f(x)$ et $g(y)$, alors le couple (X, Y) admet pour densité $f(x)g(y)$. Dans ce cas, la réciproque est également vraie.

2.1.4 Moments d'une variable aléatoire

Une loi de probabilité peut être caractérisée par certaines valeurs typiques associées aux notions de valeur centrale, de dispersion et de forme de la distribution.

2.1.4.1 L'espérance mathématique

Pour une variable discrète, on définit l'espérance $E(X)$ par la formule :

$$E(X) = \sum_i x_i P(X = x_i)$$

(si cette expression a un sens). $E(X)$ est la moyenne arithmétique des différentes valeurs de X pondérées par leurs probabilités.

Pour une variable continue admettant une densité, $E(X)$ est la valeur, si l'intégrale converge, de $\int_{\mathbb{R}} x f(x) dx$.

Ces deux expressions ne sont en fait que des cas particuliers de la définition générale suivante :

DÉFINITION

X étant une variable aléatoire réelle définie sur (Ω, \mathcal{C}, P) , l'espérance mathématique de X est, si elle existe, l'intégrale de X par rapport à la mesure P :

$$E(X) = \int_{\Omega} X dP$$

D'après le théorème de la mesure image, on a :

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

d'où, en particulier si P_X est absolument continue par rapport à la mesure de Lebesgue de \mathbb{R} , il existe une densité $f(x) : dP_X(x) = f(x) dx$ et alors on retrouve :

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

Il faut prendre garde au fait que l'espérance mathématique n'existe pas toujours. Ainsi, la variable X ayant pour densité sur \mathbb{R} :

$$f(x) = \frac{1}{\pi(1+x^2)} \quad (\text{loi de Cauchy})$$

n'a pas d'espérance car l'intégrale $\int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx$ diverge.

Les propriétés élémentaires de l'espérance mathématique sont celles des intégrales et se déduisent de la linéarité. Si a est une constante :

$E(a)$	$= a$
$E(aX)$	$= aE(X)$
$E(X + a)$	$= E(X) + a$

La plus importante propriété est l'additivité : l'espérance d'une somme de variables aléatoires (qu'elles soient ou non indépendantes) est égale à la somme de leurs espérances :

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

A. Espérance d'une fonction $\varphi(X)$ d'une variable aléatoire

Par définition, $E[\varphi(X)] = \int_{\Omega} (\varphi \circ X) dP$ si cette expression a un sens.

En utilisant à nouveau le théorème de la mesure image, on a :

$$E(\varphi(X)) = \int_{\mathbb{R}} \varphi(x) dP_X(x)$$

Ce résultat très important est d'un emploi courant et permet de calculer l'espérance d'une variable $\varphi(X)$ sans avoir à déterminer la loi de $\varphi \circ X$.

B. Inégalité de Jensen

Si φ est une fonction convexe, on peut montrer, si les espérances existent, que :

$$E(\varphi(X)) \geq \varphi(E(X))$$

On en déduit en particulier :

$$\begin{aligned} E(|X|) &\geq |E(X)| \\ E(X^2) &\geq (E(X))^2 \\ E(\exp(X)) &\geq \exp(E(X)) \end{aligned}$$

C. Espérance d'un produit

Si X et Y sont deux variables aléatoires de loi conjointe P_{XY} , on a, si l'expression a un sens :

$$E(XY) = \int_{\mathbb{R}^2} xy \, dP_{XY}(x, y)$$

Lorsque X et Y sont indépendants, $dP_{XY}(x, y) = dP_X(x) \otimes dP_Y(y)$ et l'intégrale double se factorise :

$$E(XY) = \int_{\mathbb{R}} x \, dP_X(x) \int_{\mathbb{R}} y \, dP_Y(y)$$

d'où :

$$\boxed{X \text{ et } Y \text{ indépendants} \Rightarrow E(XY) = E(X)E(Y)}$$

Attention : La réciproque est fautive et $E(X)E(Y) = E(XY)$ n'entraîne pas en général l'indépendance de X et Y .

D. Une interprétation statistique

Reprenons l'exemple du lancer de deux dés. Par raison de symétrie, $E(S) = 7$. Supposons qu'on lance n fois les deux dés et que les réalisations successives de S soient s_1, s_2, \dots, s_n .

Formons la moyenne $\bar{s} = \frac{1}{n} \sum s_i$ de ces résultats.

On montre alors que si $n \rightarrow \infty$, $\bar{s} \rightarrow 7$ en un sens qui sera précisé plus tard (loi des grands nombres, voir paragr. 2.7 et chapitre 12).

E. Espérance et fonction de répartition

Sous réserve de convergence de l'intégrale, on a pour une variable **positive** le résultat suivant :

$$E(X) = \int_0^{\infty} (1 - F(x)) \, dx$$

En effet, en intégrant par parties : $\int_0^{\infty} (1 - F(x)) \, dx = [(1 - F(x))x]_0^{\infty} + \int_0^{\infty} x f(x) \, dx$, et le crochet est nul si l'intégrale converge.

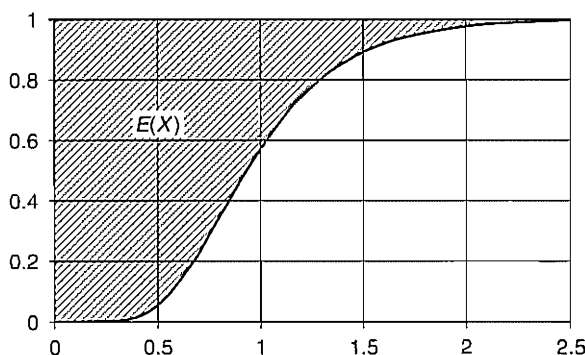


FIGURE 2.8

L'espérance d'une variable positive s'interprète donc comme l'aire située entre l'horizontale $y = 1$ et la fonction de répartition. La figure 2.8 correspond à la fonction de répartition d'une loi log-normale d'espérance 1 et d'écart-type 0.4.

2.1.4.2 La variance

On appelle variance de X notée $V(X)$ ou σ^2 la quantité définie par :

$$\sigma^2 = E((X - m)^2) = \int_{\mathbb{R}} (x - m)^2 dP_X(x)$$

où $m = E(X)$.

σ s'appelle l'*écart-type* de X .

La variance est donc le moment centré d'ordre 2 de la distribution et est une mesure de la dispersion de X autour de m .

• Propriétés de la variance

Comme $E((X - a)^2) = V(X) + (E(X) - a)^2$ (formule de König-Huyghens) on en déduit que $V(X)$ est la valeur minimale de $E((X - a)^2)$ quand a varie.

On en déduit la formule classique

$$V(X) = E(X^2) - (E(X))^2$$

Par ailleurs :

$$V(X - a) = V(X)$$

$$V(aX) = a^2 V(X) \quad \text{et} \quad \sigma(aX) = |a| \sigma(X)$$

$$V(X) = 0 \Leftrightarrow X = a \text{ (presque sûrement)}$$

L'espérance et l'écart-type sont reliés par l'*inégalité de Bienaymé-Tchebyshev* :

$$P(|X - E(X)| > k\sigma) \leq \frac{1}{k^2}$$

■ Démonstration

$$\sigma^2 = \int_{\mathbb{R}} (x - m)^2 dP_X(x) > \int_{|X - m| > k\sigma} (x - m)^2 dP_X(x)$$

car on restreint le domaine d'intégration d'une fonction positive. En minorant $(x - m)^2$ par $k^2\sigma^2$, on a :

$$\int_{|X - m| > k\sigma} (x - m)^2 dP_X(x) > k^2\sigma^2 \int_{|X - m| > k\sigma} dP_X(x)$$

Cette dernière intégrale vaut $P(|X - m| > k\sigma)$, ce qui établit la propriété.

Cette inégalité, dont l'intérêt théorique vient de ce qu'elle est valable quelle que soit la loi de X , n'a que peu d'applications pratiques, car la majoration qu'elle fournit est la plupart du temps excessive. Ainsi pour une loi normale, $P(|X - E(X)| > 2\sigma) \approx 0.05$ alors que l'inégalité de Bienaymé-Tchebyshev donne 0.25 comme majorant. Remarquons, de plus, que l'inégalité est inutilisable pour $k \leq 1$.

- Variance d'une somme de variables aléatoires

$$\begin{aligned} V(X + Y) &= E[(X + Y)^2] - (E(X) + E(Y))^2 \\ &= E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)^2 - 2E(X)E(Y) \\ &= V(X) + V(Y) + 2(E(XY) - E(X)E(Y)) \end{aligned}$$

On appelle covariance de X et Y la quantité :

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = E((X - E(X))(Y - E(Y)))$$

donc :

$$V(X + Y) = V(X) + V(Y) + 2 \text{cov}(X, Y)$$

En particulier :

$$\begin{array}{l} X \text{ et } Y \\ \text{indépendantes} \end{array} \Rightarrow V(X + Y) = V(X) + V(Y)$$

mais la réciproque est ici encore inexacte en général.

- Variance d'un produit de deux variables indépendantes

Un calcul élémentaire montre que :

$$V(XY) = V(X)V(Y) + V(X)(E(Y))^2 + V(Y)(E(X))^2$$

- Approximations de l'espérance et de la variance d'une fonction $\varphi(X)$

Un développement limité à l'ordre 2 au voisinage de l'espérance m de X donne :

$$\varphi(x) - \varphi(m) \approx (x - m)\varphi'(m) + \frac{(x - m)^2}{2} \varphi''(m)$$

En prenant l'espérance :

$$E(\varphi(X)) - \varphi(m) = E\left(\frac{(X - m)^2}{2}\right) \varphi''(m)$$

soit :

$$E(\varphi(X)) \approx \varphi(m) + \frac{1}{2} V(X) \varphi''(m)$$

En élevant au carré $\varphi(X) - \varphi(m)$ et en prenant l'espérance, on trouve également [Lejeune, 2004] :

$$V(\varphi(X)) \approx (\varphi'(m))^2 V(X)$$

2.1.4.3 Autres moments

On définit, si ils existent, les moments centrés d'ordre k :

$$\mu_k = E[(X - m)^k]$$

On a évidemment $\mu_1 = 0$ et $\mu_2 = V(X)$. Si la distribution de la variable aléatoire est symétrique, on a $\mu_{2k+1} = 0 \quad \forall k$.

Les moments μ_3 et μ_4 sont utilisés pour caractériser la forme de distribution.

Pour obtenir des quantités sans dimension, on utilise les coefficients d'asymétrie et d'aplatissement γ_1 et γ_2 (en anglais *skewness* et *kurtosis*) :

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4}$$

La figure 2.9 donne quelques allures typiques de courbes de densité correspondant à certaines valeurs de γ_1 et γ_2 .

On remarquera que γ_2 est toujours supérieur à 1 car l'inégalité classique entre moyennes d'ordre p entraîne $(\mu_4)^{1/4} > (\mu_2)^{1/2} \Rightarrow \mu_4 > (\mu_2)^2$.

De plus, on a toujours $\gamma_2 \geq 1 + (\gamma_1)^2$.

Plus que l'aplatissement, le coefficient γ_2 mesure l'importance des « queues » de distribution.

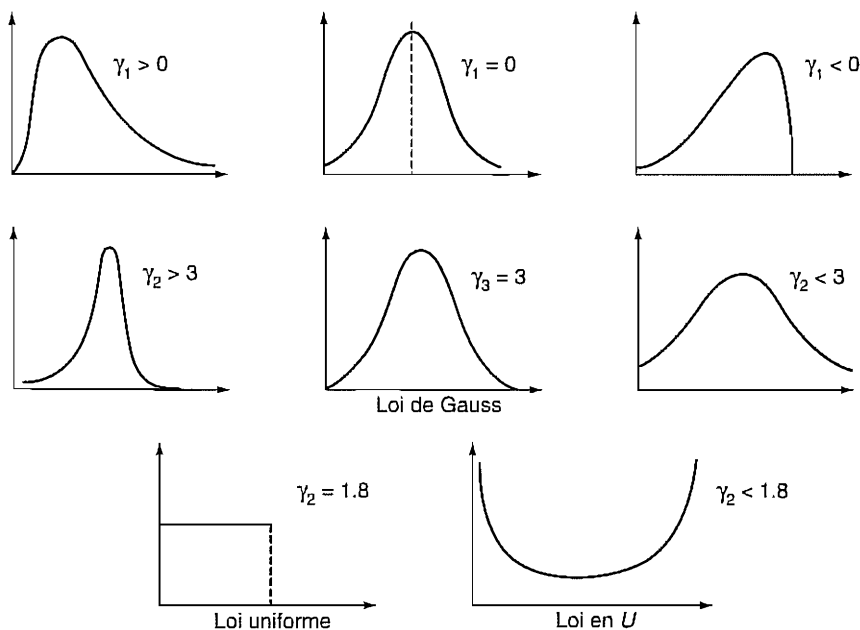


FIGURE 2.9

Inégalité de Markov : En utilisant la même méthode que pour l'inégalité de Bienaymé-Tchebyshev, on montre que :

$$P(|X| > \varepsilon) \leq \frac{E(X^k)}{\varepsilon^k}$$

2.1.4.4 Ordres stochastiques

Les concepts de dominance stochastique sont utilisés dans différents domaines, en particulier en fiabilité pour comparer des fonctions de survie, et en théorie de la décision pour comparer des risques.

A. Dominance stochastique d'ordre 1

On dit que X domine stochastiquement Y si la fonction de survie de X est supérieure à celle de Y :

$$P(X > c) \geq P(Y > c) \text{ pour tout } c$$

ce qui revient à dire que la fonction de répartition de X est toujours inférieure à celle de Y .

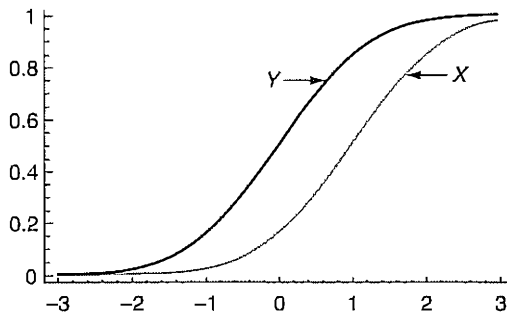


FIGURE 2.10

THÉORÈME (ADMIS)

L Pour que X domine stochastiquement Y , il faut et il suffit que $E(f(X)) \geq E(f(Y))$ pour toute fonction f croissante.

On en déduit que la dominance stochastique de X sur Y entraîne $E(X) \geq E(Y)$.

On peut montrer (exercice à faire . . .) la propriété suivante : si la fonction de hasard (ou taux de défaillance) de X est partout inférieure à celle de Y , alors X domine stochastiquement Y . C'est par exemple le cas de la durée de vie des femmes en France qui domine celle des hommes : non seulement l'espérance de vie des femmes est plus élevée que celle des hommes, mais également la probabilité de survie à tout âge.

B. Dominance stochastique d'ordre 2

La dominance d'ordre 1 implique que les fonctions de répartition de X et Y ne peuvent se croiser. Une forme plus faible de dominance, qui autorise les croisements, est définie comme suit :

DÉFINITION

X domine stochastiquement Y à l'ordre 2 si leurs fonctions de répartition sont telles que :

$$\int_{-\infty}^c F(x) dx \leq \int_{-\infty}^c G(x) dx \quad \text{pour tout } c.$$

L'inégalité porte cette fois sur les intégrales des fonctions de répartition. La dominance stochastique d'ordre 1 entraîne celle d'ordre 2.

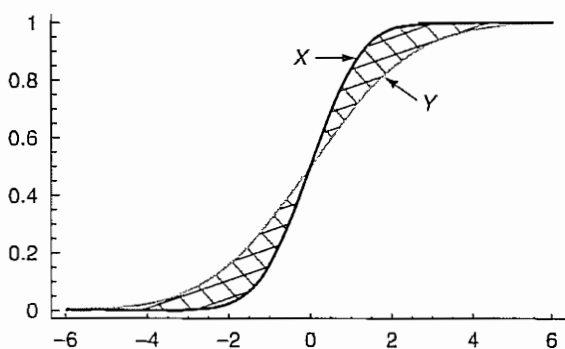


FIGURE 2.11

Cette forme de dominance est utilisée en théorie du risque pour des variables positives représentant des gains aléatoires. Supposons de plus que X et Y ont même espérance : alors les aires hachurées sur la figure précédente sont égales. On voit intuitivement que la répartition de X est moins dispersée que celle de Y . Un individu qui a de l'aversion pour le risque préférera donc X à Y . La dominance stochastique d'ordre 2 implique $V(X) < V(Y)$ mais est plus générale (la réciproque est fausse).

On montre que si X domine Y , Y a la même distribution que $X + \varepsilon$ où ε est une variable telle que $E(\varepsilon/X) = 0$. Intuitivement, Y est « plus aléatoire » que X .

Le théorème du paragraphe précédent est alors modifié comme suit [Rothschild et Stiglitz, 1970] :

THÉORÈME

Pour que X domine stochastiquement Y à l'ordre 2, il faut et il suffit que $E(f(X)) \geq E(f(Y))$ pour toute fonction f croissante concave.

2.2 LOIS DE PROBABILITÉ DISCRÈTES D'USAGE COURANT

2.2.1 Loi discrète uniforme

$$X = \{1, 2, 3, \dots, n\}$$

$$P(X = 1) = P(X = 2) = \dots = P(X = n) \quad (\text{fig. 2.12})$$

$$P(X = k) = \frac{1}{n}$$

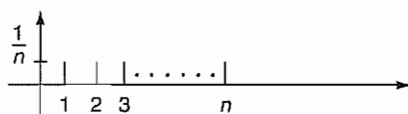


FIGURE 2.12

$$E(X) = \frac{n+1}{2} \quad \text{par symétrie}$$

$$E(X) = \frac{1}{n} (1 + 2 + \dots + n) = \frac{n+1}{2}$$

$$E(X^2) = \frac{1}{n} (1 + 4 + 9 + \dots + n^2)$$

$$E(X^2) = \frac{1}{n} \frac{n(n+1)(2n+1)}{6}$$

d'où :

$$V(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$V(X) = \frac{n+1}{12} (4n+2-3(n+1))$$

soit :

$$V(X) = \frac{n^2-1}{12}$$

2.2.2 Loi de Bernoulli de paramètre p

C'est la loi d'une variable X ne pouvant prendre que les deux valeurs 1 ou 0 avec les probabilités p et $1-p$; X est la fonction indicatrice d'un événement A de probabilité p :

$$E(X) = p$$

Comme $X^2 = X$, $E(X^2) = p$, d'où :

$$V(X) = p(1-p)$$

2.2.3 Loi binomiale $\mathcal{B}(n; p)$

A. Principe

Supposons que l'on répète n fois dans des conditions identiques une expérience aléatoire, dont l'issue se traduit par l'apparition ou la non-apparition d'un événement A de probabilité p , le résultat de chaque expérience étant indépendant des résultats précédents. Soit X le nombre d'apparitions de l'événement A parmi ces n expériences ($0 \leq X \leq n$). On dit alors que X suit une loi binomiale de paramètres n et p notée $\mathcal{B}(n; p)$. Comme à chaque expérience numérotée i ($i = 1, 2, \dots, n$), on peut associer une variable de Bernoulli X_i de paramètre p , on a : $X = \sum_{i=1}^n X_i$ d'où la deuxième définition de la loi binomiale : X suit une loi binomiale $\mathcal{B}(n; p)$ si X est une somme de n variables de Bernoulli indépendantes et de même paramètre p .

De cette définition, découlent l'espérance et la variance de X .

$E(X) = \sum E(X_i)$, donc : $E(X) = np$ $V(X) = \sum V(X_i)$ car les X_i sont indépendants ; donc :

$$V(X) = np(1 - p)$$

B. Loi de probabilité

Afin de chercher l'expression de $P(X = k)$, remarquons que toutes les configurations, telles que k variables X_i prennent la valeur 1 et $n - k$ la valeur 0, sont équiprobables et qu'il y a C_n^k configurations de cette sorte (nombre de manières de choisir k X_i parmi n).

D'autre part :

$$\begin{aligned} P(X_1 = x_1 \cap \dots \cap X_n = x_n) &= \prod_{i=1}^n P(X_i = x_i) \\ &= \prod_{i=1}^n p^{x_i} (1 - p)^{1 - x_i} \end{aligned}$$

car les X_i sont indépendants :

$$P(X_1 = x_1 \cap X_2 = x_2, \dots, \cap X_n = x_n) = p^{\sum x_i} (1 - p)^{n - \sum x_i}$$

Comme $\sum x_i = k$, on trouve :

$$P(X = k) = C_n^k p^k (1 - p)^{n - k}$$

Cette formule justifie le nom de la loi binomiale car les $P(X = k)$ sont les termes du développement de $(p + (1 - p))^n$ selon la formule du binôme de Newton (on vérifie au passage que $\sum_{k=0}^n P(X = k) = 1$).

La figure 2.13 représente quelques diagrammes en bâtons correspondant à diverses valeurs de n et p . On notera que la distribution est symétrique si $p = 1/2$ et le devient approximativement sinon, dès que n est assez élevé.

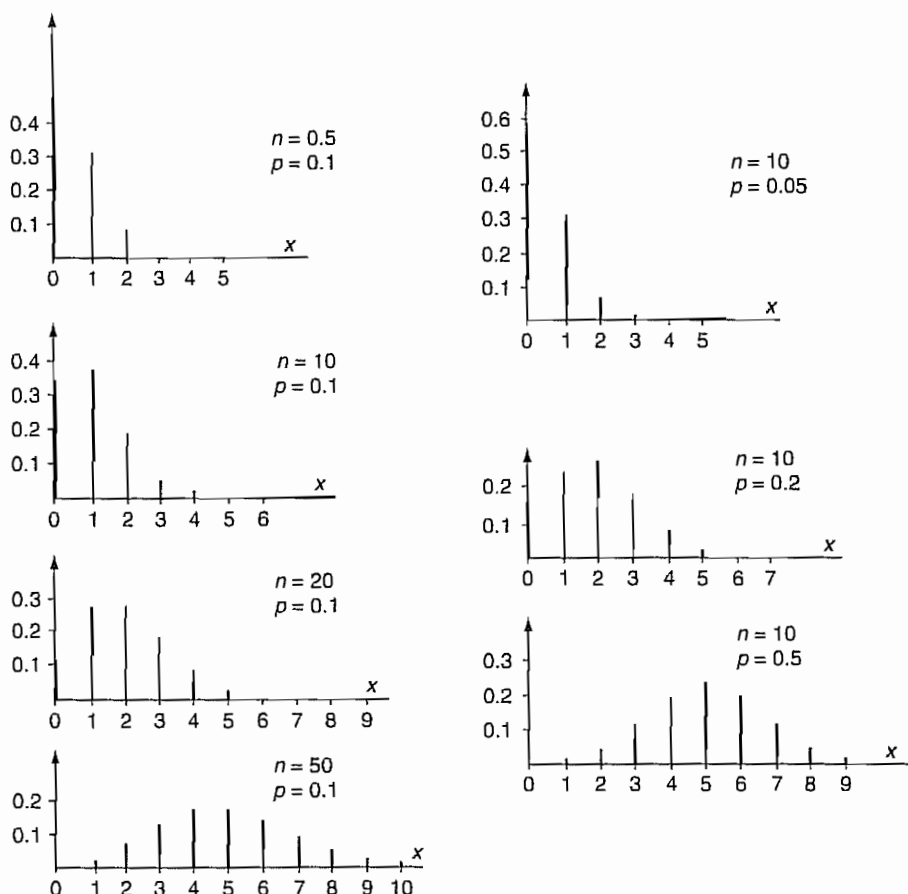


FIGURE 2.13

Un résultat utile pour l'utilisation des tables : si X suit une loi binomiale $\mathcal{B}(n; p)$, $n - X$ suit alors une loi binomiale $\mathcal{B}(n; 1 - p)$.

Pour n grand, on verra plus loin que la loi binomiale peut être approximée soit par une loi de Poisson (si p est petit) soit par une loi de Gauss.

La somme de deux variables aléatoires binomiales indépendantes et de même paramètre p est une variable aléatoire binomiale :

$$\left. \begin{array}{l} X_1 = \mathcal{B}(n_1, p) \\ X_2 = \mathcal{B}(n_2, p) \end{array} \right\} \Rightarrow X_1 + X_2 = \mathcal{B}(n_1 + n_2, p)$$

■ Démonstration

X_1 : somme de n_1 variables de Bernoulli ;

X_2 : somme de n_2 variables de Bernoulli.

$X_1 + X_2$, somme de $n_1 + n_2$ variables de Bernoulli est bien une variable binomiale d'effectif égal à la somme des effectifs.

Condition nécessaire et suffisante : X_1 et X_2 doivent être indépendantes.

2.2.4 Loi de Poisson $\mathcal{P}(\lambda)$

C'est la loi d'une variable aléatoire entière positive ou nulle qui satisfait à :

$$P(X = x) = \exp(-\lambda) \frac{\lambda^x}{x!} \quad \text{où} \quad x \in \mathbb{N}$$

On peut vérifier tout d'abord qu'il s'agit bien d'une loi de probabilité :

$$\sum_{x=0}^{\infty} P(X = x) = \exp(-\lambda) \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \exp(-\lambda) \exp(\lambda) = 1$$

À la figure 2.12, quelques diagrammes en bâtons correspondent à diverses valeurs de λ :

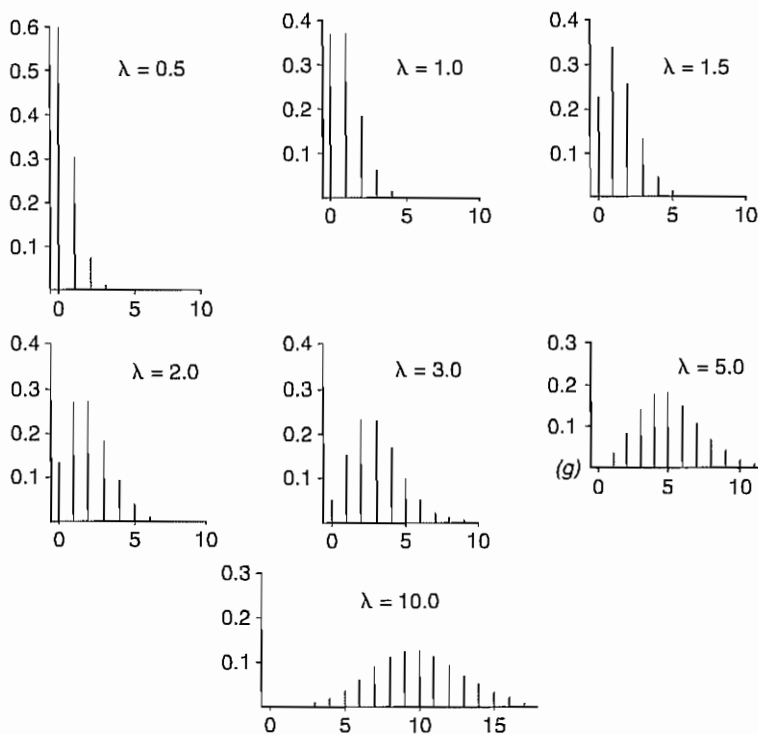


FIGURE 2.14

Le paramètre λ représente à la fois l'espérance et la variance de X .

On obtient la loi de Poisson comme approximation de la loi binomiale dans le schéma suivant :

Soit un événement A de probabilité p très faible (en pratique $p < 0.1$) que l'on essaie d'obtenir quelques fois en répétant l'expérience un grand nombre de fois (en pratique $n > 50$). Le nombre de réalisations de A suit une loi binomiale $\mathcal{B}(n; p)$ telle qu'en pratique :

$$\mathcal{B}(n, p) \sim \mathcal{P}(np)$$

c'est-à-dire :

$$C_n^k p^k (1-p)^{n-k} \approx \exp(-np) \frac{(np)^k}{k!}$$

Nous allons, en fait, établir ce résultat sous la forme mathématique suivante :

THÉORÈME

L Soit X_n une suite de variables binomiales $\mathcal{B}(n, p)$ telles que $n \rightarrow \infty$ et $p \rightarrow 0$ de manière à ce que le produit np tende vers une limite finie λ . Alors la suite de variables aléatoires X_n converge en loi vers une variable de Poisson $\mathcal{P}(\lambda)$.

Les notions de convergence seront étudiées en détail au paragraphe 2.7.

■ Démonstration

$$\begin{aligned} C_n^x p^x (1-p)^{n-x} &= \frac{n(n-1) \dots (n-x+1)}{x!} p^x (1-p)^{n-x} \\ &= \frac{(np)^x}{x!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{x-1}{n}\right) (1-p)^{n-x} \end{aligned}$$

Faisons tendre $n \rightarrow \infty$. Tous les termes $\left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{x-1}{n}\right)$ tendent vers 1, leur produit tend vers 1 car ils sont en nombre fini.

Décomposons $(1-p)^{n-x}$ en $(1-p)^n (1-p)^{-x}$

$$(1-p)^{-x} \rightarrow 1 \text{ car } p \rightarrow 0.$$

Quant à $(1-p)^n \sim \left(1 - \frac{\lambda}{n}\right)^n$ il tend vers $\exp(-\lambda)$ donc :

$$C_n^x p^x (1-p)^{n-x} \rightarrow \left(\frac{np}{x!}\right)^x \exp(-\lambda) \quad \text{c.q.f.d.}$$

La suite des espérances des binomiales $X_n : E(X_n) = np$ converge vers λ :

$$E(X) = \lambda$$

En effet :
$$E(X) = \sum_{x=0}^{\infty} \exp(-\lambda) \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} \exp(-\lambda) \frac{\lambda^x}{(x-1)!}$$

car le premier terme est nul :

$$\begin{aligned} E(X) &= \exp(-\lambda) \lambda \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \exp(-\lambda) \lambda \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= \exp(-\lambda) \lambda \exp(\lambda) = \lambda \end{aligned}$$

La suite des variances des binomiales X_n : $V(X_n) = np(1-p)$ tend aussi vers λ car $np \rightarrow \lambda$, $p \rightarrow 0$.

Montrons que $V(X) = \lambda$.

Démonstration

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 = E(X^2) - \lambda^2 \\ E(X^2) &= \sum_{x=0}^{\infty} x^2 \exp(-\lambda) \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} x \exp(-\lambda) \frac{\lambda^x}{(x-1)!} \end{aligned}$$

avec $x = x - 1 + 1$, il vient :

$$\begin{aligned} E(X^2) &= \sum_{x=2}^{\infty} \exp(-\lambda) \frac{\lambda^x}{(x-2)!} + \sum_{x=1}^{\infty} \exp(-\lambda) \frac{\lambda^x}{(x-1)!} \\ E(X^2) &= \lambda^2 \exp(-\lambda) \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda \exp(-\lambda) \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ E(X^2) &= \lambda^2 \exp(-\lambda) \exp(\lambda) + \lambda \exp(-\lambda) \exp(\lambda) = \lambda^2 + \lambda \end{aligned}$$

donc $V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Donc $\sigma = \sqrt{\lambda}$.

On verra plus loin que la somme de deux variables de Poisson indépendantes est encore une variable de Poisson. Lorsque λ est grand, on verra que la loi de Poisson peut être approximée par la loi de Gauss.

La loi de Poisson s'obtient aussi comme loi exacte du nombre d'événements survenant pendant une période donnée, sous certaines conditions (voir plus loin le paragraphe consacré au processus de Poisson).

Exemples d'application de la loi de Poisson :

- loi du nombre de suicides par an dans un pays donné ;
- loi du nombre d'appels téléphoniques pendant un intervalle de temps T ;
- loi du nombre de pièces défectueuses dans une livraison importante, la production étant de bonne qualité ;
- etc.

2.2.5 Loi hypergéométrique $\mathcal{H}(N, n, p)$ ou du tirage exhaustif

Soit une population de N individus parmi lesquels une proportion p (donc Np individus) possède un certain caractère. On prélève un échantillon de n individus parmi cette population (le tirage pouvant s'effectuer d'un seul coup ou au fur et à mesure mais sans remise). Soit X le nombre aléatoire d'individus de l'échantillon possédant la propriété envisagée. X suit la loi hypergéométrique et l'on a :

$$P(X = x) = \frac{C_{Np}^x C_{N-Np}^{n-x}}{C_N^n}$$

$\min X = \max(0; n - Nq)$;

$\max X = \min(n; Np)$;

C_N^n nombre d'échantillons possibles ;

C_{Np}^x nombre de groupes de x individus possédant la propriété ;

C_{N-Np}^{n-x} nombre de groupes de $(n - x)$ individus ne possédant pas la propriété.

Le nombre n/N est appelé taux de sondage.

On peut considérer X comme une somme de n variables de Bernoulli X_1, X_2, \dots, X_n non indépendantes correspondant aux tirages successifs de n individus.

Nous allons montrer que ces variables X_i ont toutes le même paramètre égal à p .

On sait que $E(X_1) = P(X_1 = 1)$ et il est évident que $P(X_1 = 1) = p$.

Cherchons $E(X_2) = P(X_2 = 1)$. Comme X_2 et X_1 sont liés, on a :

$$P(X_2 = 1) = P(X_2 = 1 | X_1 = 1)P(X_1 = 1) + P(X_2 = 1 | X_1 = 0)P(X_1 = 0)$$

$$\begin{aligned} \text{soit : } P(X_2 = 1) &= \frac{Np - 1}{N - 1} p + \frac{Np}{N - 1} (1 - p) \\ &= \frac{Np^2 - p + Np - Np^2}{N - 1} = p \frac{(N - 1)}{N - 1} = p \end{aligned}$$

De même $E(X_3) = E(X_4) = \dots = p$.

2.2.5.1 Espérance de l'hypergéométrique

$$E(X) = E(\sum X_i) = \sum E(X_i) \quad \boxed{E(X) = np}$$

L'espérance ne dépend pas de N et est la même que dans le cas du tirage avec remise (loi binomiale).

2.2.5.2 Variance de l'hypergéométrique

Comme il n'y a pas indépendance :

$$V(X) = \sum V(X_i) + 2 \sum_{i>j} \text{cov}(X_i, X_j) = \sum V(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j)$$

le terme $\sum V(X_i)$ vaut $np(1-p)$ (terme binomial).

$$\text{On a : } \text{cov}(X_i, X_j) = E(X_i X_j) - p^2 = P(X_i X_j = 1) - p^2$$

$$\text{et : } P(X_i X_j = 1) = P(X_j = 1 | X_i = 1) P(X_i = 1) = P(X_j = 1 | X_i = 1) p$$

$P(X_j = 1 | X_i = 1)$ ne dépend pas des indices i et j et vaut par exemple

$$P(X_2 = 1 | X_1 = 1) = \frac{Np - 1}{N - 1}.$$

$$\text{Donc : } \text{cov}(X_i, X_j) = p \frac{Np - 1}{N - 1} - p^2$$

Comme il y a $n(n-1)$ manières de prendre des couples $(X_i \text{ et } X_j)$, il vient :

$$V(X) = np(1-p) + n(n-1) \left[p \frac{Np - 1}{N - 1} - p^2 \right]$$

soit :

$$V(X) = \frac{N-n}{N-1} np(1-p)$$

2.2.5.3 Tendence vers la loi binomiale

Si $N \rightarrow \infty$, $\mathcal{H}(N, n, p)$ tend vers $\mathcal{B}(n, p)$.

■ Démonstration

$$\begin{aligned} \frac{C_{Np}^x C_{N-Np}^{n-x}}{C_N^n} &= \frac{Np!}{(Np-x)!x!} \frac{(N(1-p))!}{(n-x)!(N-Np-n+x)!} \frac{n!(N-n)!}{N!} \\ &= C_n^x \frac{Np!}{(Np-x)!} \frac{Nq!}{(Nq-n+x)!} \frac{(N-n)!}{N!} \end{aligned}$$

avec $q = 1 - p$.

$$\frac{Np!}{(Np-x)!} = \frac{1 \cdot 2 \cdot 3 \cdots Np}{1 \cdot 2 \cdot 3 \cdots (Np-x)} = Np(Np-1) \cdots (Np-x+1)$$

Si N est grand, $Np-1 \sim Np-2 \cdots \sim (Np-x+1) \sim Np$ car x est négligeable devant Np .

$$\text{Donc : } \frac{Np!}{(Np-x)!} \sim (Np)^x$$

$$\text{De même : } \frac{Nq!}{(Nq-n+x)!} \sim (Nq)^{n-x} \quad \text{et} \quad \frac{N!}{(N-n)!} \sim N^n$$

$$\text{donc : } \frac{C_{Np}^x C_{Nq}^{n-x}}{C_N^n} \sim C_n^x \frac{(Np)^x (Nq)^{n-x}}{N^n} = C_n^x p^x q^{n-x} \quad \text{c.q.f.d.}$$

En pratique, ce résultat s'applique dès que $n/N < 10\%$, c'est-à-dire dès que la population est 10 fois plus grande que l'échantillon, ce qui arrive fréquemment en sondages.

Un échantillon de 2000 individus conviendra donc aussi bien pour faire un sondage dans une ville de 200 000 habitants que dans une ville de 2 millions d'habitants.

2.2.6 Lois géométrique, de Pascal, binomiale négative

La **loi géométrique** est la loi du nombre d'essais nécessaires pour faire apparaître un événement de probabilité p :

$$P(X = x) = p(1 - p)^{x-1} \quad x = 1, 2, \dots, \infty$$

En posant $q = 1 - p$, on trouve aisément :

$$E(X) = \frac{1}{p} \quad V(X) = \frac{q}{p^2} \quad \gamma_1 = \frac{2-p}{q} \quad \gamma_2 = 9 + \frac{p^2}{q}$$

La **loi de Pascal** d'ordre n est la loi du nombre d'essais nécessaires pour observer n fois un événement A de probabilité p . L'expérience devant se terminer par A , on a :

$$P(X = x) = pC_{x-1}^{n-1} p^{n-1} q^{(x-1)-(n-1)} = C_{x-1}^{n-1} p^n q^{x-n} \quad \text{pour } x = n, n+1, \dots, \infty$$

Cette loi est la somme de n lois géométriques indépendantes (apparition de A pour la première fois, puis pour la deuxième fois, etc.), on a :

$$E(X) = \frac{n}{p} \quad V(X) = \frac{nq}{p^2} \quad \gamma_1 = \frac{2-p}{\sqrt{nq}} \quad \gamma_2 = 3 + \frac{p^2 + 6q}{nq}$$

La **loi binomiale négative** est la loi de $Y = X - n$:

$$P(Y = y) = C_{n+y-1}^{n-1} p^n q^y$$

Son nom vient du fait suivant : en posant $Q = 1/p$, $P = (1-p)/p$, on a :

$$P(Y = y) = C_{n+y-1}^{n-1} P^y Q^{-n-y}$$

terme général du développement de $(Q - P)^{-n}$ d'où :

$$E(X) = nP \quad V(Y) = nPQ \quad \gamma_1 = \frac{P+Q}{\sqrt{nPQ}} \quad \gamma_2 = 3 + \frac{1+6PQ}{nPQ}$$

que l'on comparera aux moments de la binomiale $\mathcal{B}(n, p)$.

2.3 DISTRIBUTIONS CONTINUES USUELLES

2.3.1 Loi uniforme sur $[0, a]$

Sa densité est :

$$f(x) = \frac{1}{a} \text{ sur } [0, a] ;$$

$$f(x) = 0 \text{ ailleurs ;}$$

$$F(x) = \frac{x}{a} \text{ sur } [0, a];$$

$$F(x) = 0 \text{ sur } [-\infty, 0]; F(x) = 1 \text{ sur } [a, +\infty] \text{ (voir fig. 2.13).}$$

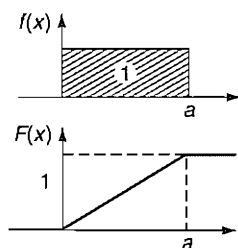


FIGURE 2.15

Son espérance vaut $E(X) = \frac{a}{2}$ car la densité est symétrique.

$$\text{Sa variance vaut } V(X) = \int_0^a x^2 \frac{1}{a} dx - \frac{a^2}{4} = \frac{a^2}{12}.$$

La somme de deux lois uniformes n'est pas une loi uniforme. Ainsi, soit X et Y deux variables uniformes sur $[0, a]$; leur somme Z , si elles sont indépendantes, est une variable de densité triangulaire (fig. 2.16).

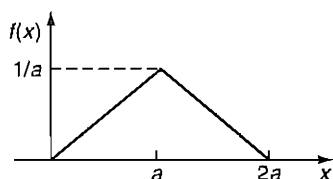


FIGURE 2.16

2.3.2 Loi exponentielle de paramètre λ

Sa densité est $f(x) = \lambda \exp(-\lambda x)$ si $x > 0$.

On trouve sans difficulté :

$$E(X) = 1/\lambda$$

$$V(X) = 1/\lambda^2$$

$$\gamma_1 = 2$$

$$\gamma_2 = 9$$

En fiabilité, cette loi est très utilisée pour représenter la durée de vie de circuits électroniques. L'espérance $1/\lambda$ est souvent appelée le MTBF (*Mean Time Between Failure*) et λ le

taux de défaillance car $h(x) = \frac{f(x)}{1 - F(x)} = \lambda$ et est constant.

2.3.3 Lois gamma

La loi exponentielle est un cas particulier d'une famille de lois appelées lois γ . Précisément, si X est une loi exponentielle de paramètre λ , λX est une variable suivant une loi γ_1 .

On dit qu'une variable aléatoire positive X suit une loi gamma de paramètre r , notée γ_r , si sa densité est donnée par :

$$f(x) = \frac{1}{\Gamma(r)} \exp(-x) x^{r-1}$$

Il s'agit bien d'une densité car $f(x)$ est > 0 et $\int_0^\infty f(x) dx = 1$ par définition de $\Gamma(r)$. Les lois γ_r avec r entier > 1 sont aussi connues sous le nom de lois d'Erlang.

2.3.3.1 Espérance

$$E(X) = r$$

En effet :

$$E(X) = \frac{1}{\Gamma(r)} \int_0^\infty x^r \exp(-x) dx = \frac{\Gamma(r+1)}{\Gamma(r)} = r$$

2.3.3.2 Variance

$$V(X) = r$$

En effet :

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{\Gamma(r)} \int_0^\infty x^{r+1} \exp(-x) dx - r^2$$

soit :

$$V(X) = \frac{\Gamma(r+2)}{\Gamma(r)} - r^2 = (r+1) \frac{\Gamma(r+1)}{\Gamma(r)} - r^2 = r(r+1) - r^2$$

Cette loi présente donc une certaine analogie avec la loi de Poisson mais en continu. Les courbes de densité sont représentées à la figure 2.17.

Les lois γ vérifient la propriété d'additivité suivante :

THÉORÈME

L Si X et Y sont des variables indépendantes suivant respectivement des lois γ_r et γ_s , alors $X + Y$ suit une loi γ_{r+s} .

Ce résultat sera démontré au paragraphe 2.5 de ce chapitre.

Les lois γ sont liées aux lois du χ^2 utilisées en statistique par une formule simple (voir chapitre 4) :

Si X suit une loi γ_r , $2X$ suit une loi χ^2_{2r} .

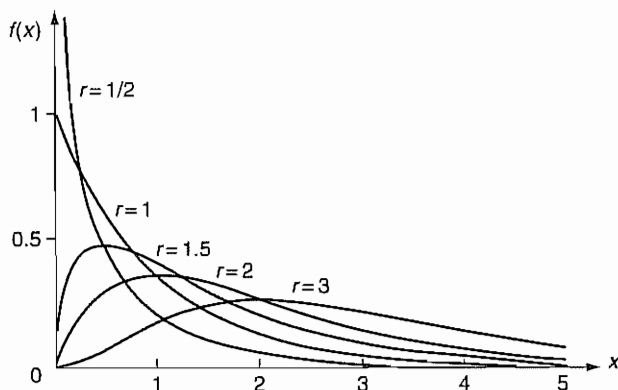


FIGURE 2.17

2.3.4 Lois bêta

2.3.4.1 Loi bêta de type I

C'est la loi d'une variable X ; $0 \leq X \leq 1$ dépendant de deux paramètres n et p dont la densité est :

$$f(x) = \frac{1}{B(n, p)} x^{n-1} (1-x)^{p-1} \quad n, p > 0 \quad \text{où } B(n, p) = \frac{\Gamma(n) \Gamma(p)}{\Gamma(n+p)}$$

On trouve :

$$E(X) = \frac{n}{n+p}$$

$$V(X) = \frac{np}{(n+p+1)(n+p)^2}$$

Ces lois sont utilisées en statistique bayésienne pour représenter la distribution *a priori* de la probabilité d'un événement.

L'allure de quelques courbes de densité est donnée par la figure 2.18.

2.3.4.2 Loi bêta de type II

Soit X une variable suivant une loi bêta $I(n, p)$; alors, par définition, $Y = X/(1-X)$ suit une loi bêta de type II dont la densité s'obtient aisément par changement de variable :

$$f(y) = \frac{1}{B(n, p)} \frac{y^{n-1}}{(1+y)^{n+p}} \quad 0 \leq Y < \infty$$

$$E(Y) = \frac{n}{p-1}$$

$$V(Y) = \frac{n(n+p-1)}{(p-1)^2(p-2)}$$

PROPRIÉTÉ

L Le rapport de deux variables indépendantes suivant des lois γ_n et γ_p respectivement suit une loi bêta $II(n, p)$.

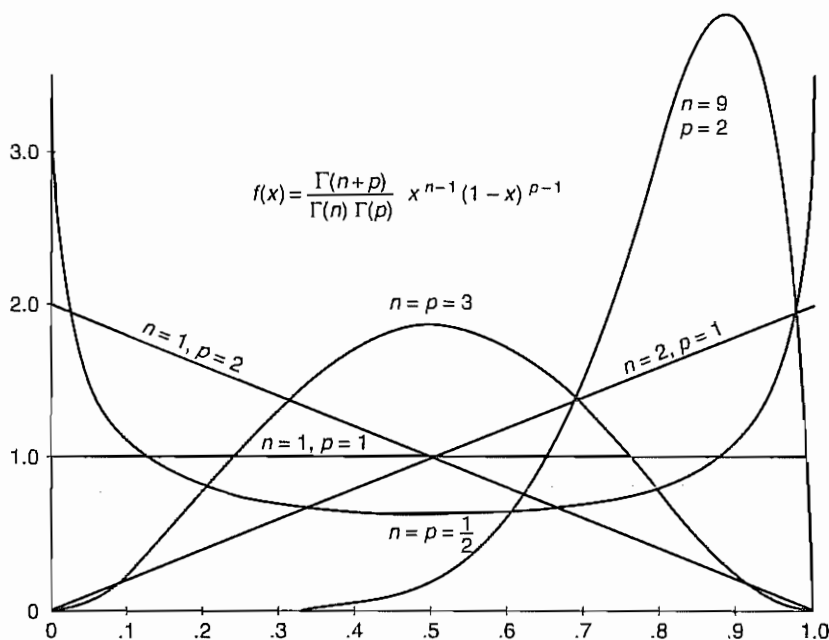


FIGURE 2.18

La démonstration est laissée au soin du lecteur.

Les diverses valeurs de n et p font que cette loi s'adapte bien à la représentation de nombreux phénomènes aléatoires positifs (temps d'attente, durées de vie, méthode Pert avec durée aléatoire).

Ces lois sont liées aux lois de Fisher-Snedecor utilisées en statistique (voir chapitre 4).

2.3.4.3 Loi de l'arc sinus

La loi bêta $I(1/2; 1/2)$ dont la densité est $f(x) = \frac{1}{\pi \sqrt{x(1-x)}}$ porte le nom de loi de l'arc sinus car sa fonction de répartition est :

$$F(x) = \frac{2}{\pi} \arcsin(\sqrt{x})$$

On a $E(X) = 1/2$, $V(X) = 1/8$, $\gamma_1 = 0$, $\gamma_2 = 1.5$.

Cette loi assez paradoxale, puisque l'espérance est la valeur la moins probable et les valeurs extrêmes sont les plus probables, s'applique en particulier dans certains phénomènes liés aux jeux de hasard.

Par exemple, deux joueurs jouent à un jeu équitable (du type pile ou face). Soit S_1, S_2, \dots, S_n la suite des gains d'un des deux joueurs ; si X désigne la proportion du temps passé en gain positif, la loi limite de X quand $n \rightarrow \infty$ est la loi de l'arc sinus. Il y a donc plus de chance d'être constamment en gain ou constamment en perte que d'être dans le cas médian (c'est la loi de la persistance de la chance ou de la malchance ...).

Cette loi a pu être appliquée à la persistance du temps en météorologie et rend compte du fait qu'il est plus fréquent de battre des records (de froid ou de chaud) que d'avoir un temps moyen.

2.3.5 La loi de Laplace-Gauss

Cette loi joue un rôle fondamental en probabilités et statistique mathématique. Elle constitue un modèle fréquemment utilisé dans divers domaines : variation du diamètre d'une pièce dans une fabrication industrielle, répartition des erreurs de mesure autour de la « vraie valeur », etc.

Malgré son appellation malencontreuse de loi normale⁽¹⁾, elle est cependant loin de décrire tous les phénomènes physiques et il faut se garder de considérer comme anormale une variable ne suivant pas la loi de Laplace-Gauss. Son rôle principal en statistique provient en réalité de ce qu'elle apparaît comme loi limite de caractéristiques liées à un échantillon de grande taille. Le théorème central-limite que nous établirons au paragraphe 2.7 montre que dans certaines conditions la somme, et donc la moyenne, de variables indépendantes et de même loi est asymptotiquement une loi normale.

X suit une loi normale $LG(m; \sigma)$ si sa densité est⁽²⁾ :

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

Par suite de la symétrie de f et comme l'intégrale de X converge, $E(X) = m$.

Avec le changement de variable aléatoire $U = \frac{X-m}{\sigma}$, on trouve que la densité de

U est :

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

U est une $LG(0, 1)$, donc toute variable X $LG(m; \sigma)$ se ramène simplement à la variable U par $X = m + \sigma U$.

Montrons que $V(U) = 1$:

$$V(U) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} u^2 \exp\left(-\frac{1}{2}u^2\right) du = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} u^2 \exp\left(-\frac{1}{2}u^2\right) du$$

Posons $t = u^2/2$, il vient $u du = dt$:

$$V(U) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \exp(-t) dt = \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right)$$

¹ Cette dénomination fut introduite par K. Pearson qui voulait éviter les querelles d'antériorité concernant son introduction en statistique et l'a d'ailleurs regretté par la suite comme l'indique cette citation : *Many years ago I called the Laplace-Gaussian curve the normal curve which name, while it avoids an international question of priority, has the disadvantage of leading people to believe that all other distributions of frequency are in one sense or another 'abnormal'. That belief is, of course, not justifiable. It has led many writers to try and force all frequency by aid of one or another process of distortion into a 'normal' curve (paper read to the Society of Biometricians and Mathematical Statisticians, June 14, 1920).*

² La notation LG sera utilisée couramment dans cet ouvrage. La notation $N(m; \sigma)$ sera également utilisée.

comme $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$: $V(U) = 1$.

Il en résulte que σ est l'écart-type de X .

La fonction de répartition et la densité de X sont représentées sur la figure 2.19.

Les points d'inflexion sont à $\pm\sigma$ de part et d'autre de m .

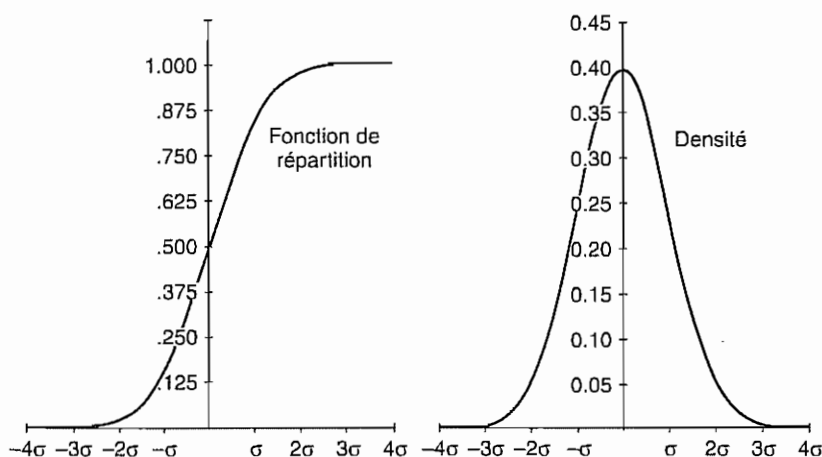


FIGURE 2.19

2.3.5.1 Valeurs remarquables

$$\begin{aligned} P(m - 1.64\sigma < X < m + 1.64\sigma) &= 0.90 \\ P(m - 1.96\sigma < X < m + 1.96\sigma) &= 0.95 \\ P(m - 3.09\sigma < X < m + 3.09\sigma) &= 0.998 \end{aligned}$$

2.3.5.2 Moments

Ils existent pour tout ordre.

Par suite de la symétrie, tous les moments d'ordre impair sont nuls. Calculons les moments d'ordre pair :

$$\mu_{2k} = \int_{\mathbb{R}} u^{2k} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right) du = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} u^{2k} \exp\left(-\frac{1}{2} u^2\right) du$$

Posons $y = u^2/2$:

$$\mu_{2k} = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (2y)^k \exp(-y) \frac{dy}{\sqrt{2y}} = \frac{2^k}{\sqrt{\pi}} \int_0^{\infty} y^{k-\frac{1}{2}} \exp(-y) dy$$

d'où :

$$\mu_{2k} = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right)$$

Comme :

$$\Gamma\left(k + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \Gamma\left(\frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots 2k-1}{2^k} \sqrt{\pi}$$

(voir annexes) il vient :

$$\mu_{2k} = 1 \cdot 3 \cdots (2k-1) = \frac{(2k)!}{2^k k!}$$

on en déduit $\mu_4 = 3$, d'où $\gamma_2 = 3$.

2.3.5.3 Additivité

Les variables de Gauss possèdent la propriété d'additivité.

THÉORÈME

L Si X_1 et X_2 sont des variables indépendantes suivant respectivement des lois $\text{LG}(m_1; \sigma_1)$ et $\text{LG}(m_2; \sigma_2)$ alors $X_1 + X_2$ est une variable $\text{LG}(m_1 + m_2; \sqrt{\sigma_1^2 + \sigma_2^2})$.

Ce résultat fondamental sera démontré au paragraphe 2.6 à l'aide des fonctions caractéristiques.

On ne peut cependant pas dire que toute combinaison linéaire de p variables gaussiennes non indépendantes soit encore gaussienne. Il faut pour cela que le p -uplet de variables suive une loi normale à p -dimensions (dont c'est précisément la définition. cf. chapitre 4).

2.3.5.4 Loi de U^2

D'après la formule établie à la fin du paragraphe 2.1.2.2, la densité de $T = U^2$ est :

$$g(t) = \frac{f(\sqrt{t})}{\sqrt{t}} = \frac{1}{\sqrt{2\pi}} t^{-1/2} \exp\left(-\frac{t}{2}\right)$$

en remplaçant $f(t)$ par $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right)$, on remarque que $U^2/2$ suit une loi $\gamma_{1/2}$ ou loi du khi-deux à un degré de liberté (voir chapitre 4).

2.3.6 La loi log-normale

C'est la loi d'une variable positive X telle que son logarithme népérien suive une loi de Laplace-Gauss :

$$\ln X \sim \text{LG}(m; \sigma)$$

Sa densité s'obtient par un simple changement de variable et on trouve :

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

$$E(X) = \exp\left(m + \frac{\sigma^2}{2}\right)$$

$$V(X) = (\exp(2m + \sigma^2))(\exp \sigma^2 - 1)$$

On utilise parfois la loi log-normale à trois paramètres γ , m , σ telle que :

$$\ln(X - \gamma) \sim \text{LG}(m; \sigma) \quad \text{avec } X > \gamma.$$

La figure 2.20 représente la densité de la loi log-normale d'espérance 2 et d'écart-type 1 :

$$(m \approx 0.58 \quad \sigma \approx 0.47)$$

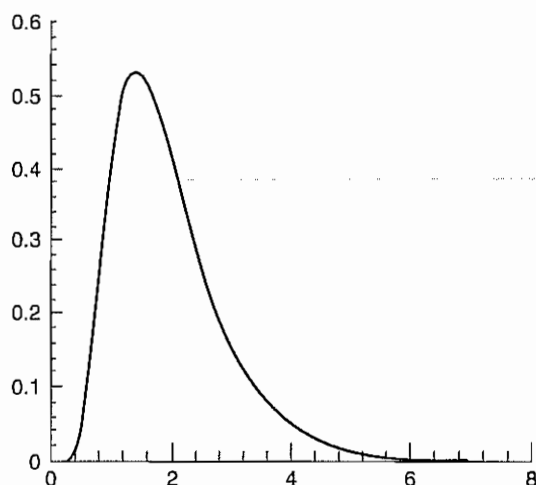


FIGURE 2.20

2.3.7 Loi de Cauchy

C'est la loi d'une variable X réelle de densité :

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Sa fonction de répartition est $F(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}$.

X ne possède aucun moment fini car l'intégrale $\int_{\mathbb{R}} \frac{x}{\pi(1+x^2)} dx$ diverge.

On montre que la loi de Cauchy est la loi du rapport de deux variables $\text{LG}(0; 1)$ indépendantes. Elle s'identifie à T_1 variable de Student de degré 1 (voir chapitre 4).

2.3.8 Loi de Weibull

Très utilisée en fiabilité, la loi de Weibull à deux paramètres donne la probabilité qu'une durée X de fonctionnement sans défaillance soit supérieure à x par :

$$P(X > x) = e^{-\left(\frac{x}{\beta}\right)^\alpha}$$

En d'autres termes, $\left(\frac{X}{\beta}\right)^\alpha$ suit une loi exponentielle.

La densité de X est : $f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$

Le paramètre α , qui est sans dimension, est appelé paramètre de forme. Selon ses valeurs, la densité de probabilité est plus ou moins dissymétrique. Le paramètre de forme est lié au vieillissement : quand il vaut 1, on a une loi exponentielle caractéristique des matériels sans usure ni fatigue. Quand il est plus grand que 1, on est en présence de fatigue : le taux instantané de défaillance $h(x)$ est alors croissant avec x :

$$h(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1}$$

Si α est inférieur à 1, on a affaire à un matériel qui se bonifie avec le temps.

Le paramètre β s'exprime dans la même unité que X (jours, heures, nombre de cycles, etc.). C'est un paramètre d'échelle lié à la durée de vie médiane par :

$$\beta = \frac{\text{médiane}}{(\ln(2))^{\frac{1}{\alpha}}}$$

La figure 2.21 donne la densité d'une loi de Weibull avec $\alpha = 2$ et $\beta = 1$.

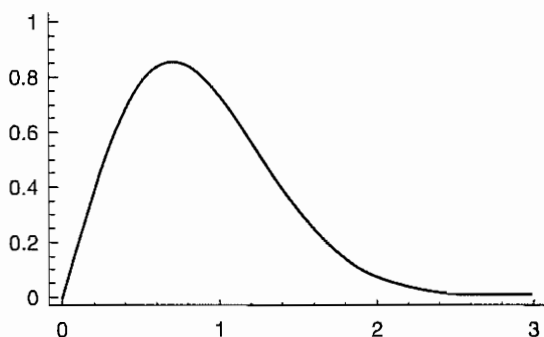


FIGURE 2.21

La relation $E\left[\left(\frac{X}{\beta}\right)^r\right] = \Gamma\left(1 + \frac{r}{\alpha}\right)$ permet de calculer les moments de X . Dans l'exemple précédent $\alpha = 2$ et $\beta = 1$, on trouve $E(X) = \frac{\sqrt{\pi}}{2}$ et $V(X) = \frac{3\pi}{4}$ (voir annexe 4).

2.3.9 Loi de Gumbel

Cette loi est utilisée pour les distributions de valeurs extrêmes (voir chapitre 12). Sous sa forme standard sa fonction de répartition est :

$$F(x) = \exp(-\exp(-x))$$

soit : $\boxed{f(x) = \exp(-x - \exp(-x))}$ (fig. 2.22)

$\exp(-X)$ suit donc une loi γ_1 .

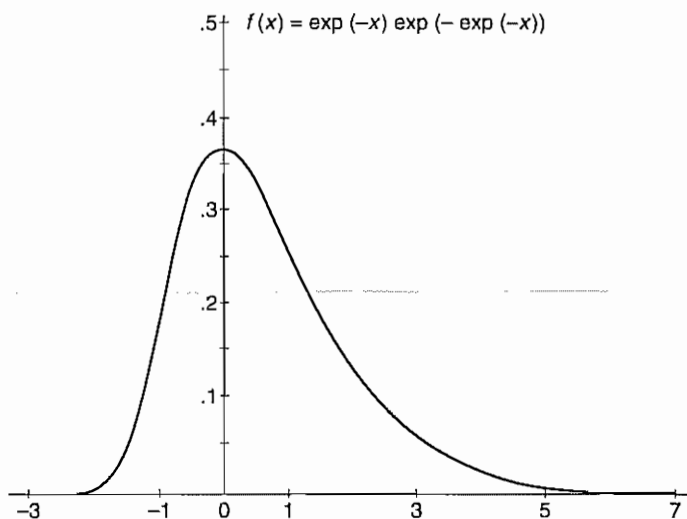


FIGURE 2.22

Ses moments sont :

$$E(X) = 0.57722 \dots \quad (\text{constante d'Euler})$$

$$V(X) = \frac{\pi^2}{6}$$

$$\gamma_1 = 1.29857$$

$$\gamma_2 = 5.4$$

La loi de Gumbel est utilisée pour modéliser des phénomènes tels que : crue maximale annuelle d'une rivière, magnitude du plus grand tremblement de terre enregistré en une année, etc.

2.4 LE PROCESSUS PONCTUEL DE POISSON

Considérons une famille X_t de variables de Bernoulli ($X_t = 1$ si un événement (arrivée d'un client, accident, appel téléphonique ...) se produit à l'instant t) : on s'intéressera à la répartition des dates d'arrivée des événements, ainsi qu'à N_t , nombre d'événements entre 0 et t .

2.4.1 Flux poissonnien d'événements

Un processus de Poisson représente l'apparition d'événements aléatoires E_1, E_2, \dots, E_n , etc., satisfaisant aux trois conditions suivantes :

- Les temps d'attente entre deux événements E_1, E_2, E_2, E_3 , etc. sont des variables indépendantes (processus sans mémoire).
- La loi du nombre d'événements arrivant dans l'intervalle $[t; t + T]$ ne dépend que de T . Si $T = 1$, on notera c son espérance, dite « cadence ».
- Deux événements ne peuvent arriver simultanément.

Soit $p_0(h)$ la probabilité qu'aucun événement ne se produise pendant une durée h ; d'après la deuxième condition, $p_0(h)$ ne dépend que de h et non de l'instant considéré.

Soient trois instants $t, t + h, t + h + k$. La probabilité qu'il ne se passe rien entre t et $t + h + k$ est $p_0(h + k)$; d'après l'axiome d'indépendance, on a :

$$p_0(h + k) = p_0(h) p_0(k) \quad \forall h, \forall k$$

D'où le résultat :

$$p_0(h) = \exp(-ch) \quad \text{avec } c > 0$$

Nous montrerons par la suite que c est bien la cadence du phénomène.

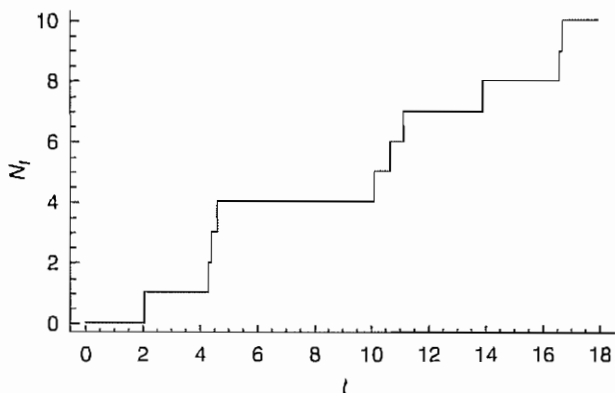


FIGURE 2.23 Une trajectoire d'un processus de Poisson avec $c = 1$; en ordonnée le nombre cumulé d'événements depuis $t = 0$.

2.4.2 Étude de la durée T séparant deux événements consécutifs E_i et E_{i+1}

Soit T cette durée qui est une variable aléatoire, la probabilité que $T > t$ est égale à la probabilité qu'il n'arrive rien pendant une durée t soit :

$$P(T > t) = \exp(-ct)$$

d'où la fonction de répartition de T : $P(T < t) = 1 - \exp(-ct)$. La densité vaut alors $f(t) = \exp(-ct)c$ il s'ensuit que cT suit une loi γ_1 , donc $E(T) = 1/c$.

2.4.3 Étude de la durée Y séparant $n + 1$ événements

Y est une variable aléatoire somme de n variables indépendantes de même loi :

$$Y = T_1 + T_2 + \dots + T_n$$

soit :

$$cY = cT_1 + cT_2 + \dots + cT_n \quad (\text{fig. 2.24})$$

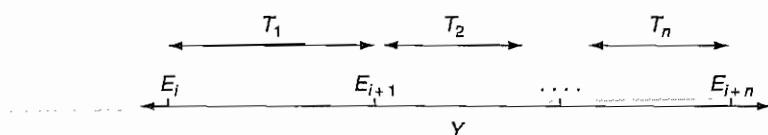


FIGURE 2.24

donc cY suit une loi γ_n ; la densité de Y est :

$$f_n(y) = \exp(-cy) \frac{(cy)^{n-1}}{(n-1)!} c$$

2.4.4 Étude du nombre d'événements se produisant pendant une période de durée T fixée

THÉORÈME

L La nombre d'événements suit une loi de Poisson de paramètre cT .

Démonstration : Soit AB la période d'étude (fig. 2.25) :



FIGURE 2.25

On a la relation évidente : $P(N = n) = P(N \geq n) - P(N \geq n + 1)$.

La probabilité $P(N \geq n)$ est aussi la probabilité que la durée AE_n soit inférieure à T ; cette durée est constituée de $AE_1 + E_1E_2 + \dots + E_{n-1}E_n$ qui sont des lois exponentielles indépendantes ; donc cAE_n suit une loi γ_n et l'on a :

$$P(N = n) = \int_0^T \exp(-ct) \frac{(ct)^{n-1}}{(n-1)!} c dt - \int_0^T \exp(-ct) \frac{(ct)^n}{n!} c dt$$

En intégrant par parties la première intégrale, il vient :

$$\begin{aligned} \int_0^T \exp(-ct) \frac{(ct)^{n-1}}{(n-1)!} c \, dt &= \int_0^T \exp(-ct) d\left(\frac{(ct)^n}{n!}\right) \\ &= \exp(-cT) \frac{(cT)^n}{n!} + \int_0^T \exp(-ct) \frac{(ct)^n}{n!} c \, dt \end{aligned}$$

donc :

$$P(N = n) = \exp(-cT) \frac{(cT)^n}{n!}$$

$E(N) = cT$, en particulier si $T = 1$.

On trouve $E(N) = c$; c est donc bien la cadence définie au début de cette partie.

Application importante : Relation entre loi de Poisson et loi du χ^2

Si N suit une loi $\mathcal{P}(\lambda)$ on a :

$$P(N \leq n) = P(\chi_{2(n+1)}^2 > 2\lambda)$$

Il suffit de considérer un processus de Poisson de cadence $c = 1$, observé sur une durée λ :

$$\begin{aligned} P(N \leq n) &= P(T_1 + T_2 + \dots + T_{n+1} > \lambda) = P(\gamma_{n+1} > \lambda) \\ &= P(2\gamma_{n+1} > 2\lambda) = P(\chi_{2(n+1)}^2 > 2\lambda) \end{aligned}$$

2.4.5 Étude de la répartition des dates E_1, E_2, \dots, E_n dans l'intervalle AB

Posons $A = 0$ et cherchons la loi de probabilité conjointe des dates E_1, E_2, \dots, E_n et de N nombre d'événements survenus.

La probabilité pour que le premier événement se passe entre t_1 et $t_1 + dt_1$ est : $c \exp(-ct_1) dt_1$.

La probabilité conditionnelle que E_2 arrive entre t_2 et $t_2 + dt_2$ sachant E_1 est : $c \exp(-c(t_2 - t_1)) dt_2$, etc.

La probabilité qu'aucun événement n'arrive après E_n sachant la date de E_n est : $\exp(-c(T - t_n))$; d'où :

$$f(t_1, t_2, \dots, t_n, n) = c^n \exp(-cT)$$

La loi conditionnelle :

$$f(t_1, t_2, \dots, t_n / N = n) = \frac{c^n \exp(-cT)}{\exp(-cT) \frac{(cT)^n}{n!}} = \frac{n!}{T^n}$$

ce qui prouve que les instants t_1, t_2, \dots, t_n constituent un échantillon ordonné de la loi uniforme sur $[0, T]$: en effet, si l'on s'intéresse seulement aux dates et non à leur ordre, il faut diviser par $n!$ qui est le nombre d'ordres possibles.

2.4.6 Le processus (N_t)

D'après ce qui précède, N_t suit pour tout t une loi de Poisson $\mathcal{P}(ct)$. Comme $E(N_t) = ct = V(N_t)$, ce processus n'est pas stationnaire mais il est à accroissements stationnaires et indépendants puisque $\forall h, N_{t+h} - N_t = \mathcal{P}(h)$.

La fonction de covariance de ce processus est facile à obtenir :

- si $s > t$: $C(t, s) = \text{cov}(N_t; N_s) = \text{cov}(N_t; N_t + X) = V(N_t) + \text{cov}(N_t; X)$: or X est une variable indépendante de N_t (accroissements indépendants) donc :
- si $s \geq t$: $C(t; s) = V(N_t) = ct$; et on trouve de même si $t > s$: $C(t, s) = cs$; d'où :

$$C(t; s) = c \inf(t; s).$$

Cette fonction est continue en $t = s$ donc le processus est continu en moyenne quadratique. Cependant, aucune trajectoire n'est continue puisque (N_t) est une fonction aléatoire en escalier (incrément de 1 à chaque événement).

2.5 CONVOLUTION

Un problème courant consiste à trouver la loi de probabilité d'une somme de deux variables indépendantes $Z = X + Y$.

2.5.1 Cas discret

Le théorème des probabilités totales donne la solution du problème :

$$P(Z = z) = \sum_x P(X = x \cap Y = z - x) = \sum_y P(X = z - y \cap Y = y)$$

Lorsque X et Y sont indépendantes, on a :

$$P(Z = z) = \sum_x P(X = x)P(Y = z - x)$$

Sinon, on peut toujours écrire :

$$P(Z = z) = \sum_x P(X = x)P(Y = z - x/X = x)$$

Remarquons que, pour la sommation, x ne prend pas nécessairement toutes les valeurs possibles de X mais uniquement celles compatibles avec l'événement $Z = z$.

■ **Exemple :** Soit X et Y , deux variables de Poisson indépendantes de paramètres λ et μ respectivement :

$$P(X = x) = \exp(-\lambda) \frac{\lambda^x}{x!} \quad P(Y = y) = \exp(-\mu) \frac{\mu^y}{y!}$$

On a donc :

$$P(Z = z) = \sum_{x=0}^{x=z} \exp(-\lambda) \frac{\lambda^x}{x!} \exp(-\mu) \frac{\mu^{z-x}}{(z-x)!}$$

soit en multipliant et divisant par $z!$:

$$\begin{aligned} P(Z = z) &= \frac{\exp(-(\lambda + \mu))}{z!} \sum_{x=0}^{x=z} C_z^x \lambda^x \mu^{z-x} \\ &= \frac{\exp(-(\lambda + \mu))}{z!} (\lambda + \mu)^z \end{aligned}$$

$Z = X + Y$ est donc une variable de Poisson $\mathcal{P}(\lambda + \mu)$. ■

2.5.2 Cas général

La loi de probabilité de $Z = X + Y$ s'obtient grâce au théorème de la mesure image : en effet, la loi de Z n'est autre que la mesure image de $P_{X,Y}$ par l'application de \mathbb{R}^2 dans \mathbb{R} définie par $(x, y) \rightarrow x + y$.

Lorsque X et Y sont indépendants, on a donc le résultat suivant :

THÉORÈME

L La loi de probabilité de la somme Z de deux variables indépendantes est la mesure image de $P_X \otimes P_Y$ par l'application $(x, y) \rightarrow x + y$ de \mathbb{R}^2 dans \mathbb{R} .

Notée $P_X * P_Y = P_Z$ (produit de convolution de deux mesures), elle est telle que pour tout borélien B :

$$P_Z(B) = \int_{\mathbb{R}^2} \mathbb{1}_B(x + y) dP_X(x) \otimes dP_Y(y)$$

On remarquera le caractère symétrique en x et y de la formule précédente.

En particulier, si X et Y admettent des densités, on a :

$$P_Z(B) = \int_{\mathbb{R}^2} \mathbb{1}_B(x + y) f(x) g(y) dx dy$$

Posons $x + y = z$, $x = u$ et appliquons le théorème de Fubini :

$$\begin{aligned} P_Z(B) &= \int_{\mathbb{R}^2} \mathbb{1}_B(z) f(u) g(z - u) du dz \\ &= \int_{\mathbb{R}} \mathbb{1}_B(z) dz \int_{D_X} f(u) g(z - u) du \end{aligned}$$

D'après la définition des variables continues, on en déduit que Z admet pour densité :

$$k(z) = \int_{D_X} f(u) g(z - u) du = \int_{D_Y} g(y) f(z - y) dy$$

les domaines D_X et D_Y étant les ensembles de valeurs de X et de Y respectivement compatibles avec l'événement $Z = z$.

Par intégration, on en déduit :

$$P(Z < z) = K(z) = \int_{D_x} f(x) G(z - x) dx = \int_{D_y} g(y) F(z - y) dy$$

Géométriquement, $K(z)$ représente la mesure du domaine hachuré (fig. 2.26).

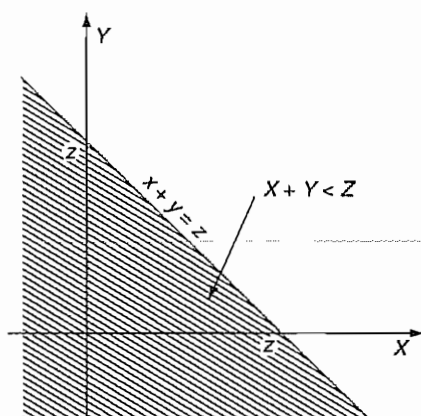


FIGURE 2.26

2.5.3 Applications

2.5.3.1 Somme de lois γ

Soit X de loi γ_r $f(x) = \frac{1}{\Gamma(r)} \exp(-x) x^{r-1}$ et Y de loi γ_s $g(y) = \frac{1}{\Gamma(s)} \exp(-y) y^{s-1}$ indépendante.

$$\begin{aligned} k(z) &= \int_0^z \frac{1}{\Gamma(r)} \exp(-x) x^{r-1} \frac{1}{\Gamma(s)} \exp(-(z-x)) (z-x)^{s-1} dx \\ &= \frac{\exp(-z)}{\Gamma(r)\Gamma(s)} \int_0^z x^{r-1} (z-x)^{s-1} dx \end{aligned}$$

Posons $x = tz$, il vient :

$$k(z) = \frac{\exp(-z)}{\Gamma(r)\Gamma(s)} \int_0^1 t^{r-1} z^{r-1} (z-tz)^{s-1} dt$$

d'où :

$$k(z) = \frac{\exp(-z) z^{r+s-1}}{\Gamma(r)\Gamma(s)} \int_0^1 t^{r-1} (1-t)^{s-1} dt$$

$$k(z) = \exp(-z) z^{r+s-1} c$$

$k(z)$ étant une densité, la constante c vaut nécessairement $\frac{1}{\Gamma(r+s)}$ puisqu'on reconnaît l'expression de la densité d'une loi γ . On en déduit une preuve (probabiliste) de la formule :

$$\int_0^1 t^{r-1} (1-t)^{s-1} dt = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

Donc si X est une γ_r et Y une γ_s indépendante, $X+Y$ est une γ_{r+s} .

2.5.3.2 Somme de lois uniformes sur $[0, 1]$

Soient X et Y deux variables continues uniformes sur $[0, 1]$. La loi de leur somme s'obtient par l'argument géométrique suivant : le couple (X, Y) est uniformément réparti sur le carré unité et l'événement $Z < z$ correspond à la zone hachurée dont il suffit alors de trouver la surface. K et k ont deux déterminations mais sont continues (fig. 2.27).

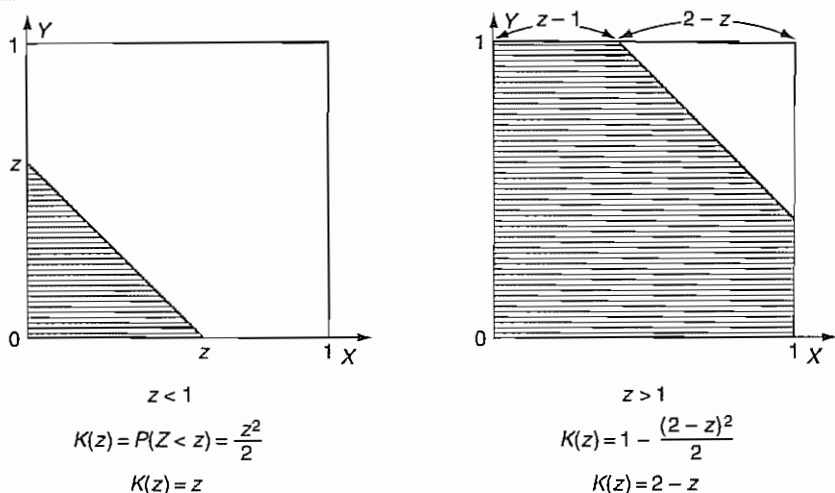


FIGURE 2.27

2.6 FONCTIONS CARACTÉRISTIQUES

2.6.1 Définitions et principales propriétés

2.6.1.1 Définition

La fonction caractéristique d'une variable aléatoire réelle X est la transformée de Fourier de sa loi de probabilité. Elle est notée φ_X et on a :

$$\varphi_X(t) = E[\exp(itX)] = \int_{\mathbb{R}} \exp(itx) dP_X(x)$$

Cette fonction **existe toujours** car P_X est une mesure bornée et $|\exp(itX)| = 1$. Il s'ensuit que la **fonction caractéristique est continue**.

Lorsque X possède une densité :

$$\varphi_X(t) = \int_{\mathbb{R}} \exp(itx) f(x) dx$$

2.6.1.2 Fonction caractéristique d'une forme linéaire

$$\begin{aligned}\varphi_{\lambda X}(t) &= \varphi_X(\lambda t) \\ \varphi_{X+a}(t) &= \exp(it a) \varphi_X(t)\end{aligned}$$

et on en déduit, si X est une variable d'espérance m et d'écart-type σ , en posant $U = (X - m)/\sigma$:

$$\begin{aligned}\varphi_{\frac{X-m}{\sigma}}(t) &= \varphi_U(t) = \exp\left(-\frac{itm}{\sigma}\right) \varphi_X\left(\frac{t}{\sigma}\right) \\ \varphi_X(t) &= \exp(itm) \varphi_U(\sigma t)\end{aligned}$$

2.6.1.3 Convolution

La fonction caractéristique se prête bien aux additions de variables aléatoires indépendantes : la fonction caractéristique d'une somme de variables indépendantes est égale au produit de leurs fonctions caractéristiques :

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

En effet :

$$\varphi_{X_1+X_2}(t) = E[\exp(it(X_1 + X_2))] = E[\exp(itX_1) \exp(itX_2)]$$

si X_1 et X_2 sont indépendantes, il en est de même pour $\exp(itX_1)$ et $\exp(itX_2)$ et l'espérance du produit est alors égal au produit des espérances. Notons au passage qu'il ne s'agit donc pas d'une condition nécessaire et suffisante d'indépendance.

2.6.1.4 Cas d'une distribution symétrique

Supposons la loi de X symétrique par rapport à l'origine. Alors la fonction caractéristique de X est réelle :

$$\varphi_X(-t) = \int_{\mathbb{R}} \exp(-itx) dP_X(x) = \int_{\mathbb{R}} \exp(itx) dP_X(-x)$$

La première intégrale vaut $\overline{\varphi_X(t)}$ et la deuxième est égale à $\varphi_X(t)$ à cause de la symétrie car $dP_X(x) = dP_X(-x)$.

2.6.1.5 Dérivées à l'origine et moments non centrés

Notons tout d'abord que $\varphi_X(0) = 1$ car $\varphi_X(0) = \int_{\mathbb{R}} dP_X(x)$ P_X est une mesure de masse totale égale à 1.

Si les dérivées existent jusqu'à l'ordre k , on a :

$$\varphi_X^{(k)}(0) = i^k E(X^k)$$

En effet, $\varphi_X^{(k)}(t) = \int_{\mathbb{R}} (ix)^k \exp(itx) dP_X(x)$ par dérivation sous le signe somme. En particulier :

$$\varphi_X'(0) = iE(X)$$

$$\varphi_X''(0) = -E(X^2)$$

Si $\varphi_X(t)$ est indéfiniment dérivable, la formule de Mac-Laurin donne :

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} i^k E(X^k)$$

2.6.1.6 Unicité et inversion de la fonction caractéristique

D'après les propriétés des transformées de Fourier, deux variables ayant même fonction caractéristique ont même loi de probabilité : la fonction caractéristique détermine donc de manière unique une distribution de probabilité d'où son nom.

Les formules d'inversion de la transformée de Fourier permettent d'obtenir la loi de X connaissant $\varphi_X(t)$.

THÉORÈME

Si $\int_{\mathbb{R}} |\varphi_X(t)| dt < \infty$ alors X admet une densité $f(x)$ continue et :

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) \exp(-itx) dt$$

Sinon, on a toujours le résultat suivant (admis) :

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

Une fonction quelconque n'est pas nécessairement une fonction de répartition ; de même, pour qu'une fonction $\varphi(t)$ soit une fonction caractéristique elle doit vérifier certaines propriétés. Le théorème suivant, que nous ne démontrerons pas, identifie les fonctions caractéristiques aux fonctions de « type positif ».

THÉORÈME (BOCHNER)

Pour qu'une fonction continue φ soit une fonction caractéristique, il faut et il suffit que pour toute famille finie t_1, t_2, \dots, t_n de réels et pour toute famille finie de complexes z_1, z_2, \dots, z_n on ait :

$$\sum_{i=1}^n \sum_{j=1}^n \varphi(t_i - t_j) z_i \bar{z}_j \geq 0$$

2.6.2 Fonctions caractéristiques des lois usuelles

2.6.2.1 Lois discrètes

- Loi de Bernoulli : $\varphi_X(t) = p \exp(it) + q$ avec $q = 1 - p$.

• Loi binomiale : Comme X est une somme de n variables de Bernoulli indépendantes, on trouve :

$$\varphi_X(t) = (p \exp(it) + q)^n$$

- Loi de Poisson :

$$\varphi_X(t) = \exp(\lambda (\exp(it) - 1))$$

$$\begin{aligned} \text{En effet : } E[\exp(itX)] &= \sum_{x=0}^{\infty} \exp(itx) \exp(-\lambda) \frac{\lambda^x}{x!} = \exp(-\lambda) \sum_{x=0}^{\infty} \left(\frac{\lambda \exp(it)}{x!} \right)^x \\ &= \exp(-\lambda) \exp(\lambda \exp(it)) \end{aligned}$$

2.6.2.2 Lois continues

- Loi uniforme sur $[-a, a]$:

$$\varphi_X(t) = \frac{\sin at}{at}$$

$$\text{En effet : } E[\exp(itX)] = \frac{1}{2a} \int_{-a}^a \exp(itx) dx = \frac{1}{2ait} [\exp(iat) - \exp(-iat)]$$

d'où le résultat avec $\exp(iat) = \cos at + i \sin at$.

- Lois gamma : Si X suit une loi γ_1 , c'est-à-dire une loi exponentielle de paramètre 1, on a :

$$\varphi_{\gamma_1}(t) = \frac{1}{1 - it}$$

$$\text{En effet : } \varphi_{\gamma_1}(t) = \int_0^{\infty} \exp(itx) \exp(-x) dx = \int_0^{\infty} \exp(-(1 - it)x) dx$$

D'où, pour tout n entier :

$$\varphi_{\gamma_n}(t) = \frac{1}{(1 - it)^n}$$

car une γ_n est une somme de n γ_1 indépendantes.

Pour r quelconque, cette formule se généralise et $\varphi_{\gamma_r}(t) = \frac{1}{(1 - it)^r}$.

Remarquons que le calcul formel suivant conduit au résultat :

$$\int_0^{\infty} \exp(itx) \frac{1}{\Gamma(r)} \exp(-x) x^{r-1} dx = \frac{1}{\Gamma(r)} \int_0^{\infty} \exp(-(1 - it)x) x^{r-1} dx$$

en posant $(1 - it)x = u$:

$$= \frac{1}{\Gamma(r)} \int_0^\infty \exp(-u) u^{r-1} \frac{1}{(1 - it)^r} du = \frac{\Gamma(r)}{\Gamma(r)(1 - it)^r} = \frac{1}{(1 - it)^r}$$

Il convient cependant de justifier ce résultat car il s'agit d'une intégrale dans le champ complexe. Nous le laisserons au soin du lecteur.

- Loi de Laplace-Gauss : Si U est la loi $\text{LG}(0 ; 1)$:

$$\varphi_u(t) = \exp(-t^2/2)$$

On peut obtenir ce résultat directement car on sait que $E(U^k) = 0$ si k est impair et

$$E(U^{2k}) = \frac{(2k)!}{2^k k!}.$$

D'après la formule de Mac-Laurin :

$$\begin{aligned} \varphi_u(t) &= \sum_{k=0}^{\infty} \frac{t^{2k}}{2k!} (-1)^k \frac{2k!}{2^k k!} \\ &= \sum_{k=0}^{\infty} \frac{\left(-\frac{t^2}{2}\right)^k}{k!} = \exp(-t^2/2) \end{aligned}$$

Remarquons qu'ici aussi un calcul formel (qui devrait être justifié par une intégration dans le plan complexe) donne le même résultat :

$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \exp(itx) dx &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}[x - it]^2 t^2/2\right) dx \\ &= \exp(-t^2/2) \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}[x - it]^2\right) dx \end{aligned}$$

et l'intégrale vaut 1 car c'est l'intégrale de la densité d'une variable de Gauss imaginaire (!) de moyenne it et de variance 1.

Si X est une $\text{LG}(m ; \sigma)$:

$$\varphi_X(t) = \exp(itm) \exp\left(-\frac{t^2\sigma^2}{2}\right)$$

on en déduit que la somme de deux variables de Gauss indépendantes est encore une variable de Gauss :

$$\begin{aligned} X_1 \text{ LG}(m_1 ; \sigma_1) \quad X_2 \text{ LG}(m_2 ; \sigma_2) \\ \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t) = \exp(it(m_1 + m_2)) \exp\left(-t^2\left(\frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2}\right)\right) \end{aligned}$$

donc $X_1 + X_2$ suit une $\text{LG}(m_1 + m_2 ; \sqrt{\sigma_1^2 + \sigma_2^2})$.

2.6.3 Fonctions génératrices

Il en existe deux formes assez voisines ; elles servent essentiellement à calculer les moments de variables aléatoires et de sommes de variables indépendantes car la fonction génératrice d'un produit de variables indépendantes est égale au produit de leurs fonctions génératrices.

- Pour des variables à valeurs entières positives, on utilisera la forme suivante :

$$g_X(t) = E(t^X) = \sum_{n \geq 0} t^n P(X = n)$$

Par dérivations successives en zéro, on trouve facilement que $g_X^{(n)}(0) = n! P(X = n)$, ce qui prouve que la fonction génératrice détermine la loi de probabilité de X .

Sous réserve d'existence, les dérivées successives en 1 sont égales aux moments factoriels :

$$g_X'(1) = E(X)$$

$$g_X''(1) = E(X(X - 1))$$

$$g_X^{(n)}(1) = E(X(X - 1)(X - 2) \dots (X - n + 1))$$

- Pour des variables quelconques, on appelle fonction génératrice des moments :

$$M_X(t) = E(e^{tX})$$

qui est donc la transformée de Laplace de $-X$. Sous réserve d'existence, on a :

$$E(X^n) = M_X^{(n)}(0)$$

Les fonctions génératrices sont liées à la fonction caractéristique par :

$$g_X(t) = \varphi_X(-i \ln(t))$$

$$M_X(t) = \varphi_X(-it)$$

2.7 CONVERGENCES DES SUITES DE VARIABLES ALÉATOIRES

2.7.1 Les différents types de convergence

Une suite (X_n) de variables aléatoires étant une suite de fonctions de Ω dans \mathbb{R} , il existe diverses façons de définir la convergence de (X_n) dont certaines jouent un grand rôle en calcul des probabilités.

2.7.1.1 La convergence en probabilité

DÉFINITION

L La suite (X_n) converge en probabilité vers la constante a si, $\forall \varepsilon$ et η (arbitrairement petits), il existe n_0 tel que $n > n_0$ entraîne :

$$P(|X_n - a| > \varepsilon) < \eta$$

On note alors $(X_n) \xrightarrow{P} a$.

On définit alors la convergence en probabilité vers une variable aléatoire X comme la convergence vers 0 de la suite $X_n - X$.

Lorsque $E(X_n) \rightarrow a$, il suffit de montrer que $V(X_n) \rightarrow 0$ pour établir la convergence en probabilité de X_n vers a . En effet, d'après l'inégalité de Bienaymé-Tchebycheff :

$$P(|X_n - E(X_n)| > \varepsilon) < \frac{V(X_n)}{\varepsilon^2}$$

On en déduit donc sans difficulté que $X_n - E(X_n) \xrightarrow{P} 0$, ce qui établit le résultat.

2.7.1.2 La convergence presque sûre ou convergence forte

Définissons d'abord l'égalité presque sûre de deux variables aléatoires :

DÉFINITION

L X et Y sont égales presque sûrement si $P(\{\omega | X(\omega) \neq Y(\omega)\}) = 0$.

C'est l'égalité presque partout des fonctions mesurables. On définit donc ainsi des classes de variables aléatoires presque sûrement égales.

La convergence presque sûre se définit alors par :

DÉFINITION

L La suite (X_n) converge presque sûrement vers X si :

$$P(\{\omega | \lim_{n \rightarrow \infty} X_n(\omega) \neq X(\omega)\}) = 0$$

L et on note $X_n \xrightarrow{ps} X$.

En d'autres termes, l'ensemble des points de divergence est de probabilité nulle. Remarquons que la limite de (X_n) n'est pas unique mais que deux limites sont presque sûrement égales.

Il est immédiat de montrer que la convergence presque sûre implique la convergence en probabilité.

2.7.1.3 La convergence en moyenne d'ordre p

Si $E[(X_n - X)^p]$ existe, on a :

DÉFINITION

L $(X_n) \rightarrow X$ en moyenne d'ordre p si $E[|X_n - X|^p] \rightarrow 0$.

La plus utilisée est la convergence en moyenne quadratique si $p = 2$.

La convergence en moyenne d'ordre p implique la convergence en probabilité.

2.7.1.4 La convergence en loi

Bien que la plus faible, elle est très utilisée en pratique car elle permet d'approximer la fonction de répartition de X_n par celle de X .

DÉFINITION

L La suite (X_n) converge en loi vers la variable X de fonction de répartition F si, en tout point de continuité de F , la suite (F_n) des fonctions de répartition des X_n converge vers F . On note $X_n \xrightarrow{f} X$.

Un théorème dû à Polya établit que si F est continue alors la convergence est uniforme.

Pour des variables discrètes, la convergence en loi vers une variable discrète s'exprime par $P(X_n = x) \rightarrow P(X = x)$.

C'est ainsi qu'on a établi la convergence de la loi binomiale vers la loi de Poisson.

Une suite de variables discrètes peut cependant converger en loi vers une variable continue (voir plus loin).

On montre également que, si (X_n) est une suite de variables de densités f_n et X une variable de densité f , alors :

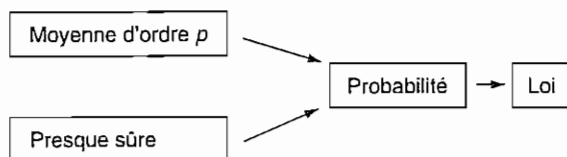
$$X_n \xrightarrow{f} X \Rightarrow f_n(x) \rightarrow f(x) \quad \forall x$$

La convergence en loi est intimement liée à la convergence des fonctions caractéristiques comme le précise le résultat fondamental suivant, que nous énoncerons sans démonstration :

THÉORÈME (LEVY-CRAMER-DUGUÉ)

L Si $X_n \xrightarrow{f} X$ alors $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$ uniformément dans tout intervalle fini $[-u, u]$. Si la suite des fonctions caractéristiques $\varphi_{X_n}(t)$ converge vers une fonction φ dont la partie réelle est continue à l'origine, alors φ est une fonction caractéristique et la suite X_n converge en loi vers une variable aléatoire X dont φ est la fonction caractéristique.

La convergence en probabilité entraîne la convergence en loi et on a, pour résumer, la hiérarchie suivante des convergences :



2.7.2 Convergence en loi de la binomiale vers la loi de Laplace-Gauss (théorème de De Moivre-Laplace)

THÉORÈME

L X_n étant une suite de variables binomiales $\mathcal{B}(n; p)$, alors $\frac{X_n - np}{\sqrt{npq}} \xrightarrow{f} \text{LG}(0; 1)$ en notant $q = 1 - p$.

■ **Démonstration :** La fonction caractéristique de X_n vaut $(p \exp(it) + 1 - p)^n$ donc celle

de $\frac{X_n - np}{\sqrt{npq}}$ vaut :

$$\varphi(t) = \left(p \exp\left(\frac{it}{\sqrt{npq}}\right) + 1 - p \right)^n \exp\left(-\frac{itnp}{\sqrt{npq}}\right)$$

$$\ln \varphi = n \ln \left(p \left(\exp\left(\frac{it}{\sqrt{npq}}\right) - 1 \right) \right) - \frac{itnp}{\sqrt{npq}}$$

Développons au deuxième ordre l'exponentielle ; il vient :

$$\ln \varphi \approx n \ln \left(1 + p \left(\frac{it}{\sqrt{npq}} - \frac{t^2}{2npq} \right) \right) - \frac{itnp}{\sqrt{npq}}$$

puis le logarithme :

$$\ln \varphi \approx n \left[\frac{pit}{\sqrt{npq}} - \frac{pt^2}{2npq} + \frac{p^2 t^2}{2npq} \right] - \frac{itnp}{\sqrt{npq}}$$

soit :

$$\ln \varphi \approx -\frac{t^2}{2q} + \frac{pt^2}{2q} = \frac{t^2}{2q} (p - 1) = -\frac{t^2}{2}$$

car $p = 1 - q$.

$\varphi(t) \rightarrow \exp(-t^2/2)$ qui est la fonction caractéristique de la loi normale centrée-réduite. ■

Application : Lorsque n est assez grand, on peut donc approximer la loi binomiale par la loi de Gauss. On donne généralement comme condition np et $nq > 5$.

Il convient cependant d'effectuer ce que l'on appelle la correction de continuité : la convergence de la loi binomiale vers la loi de Gauss se traduit par le fait que les extrémités des bâtons du diagramme de la binomiale $\mathcal{B}(n; p)$ sont voisines de la courbe de densité de la loi LG $(np; \sqrt{npq})$.

On obtient donc une valeur approchée de $P(X = x)$ par la surface sous la courbe de densité comprise entre les droites d'abscisse $x - \frac{1}{2}$ et $x + \frac{1}{2}$ (fig. 2.28).

$$P(X = x) \approx P \left(\frac{x - \frac{1}{2} - np}{\sqrt{npq}} < U < \frac{x + \frac{1}{2} - np}{\sqrt{npq}} \right)$$

On aura alors :

$$P(X \leq x) \approx P \left(U < \frac{x + \frac{1}{2} - np}{\sqrt{npq}} \right)$$

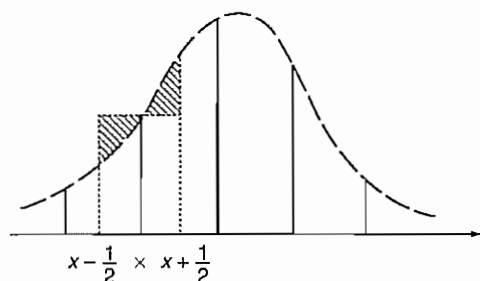


FIGURE 2.28

■ **Exemple :** $X \mathcal{B}(40; 0.3)$ $np = 12$; $npq = 8.4$. La valeur exacte pour $P(X = 11)$ est 0.1319. La formule d'approximation avec une loi LG(12 ; $\sqrt{8.4}$) donne :

$$P\left(\frac{10.5 - 12}{\sqrt{8.4}} < U < \frac{11.5 - 12}{\sqrt{8.4}}\right)$$

soit : $P(-0.52 < U < -0.17) = P(0.17 < U < 0.52) = 0.6895 - 0.5675 = 0.122$

Soit une erreur de moins de 1 %.

Quant à $P(X \leq 11)$ qui vaut exactement 0.4406, l'approximation normale fournit $1 - P(U < 0.17)$ soit 0.4325. En l'absence de correction de continuité, on aurait trouvé

$P\left(U < \frac{11-12}{\sqrt{8.4}}\right) = P(U < -0.35) = 1 - P(U < 0.35) = 0.3632$, ce qui est très imprécis. ■

2.7.3 Convergence de la loi de Poisson vers la loi de Gauss

THÉORÈME

Soit (X_λ) une famille de variables $\mathcal{P}(\lambda)$ alors si $\lambda \rightarrow \infty$, $\frac{X - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{L}} \text{LG}(0; 1)$.

■ Démonstration

$$\varphi_X(t) = \exp(\lambda)(\exp(it - 1))$$

d'où :

$$\varphi_{\frac{X-\lambda}{\sqrt{\lambda}}}(t) = \exp\left(\lambda\left(\exp\left(\frac{it}{\sqrt{\lambda}}\right) - 1\right)\right) \exp\left(-\frac{it\lambda}{\sqrt{\lambda}}\right)$$

$$= \exp\left(\lambda \exp\left(\frac{it}{\sqrt{\lambda}}\right) - \lambda - it\sqrt{\lambda}\right)$$

comme :

$$\exp\left(\frac{it}{\sqrt{\lambda}}\right) \approx 1 + \frac{it}{\sqrt{\lambda}} - \frac{t^2}{2\lambda}$$

il vient :

$$\varphi_{X-\lambda}(t) \approx \exp\left(\lambda + it\sqrt{\lambda} - \frac{t^2}{2} - \lambda - it\sqrt{\lambda}\right) = \exp\left(-\frac{t^2}{2}\right)$$

La figure 2.29 illustre l'approximation de la loi de Poisson $\mathcal{P}(\lambda)$ par la loi de Gauss de même espérance λ et de même écart-type $\sqrt{\lambda}$.

L'approximation est très satisfaisante pour $\lambda > 18$. On trouvera en annexe d'autres formules d'approximation plus précises. On a, ici encore, intérêt à effectuer la correction de continuité.

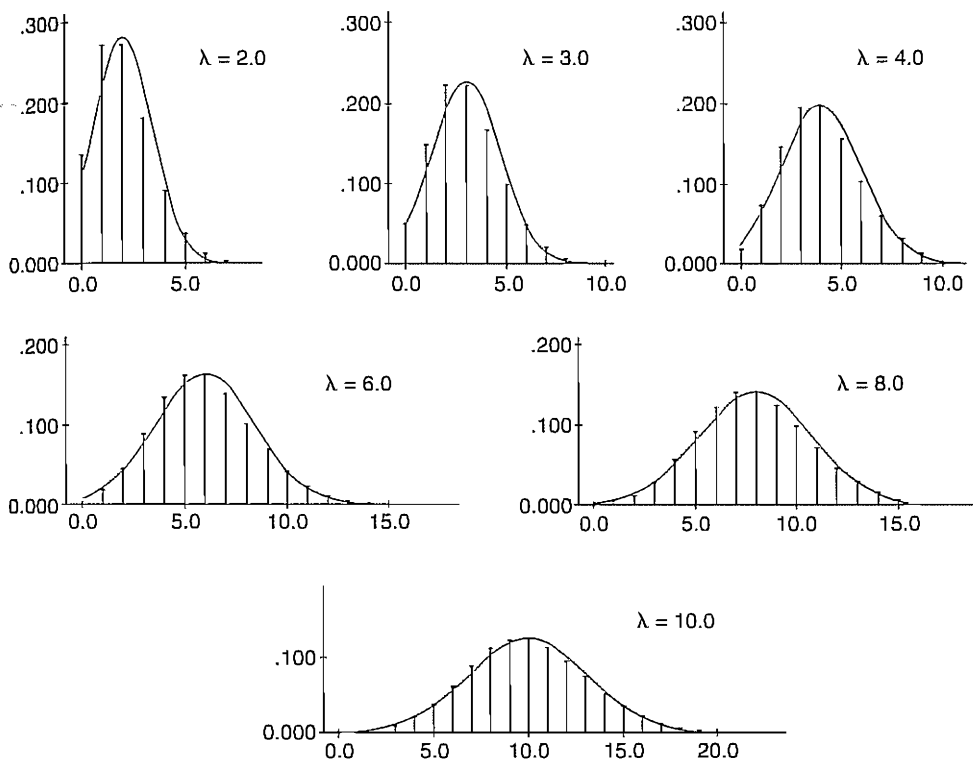


FIGURE 2.29

2.7.4 Le théorème central-limite

L'étude de sommes de variables indépendantes et de même loi joue un rôle capital en statistique.

Le théorème suivant connu sous le nom de **théorème central-limite** (il vaudrait mieux dire théorème de la limite centrée) établit la convergence vers la loi de Gauss sous des hypothèses peu contraignantes.

THÉORÈME

Soit (X_n) une suite de variables aléatoires indépendantes de même loi d'espérance μ et d'écart-type σ . Alors :

$$\frac{1}{\sqrt{n}} \left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma} \right) \xrightarrow{\mathcal{L}} \text{LG}(0; 1).$$

■ Démonstration

$$\frac{1}{\sqrt{n}} \left(\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma} \right) = \sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$$

Soit $\varphi_X(t)$ la fonction caractéristique de X ; la fonction caractéristique de $\sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$ est donc égale à $[\varphi_{\frac{X-\mu}{\sigma\sqrt{n}}}(t)]^n$. Or $\frac{X - \mu}{\sigma\sqrt{n}}$ est une variable d'espérance nulle et de variance $1/n$.

Le développement en série de la fonction caractéristique de $\frac{X - \mu}{\sigma\sqrt{n}}$ commence par $1 - \frac{t^2}{2n}$, les termes suivants sont des infiniments petits d'ordre $1/n^2$.

Donc, en élevant à la puissance n , la fonction caractéristique de $\sum_{i=1}^n \frac{X_i - \mu}{\sigma\sqrt{n}}$ est équivalente à $\left(1 - \frac{t^2}{2n}\right)^n$ et tend si $n \rightarrow \infty$ vers $\exp\left(-\frac{t^2}{2}\right)$ selon un résultat classique. ■

On remarque que, si les variables X_i sont des variables de Bernoulli, on retrouve comme cas particulier la convergence de la loi binomiale vers la loi de Gauss.

On peut démontrer un théorème encore plus général dû à Lindeberg :

THÉORÈME

Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes pas forcément de même loi et d'espérance m_i et de variance σ_i^2 . Soit $S_n^2 = \sum_{i=1}^n \sigma_i^2$ et $F_i(x)$ la fonction de répartition de $(X_i - m_i)$.

Si la condition suivante est réalisée :

$$\lim_{n \rightarrow \infty} \left[\frac{1}{S_n^2} \sum_{i=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_i(x) \right] = 0$$

alors :

$$\frac{\sum_{i=1}^n (X_i - m_i)}{S_n} \xrightarrow{\mathcal{L}} U \in \text{LG}(0; 1)$$

La condition de Lindeberg exprime que les variables $\frac{X_i - m_i}{S_n}$ sont « uniformément petites » avec une grande probabilité. Le résultat veut dire qu'à force d'ajouter de telles variables, on finit par obtenir une loi de Gauss.

Ce phénomène est souvent exprimé de la manière suivante : si une variable est la résultante d'un grand nombre de causes, petites, à effet additif, cette variable suit une loi de Gauss. On peut y voir la justification de l'emploi abondant et souvent abusif de la loi de Laplace-Gauss comme modèle.

Pour terminer, notons que l'existence des moments $E(X)$ et $V(X)$ est indispensable. La loi de Cauchy de densité $\frac{1}{\pi(1+x^2)}$ sur \mathbb{R} n'a aucun moment et fournit un contre-exemple classique : on montre que $\frac{X_1 + X_2 + \dots + X_n}{n}$ a même loi que X quel que soit n .