Contenu

1 Espérance d'une variable aléatoire

1.1 Variable discrète

Définition 1 *L'espérance (ou la moyenne) d'une variable aléatoire discrète X est définie par*

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

si la série est absolument convergente (c'est-à-dire si la série $\sum_i |x_i| p(x_i)$ est convergente).

Une notation courante pour l'espérance est $E(X) = \mu$ (ou μ_X).

Exemple 1 *On lance une pièce de monnaie équilibrée une fois. Si X est le nombre de piles obtenu, la distribution (loi) de X est donnée par*

X	0	1	*Total*
$p(x)$	$\frac{1}{2}$	$\frac{1}{2}$	1

La moyenne de X est alors

$$E(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

On a donc en moyenne "pile" la moitié du temps (si on réfléchit en termes de fréquences).

Remarque 1 *Le concept d'espérance a pour origine les jeux de hasard. On dit qu'un jeu est équitable si l'espérance du gain (aléatoire) est nulle. Un observateur attentif (exemple : le Chevalier de Méré) peut décider si un jeu est équitable (penser au jeu de la roulette dans les casinos, par exemple, ou aux primes d'assurances). Aujourd'hui, à défaut d'un calcul exact de l'espérance (qui peut être impossible), on effectue une simulation informatique pour en obtenir une estimation.*

Exemple 2 *Paradoxe de Saint-Petersbourg. On vous propose le jeu suivant. Vous lancez une pièce de monnaie équilibrée jusqu'à obtention de Pile. On vous offre alors 2^X dollars, o**ù** X est le nombre de lancers.*

1. *Combien êtes vous prêts à payer pour que le jeu soit équitable ?*
2. *On change les règles comme suit : après 20 essais infructueux, la récompense sera fixée à 2^{20} dollars (un peu plus d'un million : $2^{20} = 1048\,576$), quel que soit le nombre de lancers subséquents jusqu'à obtention de Pile. Combien faut-il payer pour que le jeu soit équitable ? Note. En science économiques, on postule que chacun d'entre nous a une fonction utilité $u(x)$, c-à-d que si nous gagnons x dollars, alors nous avons une quantité d'utilité $u(x)$ (mesure de notre plaisir, joie, récompense). Cette question signifie que notre joie est pleinement acquise une fois qu'on a gagné un million de dollars (gagner plus que ça ne nous apporte pas plus de satisfaction). L'espérance calculée donne la valeur monétaire que vous attribuez au Paradoxe, même s'il est susceptible de vous rapporter un million. Faites une simulation (ou lancez une pièce) !*

Exemple 3 *Soit la variable X concentrée sur les entiers $n = \pm 2^k, k = 1, 2, \cdots$ et dont la fonction de masse est*

$$p\left(\pm 2^k\right) = \frac{1}{2^{k+1}}, \qquad k = 1, 2, \cdots$$

(0 ailleurs). Calculer l'espérance de X.

Commençons par vérifier que $\sum_{k=1}^{\infty} p\left(\pm 2^k\right) = 1$:

$$\sum_{k=1}^{\infty} p\left(\pm 2^k\right) = \sum_{k=1}^{\infty} 2\left(\frac{1}{2^{k+1}}\right) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1$$

On a utilisé le fait que la série géométrique

$$\sum_{k=1}^{\infty} x^k = \frac{1}{1-x} - 1 = \frac{x}{1-x}$$

converge pour $|x| < 1$. Maintenant,

$$E(X) = \sum_{k=1}^{\infty} \frac{-2^k}{2^{k+1}} + \sum_{k=1}^{\infty} \frac{2^k}{2^{k+1}} = \sum_{k=1}^{\infty} \frac{-1}{2} + \sum_{k=1}^{\infty} \frac{1}{2} = -\infty + \infty$$

Cette série diverge (pas d'espérance). Noter qu'on aurait pu directement utiliser le fait qu'elle ne converge pas absolument :

$$E(|X|) = \sum_{k=1}^{\infty} \frac{2 \times 2^k}{2^{k+1}} = \sum_{k=1}^{\infty} 1 = +\infty$$

Remarque 2 *Si X est concentrée sur les entiers ≥ 0, alors*

$$E(X) = \sum_{k \geq 0} P(X > k)$$

Cette propriété peut être très utile. En voici une démonstration. Posons $p(k) = P(X = k)$. On a

$$\begin{array}{rllllll} E(X) = & 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + 3 \cdot p(3) + \ldots \\ = & p(1) & +p(2) & +p(3) & +p(4) & +\cdots \\ & & +p(2) & +p(3) & +p(4) & +\cdots \\ & & & +p(3) & +p(4) & +\cdots \\ & & & & +p(4) & +\cdots \\ & & & & & +\vdots \\ = & P(X > 0) + P(X > 1) + P(X > 2) + \cdots \end{array}$$

Exemple 4 *Un joueur paye la somme $S = 5\$$ pour participer à un jeu. Il lance un dé régulier. Si la face 1 ou 2 apparaît, le jeu s'arrête (et le joueur perd sa mise : son gain est Gain $= -5\$$). Sinon (si l'une des faces 2, 4, 5 ou 6 apparaît), il lance une pièce régulière. S'il obtient Pile, il gagne $1\$$ et le jeu s'arrête (le gain de notre joueur sera alors Gain $= -5 + 1 = -4\$$). Mais si c'est Face qui sort, il tire deux cartes (sans remise) d'un jeu de 52 cartes. Si aucun as n'est tiré, il reçoit $2\$$ et le jeu s'arrête (Gain $= -5 + 2 = -3\$$). Si un as est tiré, il reçoit $10\$$ (Gain $= -5 + 10 = 5\$$). Enfin, si deux as sont tirés, il reçoit $50\$$ (Gain $= -5 + 50 = 45\$$). Trouvons l'espérance du gain.*

Gain	-5	-4	-3	5	45	*Total*
Pr	$\frac{2}{6} = \frac{1}{3}$	$\frac{4}{6} \times \frac{1}{2} = \frac{1}{3}$	$\frac{4}{6} \times \frac{1}{2} \frac{\binom{4}{0}\binom{48}{2}}{\binom{52}{2}} = \frac{188}{663}$	$\frac{4}{6} \times \frac{1}{2} \frac{\binom{4}{1}\binom{48}{1}}{\binom{52}{2}} = \frac{32}{663}$	$\frac{4}{6} \times \frac{1}{2} \frac{\binom{4}{2}\binom{48}{0}}{\binom{52}{2}} = \frac{1}{663}$	1

L'espérance du gain est alors

$$E(Gain) = -5\left(\frac{1}{3}\right) + (-4)\left(\frac{1}{3}\right) + (-3)\left(\frac{188}{663}\right) + 5\left(\frac{32}{663}\right) + 45\left(\frac{1}{663}\right) = -\frac{2348}{663} = -3.54\$$$

Le jeu n'est donc pas équitable. Faites une simulation de ce jeu.

1.1.1 Exercice

Exercice 1 *Soit la variable X concentrée sur les entiers $n = 1, 2, \cdots$ et dont la fonction de masse est*

$$p(n) = \frac{1}{n(n+1)}$$

On vérifie facilement que c'est bien une fonction de masse. Montrer que $E(X) = +\infty$.

1.2 Variable continue

L'espérance d'une variable aléatoire continue X est définie par

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

si l'intégrale est absolument convergente (c'est-à-dire si l'intégrale $\int_{\mathbb{R}} |x| f(x)\, dx$ est convergente). En particulier, si X est positive, on montre par intégration par parties que

$$E(X) = \int_0^{\infty} P(X > x)\, dx$$

Exemple 5 *Soit*

$$f(x) = \begin{cases} 1/\left(2x^2\right) & si \quad x \in \left]-\infty; -1\right] \cup \left[1; +\infty\right[\\ 0 & si \quad x \in \left]-1; 1\right[\end{cases}$$

Alors

$$\begin{aligned} E(X) &= \int_{-\infty}^{-1} \frac{x}{x^2} dx + \int_1^{+\infty} \frac{x}{x^2} dx \\ &= -\infty + \infty \text{ (nondéfinie)} \end{aligned}$$

Exemple 6 *Soit $f(x) = 2(1-x)$ si $x \in [0; 1]$ (0 ailleurs). Alors*

$$E(X) = 2 \int_0^1 x(1-x) = \frac{1}{3}$$

Noter que, puisque $X \geq 0$, on a $P(X > x) = \int_x^1 2(1-u)\, du = (x-1)^2$ (pour $x \in [0; 1]$), et donc

$$E(X) = \int_0^{\infty} P(X > x)\, dx = \int_0^1 (x-1)^2\, dx = \int_0^1 x^2 dx = \frac{1}{3}$$

2 Propriétés de l'espérance

Théorème 1 *Espérance d'une fonction d'une variable aléatoire $Y = g(X)$. Si $E(Y)$ existe, alors*

$$E(Y) = \begin{cases} \sum_i g(x_i)\, p(x_i) & si\ X \text{ est discrète} \\ \\ \int_{\mathbb{R}} g(x)\, f_X(x)\, dx & si\ X \text{ est continue} \end{cases}$$

On généralise à une fonction de plusieurs variables.

Théorème 2 *Soit $Y = g(X_1, \ldots, X_n)$. Si $E(Y)$ existe, alors*

$$E(Y) = \begin{cases} \sum_i g(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n) & si\ X_i \text{ sont discrètes} \\ \\ \int_{\mathbb{R}^n} g(x_1, \ldots, x_n)\, f_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n & si\ X_i \text{ sont continues} \end{cases}$$

Corollaire 1 *Si X et Y sont indépendantes, alors*

$$E\left(g\left(X\right)h\left(Y\right)\right) = E\left(g\left(X\right)\right)E\left(h\left(Y\right)\right)$$

pour toutes fonctions g et h. En particulier,

$$E\left(XY\right) = E\left(X\right)E\left(Y\right)$$

Exemple 7 *Soit*

$$f_{(X,Y)}\left(x,y\right) = \begin{cases} x+y & si \quad (x,y) \in [0;1]^2 \\ \\ 0 & ailleurs \end{cases}$$

alors

$$E\left(XY^2\right) = \int_{\mathbb{R}^2} xy^2 f_{(X,Y)}\left(x,y\right)dxdy = \int_0^1 \int_0^1 xy^2\left(x+y\right)dxdy \overset{calculs}{=} \frac{17}{72}$$

Exemple 8 *Le corollaire est faux si X et Y ne sont pas indépendantes. Prenons par exemple $X \sim Bernoulli\left(p\right)$ $(0 < p < 1)$ et $Y = 1 - X$. On a*

$$E\left(XY\right) = E\left(X\left(1-X\right)\right) = 0 \times \left(1-p\right) + 0 \times p = 0$$

alors que

$$E\left(X\right)E\left(Y\right) = p\left(1-p\right) \neq 0$$

Théorème 3 *Espérance d'une combinaison linéaire de variables aléatoires. Soit $Y = a + \sum_{i=1}^n b_i X_i$. Si $E\left(X_i\right) = \mu_i$ est défini pour tout i, alors*

$$E\left(Y\right) = a + \sum_{i=1}^n b_i \mu_i$$

Théorème 4 *Propriété de monotonicité. Si la variable aléatoire X est positive $(X \geq 0)$, alors $E\left(X\right) \geq 0$ (peut être infinie).*
Preuve. *Évident.* ■

Corollaire 2 *Soit X et Y deux variables aléatoires telles que $X \leq Y$. Alors $E\left(X\right) \leq E\left(Y\right)$.*
Preuve. *Il suffit d'appliquer le théorème à la variable $D = Y - X$. Attention, que peut-on dire si $E\left(X\right) = +\infty$?* ■

2.1 Variance et écart-type

La variance de X est définie par

$$VAR\left(X\right) = \sigma_X^2 = E\left(\left(X - E\left(X\right)\right)^2\right) = E\left(\left(X - \mu_X\right)^2\right)$$

Noter que cette quantité peut être infinie (on considérera qu'elle est quand même définie, contrairement à l'espérance où on peut rencontrer la situation non définie $-\infty + \infty$). L'écart-type de X est la racine carrée de la variance :

$$\sigma_X = \sqrt{VAR\left(X\right)}$$

C'est une mesure de la variation de la variable X par rapport à son espérance (sa moyenne). On montre facilement que

$$VAR\left(X\right) = E\left(X^2\right) - \mu_X^2$$

Une autre mesure de variation parfois utilisée est l'écart moyen : $E\left(\left|X - \mu_X\right|\right)$.

On a les résultats immédiats suivants.

Proposition 1 *1.*

$$\boxed{VAR\,(aX+b) = a^2 VAR\,(X)}$$

$$\sigma_{aX+b} = |a|\,\sigma_X$$

pour toutes constantes a et b.

2. Si X et Y sont indépendantes, alors

$$\boxed{VAR\,(X+Y) = VAR\,(X) + VAR\,(Y)}$$

Remarque 3 *En général,*

$$Var\,(X+Y) \neq Var\,(X) + Var\,(Y)$$

En effet, il faut ajouter un terme de correction (voir covariance plus loin).

Exemple 9 *On vous propose le jeu suivant : dans une urne, il y a une proportion p de boules blanches et une proportion $q = 1-p$ de boules rouges ($0 < p < 1$). Vous avez le choix entre deux stratégies : 1) vous tirez au hasard une boule de l'urne. Si elle est blanche, vous gagnez N\$ (sinon 0\$). 2) Vous tirez au hasard et avec remise n boules. À chaque fois que vous tirez une blanche, on vous donne (N/n) \$. Laquelle des deux stratégies est préférable ?*

Soit X le gain pour chacune des deux stratégie. Cherchons la moyenne et la variance.

Stratégie 1. Clairement,

$$\begin{aligned} E\,(X) &= 0q + Np = Np\$ \\ E\left(X^2\right) &= 0^2 q + N^2 p = N^2 p \\ Var\,(X) &= N^2 p - (Np)^2 = N^2 p\,(1-p) \end{aligned}$$

Stratégie 2. Puisque le tirage se fait avec remise, les tirages successifs sont indépendants et $X = X_1 + \cdots + X_n$ avec

$$\begin{aligned} E\,(X_i) = (N/n)\,p\$ &\Rightarrow E\,(X) = n\,(N/n)\,p = Np\$ \\ Var\,(X_i) = (N/n)^2\,p\,(1-p) &\Rightarrow Var\,(X) = n\,(N/n)^2\,p\,(1-p) = N^2 p\,(1-p)\,/n \end{aligned}$$

*Nous constatons que l'espérance du gain est la même dans les deux stratégies. Ceci peut être le droit d'entrée dans le jeu. Mais la variance dans la première stratégie est divisée par n dans la seconde stratégie. La variance est appelée **risque** dans le contexte financier et la seconde stratégie s'appelle **diversification** (quand vous investissez dans le monde des finances, il est moins risqué de diversifier votre portefeuille).*

AN. Si $N = 100\$$, $n = 10$ et $p = 1/10$, alors $E\,(X) = 100/10 = 10\$$ et $Var\,(X) = 900$ (stratégie 1), $Var\,(X) = 90$ (stratégie 2).

2.1.1 Exercice

Exercice 2 *Refaire l'exemple précédent si dans la stratégie 2 vous tirez les n boules sans remise (par exemple simultanément).*

2.1.2 Biais et erreur quadratique moyenne

Supposons que $E\left(X^2\right)$ existe et considérons la fonction d'une variable réelle $h\,(y) = E\left((X-y)^2\right)$. Alors, cette fonction atteint son minimum pour $y = \mu = E\,(X)$ et $h\,(\mu) = Var\,(X)$.

En effet, on a

$$\begin{aligned} h(y) &= E\left((X-y)^2\right) = E\left(((X-\mu)+(\mu-y))^2\right) \\ &= E\left((X-\mu)^2 + 2(X-\mu)(\mu-y) + (\mu-y)^2\right) \\ &= E\left((X-\mu)^2\right) + E\left(2(X-\mu)(\mu-y)\right) + E\left((\mu-y)^2\right) \\ &= Var(X) + 2(\mu-y)E(X-\mu) + (\mu-y)^2 \\ &= Var(X) + (\mu-y)^2. \end{aligned}$$

Les deux dernières quantités étant positives, le minimum est atteint pour $y=\mu$.

Biais et erreur quadratique moyenne (EQM). Supposons qu'on veuille effectuer la mesure d'une certaine quantité physique (par exemple la longueur ou le poids d'une certaine constante c). À cause le l'imprécision de l'instrument de mesure, la mesure est entachée d'une erreur aléatoire ε de moyenne (espérance) nulle. Si de plus notre instrument est mal reglé, on fait une erreur systématique constante β. Au final, on se retrouve avec une variable aléatoire $X = c+\beta+\varepsilon$, avec c et β constantes et $E(\varepsilon)=0$, $VAR(X) = VAR(\varepsilon) = \sigma^2$. Alors $E(X) = c+\beta$. La constante β est le biais. L'EQM est

$$EQM(X) = E\left((X-c)^2\right) = \beta^2+\sigma^2$$

Exercices

Exercice 3 *Pourquoi l'hypothèse de l'existence de* $E\left(X^2\right)$ *est-elle nécéssaire ?*

Exercice 4 *Refaire la démonstration en dérivant la fonction* $h(y)$.

2.1.3 Moment d'ordre k

Définition 2 *Moment d'ordre* k *(k entier non négatif).* $\text{seq}_{k=1..10}\left(\left(3^{2k+1}+2^{2k+2}\right) \bmod 7\right) = 1,6,0,1,6,0,1,6,0,1$

$$\mu_k = E\left(X^k\right)$$

(sous réserve de convergence absolue). En particulier, $\mu_1 = \mu$ *est l'espérance de* X*. Le moment centré d'ordre* k *est*

$$\mu'_k = E\left((X-\mu)^k\right)$$

En particulier, $\mu'_2 = VAR(X)$*. Le moment factoriel d'ordre* k *($k \geq 1$) est*

$$E[X(X-1)\dots(X-k+1)]$$

Théorème 5 *Soit* X *une variable aléatoire et* r *un entier positif. Si* $\mu_r = E(X^r)$ *existe. alors* $\mu_k = E\left(X^k\right)$ *existe pour tout* $0 \leq k \leq r$.

Preuve. *On montre facilement que* $E\left(|X|^k\right) \leq 1 + E\left(|X|^r\right)$ *(exercice).* ■

2.1.4 Quantiles, etc.

Quantiles Déjà vus.

Paramètre de position Soit donné une fonction de répartition $F_0(x)$ avec densité $f_0(x)$. Un nombre a est un paramètre de position pour la distribution de la variable aléatoire X si la fonction de répartition de X est de la forme

$$F(x;a) = F_0(x-a)$$
$$f(x;a) = f_0(x-a)$$

Par exemple,

$$f(x;a) = \begin{cases} e^{-(x-a)} & si \quad x \geq a \\ 0 & si \quad x < a \end{cases}$$

Ici, $f_0(x)$ est la densité d'une variable exponentielle standard. Un autre exemple est la loi normale $N(\mu, \sigma^2)$. L'espérance μ est un paramètre de position.

Paramètre d'echelle Avec les notations ci-dessus, $b > 0$ est un paramètre d'échelle si

$$F(x;b) = F_0(x/b)$$
$$f(x;b) = \frac{1}{b} f_0\left(\frac{x}{b}\right)$$

Combinant les deux (position et échelle), on a

$$F(x;a,b) = F_0\left(\frac{x-a}{b}\right)$$
$$f(x;a,b) = \frac{1}{b} f_0\left(\frac{x-a}{b}\right)$$

Exemple 10 *Loi normale ; loi de Cauchy.*

2.1.5 Exemples : lois discrètes usuelles

- Variable binomiale $X \sim Bin(n,p)$. $E(X) = np$ et $Var(X) = np(1-p)$.
- Variable de Poisson $X \sim Poi(\lambda)$. $E(X) = Var(X) = \lambda$.
- Variable géométrique $X \sim Geo(p)$. $E(X) = 1/p$ et $Var(X) = \frac{1-p}{p^2}$.
- Variable négative binomiale $X \sim NB(r,p)$. $E(X) = (1-p)r/p^2$.
- Variable géométrique : $P(X=k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$. $E(X) = n\frac{r}{b+r}$ et $Var(X) = -n\frac{r}{b+r}\frac{b}{b+r}\frac{b+r-n}{b+r-1}$.

2.1.6 Exemples : lois continues usuelles

- $X \sim U[a;b]$. $E(X) = (b+a)/2$ et $Var(X) = (b-a)^2/12$.
- $X \sim Exp(\lambda)$. $E(X) = 1/\lambda$ et $Var(X) = 1/\lambda^2$.
- $X \sim Gamma(\alpha, \lambda)$. $E(X) = \alpha/\lambda$ et $Var(X) = \alpha/\lambda^2$
- $X \sim N(\mu, \sigma^2)$. $E(X) = \mu$ et $Var(X) = \sigma^2$. En particulier, $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
- $X \sim Cauchy$: $f(x) = \frac{1}{\pi(1+x^2)}$. $E(X)$ non définie.

3 Inégalités importantes

Notons d'abord que si $X \geq 0$, alors $E(X) \geq 0$ (ou est infinie). Plus généralement, si $Y \leq X$, alors $E(Y) \leq E(X)$ (en supposant l'existence). Il suffit, pour le voir, de considérer la différence $D = X - Y \geq 0$.

Théorème 6 *Inégalité de Markov. Soit $X \geq 0$ avec $E(X)$ définie. Alors, pour tout nombre $t > 0$,*

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Plus généralement, si $g(x) \geq 0$*, alors*

$$P(g(X) \geq t) \leq \frac{E(g(X))}{t}$$

Preuve. *Introduisons la variable tronquée*

$$Y = \begin{cases} 0 & si \quad X < t \\ t & si \quad X \geq t \end{cases}$$

On a clairement $Y \leq X$ *et, par conséquent,* $E(Y) \leq E(X)$*. Maintenant,*

$$E(Y) = 0 \times P(X < t) + t \times P(X \geq t) = tP(X \geq t) \leq E(X)$$

Le résultat suit. La seconde partie est montrée de manière similaire. ■

Exemple 11 *Soit* $X \sim Exp(\lambda)$*. On a* $E(X) = 1/\lambda$*. L'inégalité de Markov dit*

$$P(X \geq t) = e^{-\lambda t} \leq \frac{1}{\lambda t}$$

pour tout $t > 0$*.*

Remarque 4 *Sans information supplémentaire sur la variable* X*, on ne peut pas améliorer l'inégalité de Markov comme le montre l'exemple qui suit.*

Exemple 12 *Soit* $P(X = 0) = P(X = 2) = 1/2$ *et prenons* $t = 2$*. On a* $E(X) = 0 \times \frac{1}{2} + 2 \times \frac{1}{2} = 1$ *et l'inégalité de Markov nous dit*

$$P(X \geq 2) \leq \frac{E(X)}{2} = \frac{1}{2} = P(X = 2)$$

Théorème 7 *Inégalité de Chebyshev. Soit* X *une variable aléatoire ayant pour espérance* μ *et variance* σ^2*. Alors, pour toute constante* $t > 0$*,*

$$\boxed{P(|X - \mu| > t) \leq \sigma^2/t^2}$$

ou, en passant à l'événement complémentaire,

$$P(|X - \mu| \leq t) \geq 1 - \sigma^2/t^2$$

Preuve. *Conséquence de l'inégalité de Markov appliquée à la variable* $Y = |X - \mu|$*.* ■

Exemple 13 *Reprenons l'exemple* $X \sim Exp(\lambda)$*. On a* $Var(X) = 1/\lambda^2$*. L'inégalité de Markov dit*

$$P(X \geq t) = e^{-\lambda t} \leq \frac{1}{\lambda^2 t}$$

pour tout $t > 0$*.*

Exemple 14 *Soit* $X \sim Bin(n, p)$ *et soit* $\overline{X} = X/n$ *(fréquence relative du nombre de succès). On a* $E\left(\overline{X}\right) = np/n = p$ *et* $Var\left(\overline{X}\right) = np(1-p)/n^2 = p(1-p)/n$*. Alors (Chebyshev)*

$$P\left(\left|\overline{X} - p\right| > t\right) \leq \frac{p(1-p)}{nt^2} \leq \frac{1}{4nt^2}$$

(car $p(1-p) \leq 1/4$ *pour tout* $p \in [0;1]$*). On voit que pour* n *suffisamment grand, cette probabilité est petite. Ceci justifie pourquoi* $\overline{X}$ *est l'estimateur naturel de* p*. Nous y reviendrons dans le chapitre sur les théorèmes limites.*

Corollaire 3 *Soit X une variable aléatoire telle que $VAR(X) = 0$. Alors*

$$X = \mu \text{ } pp$$

En d'autres termes, la variable X est concentrée sur un seul point μ :

$$P(X = \mu) = 1$$

Preuve. *En effet, pour tout entier positif n, on a $\left\{|X-\mu| \geq \frac{1}{n}\right\} \subset \{X \neq \mu\}$ et $\left\{|X-\mu| \geq \frac{1}{n}\right\} \subset \left\{|X-\mu| \geq \frac{1}{n+1}\right\}$. La suite d'événements $\left\{|X-\mu| \geq \frac{1}{n}\right\}$ est donc croissante et en fait (exercice),*

$$\{X \neq \mu\} = \cup_{n=1}^{\infty} \left\{|X-\mu| \geq \frac{1}{n}\right\}$$

$$\Rightarrow P(X \neq \mu) \leq \sum_{n \geq 1} P\left(|X-\mu| \geq \frac{1}{n}\right) \leq 0$$

■

Remarque 5 *En l'absence d'information supplémentaire sur la variable X (existence des deux premiers moments), on ne peut pas améliorer l'inégalité de Chebyshev comme le montre l'exemple suivant.*

Exemple 15 *Soit X une variable aléatoire discrète de loi*

X	-1	0	1	*Total*
P	$1/8$	$6/8$	$1/8$	1

On a $\mu = 0$ et $\sigma^2 = 1/4$. On a

$$P(|X| \geq 1) = \frac{1}{4} \geq 1 - \sigma^2/t^2 = \frac{1}{4}$$

La borne est atteinte.

3.1 Autres inégalités*

Théorème 8 *Inégalité de Hoeffding. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et bornées, avec $a_i \leq X_i \leq b_i$. Soit $S_n = \sum_{i=1}^{n} X_i$ leur somme et $\overline{X} = S/n$ leur moyenne empirique. Alors, pour toute constante $t > 0$,*

$$P(|S_n - E(S_n)| \geq t) \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2}\right)$$

Si, de plus les X_i ont la même loi concentrée sur $[a; b]$, l'inégalité devient

$$P\left(\left|\overline{X} - \mu\right| \geq t\right) \leq 2 \exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

où $\mu = E\left(\overline{X}\right)$ est la moyenne des X_i.
Preuve. *Omise.* ■

Exemple 16 *Reprenons l'exemple $X \sim Bin(n, p)$. Ici, X est la somme de n variables iid de Bernoulli. On a alors $(a = 0, b = 1)$*

$$P\left(\left|\overline{X} - p\right| \geq t\right) \leq 2 \exp\left(-2nt^2\right)$$

Cette inégalité est meilleure que celle de Chebyshev trouvée plus haut dès que $nt^2 \geq 1.08$. A titre d'illustration numérique, prenons $t = 0.1$ et $n = 500$. Alors Chbyshev donne

$$P\left(\left|\overline{X} - p\right| \geq 0.1\right) \leq \frac{1}{4nt^2} = \frac{1}{4 \times 500 \times 0.1^2} = 0.05$$

alors que Hoeffding donne

$$P\left(\left|\overline{X} - p\right| \geq 0.1\right) \leq 2 \exp\left(-2nt^2\right) = 2 \exp\left(-2 \times 500 \times 0.1^2\right) = 0.000090800$$

Nettement meilleure.

Remarque 6 *On peut utiliser l'inégalité de Hoeffding fournit une méthode pour construire un intervalle de confiance (chapitre 8) pour le paramètre p d'une loi binomiale. Soit $0 < \alpha < 1$ petit (généralement autour de 5%). Prenons*

$$t = \sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}$$

Alors (Hoeffding),

$$\begin{aligned} P\left(\left|\overline{X} - p\right| \geq t\right) &\leq 2\exp\left(-2nt^2\right) = 2\exp\left(-2n\frac{1}{2n}\ln\frac{2}{\alpha}\right) \\ &= 2\exp\left(-\ln\frac{2}{\alpha}\right) = \alpha \end{aligned}$$

Posons

$$IC_{1-\alpha} = \left]\overline{X} - t, \overline{X} + p\right[$$

Alors

$$\begin{aligned} &P\left(p \notin IC_{1-\alpha}\right) = P\left(\left|\overline{X} - p\right| \geq t\right) \leq \alpha \\ &\qquad \Rightarrow P\left(p \in IC_{1-\alpha}\right) = P\left(p \in \left]\overline{X} - t, \overline{X} + p\right[\right) \geq 1 - \alpha \end{aligned}$$

Une limitation de l'inégalité de Hoeffding est le fait que la variable aléatoire X soit bornée. On peut parfois obtenir des bornes exponentielles explicites si on suppose que X admet des moments exponentiels :

$$E\left(e^{\lambda X}\right) < \infty$$

Alors, pour toute constante $t > 0$ et tout $\lambda > 0$, l'inégalité de Markov donne

$$P\left(X \geq t\right) = P\left(e^{\lambda X} \geq e^{\lambda t}\right) \leq e^{-\lambda t} E\left(e^{\lambda X}\right) = g\left(\lambda\right)$$

Si la fonction g atteint son minimum en λ_0, alors

$$P\left(X \geq t\right) \leq g\left(\lambda_0\right) = e^{-\lambda_0 t} E\left(e^{\lambda_0 X}\right)$$

Ce résultat s'appelle borne de Chernoff.

Définition 3 *Fonction convexe. Rappelons qu'une fonction g définie sur un intervalle I (I peut être infini) est convexe (concave vers le haut) si*

$$g\left(\alpha x + (1-\alpha) y\right) \leq \alpha g\left(x\right) + (1-\alpha) g\left(y\right), \qquad x, y \in I, 0 \leq \alpha \leq 1$$

(ou encore $g''(x) > 0$ sur I, si cette dérivée existe). On dit que g est strictement convexe si l'inégalité est stricte.

Ceci nous dit que le segment de droite joignant les points $(x, g(x))$ et $(y, g(y))$ est au dessus du graphe de g.

Exemple 17 *Des exemples de fonctions convexes sont $g(x) = x^2$, $g(x) = e^x$ et $g(x) = \max(0, x)$.*

Théorème 9 *Inégalité de Jensen. Soit g convexe sur un intervalle I. Alors*

$$g\left(E\left(X\right)\right) \leq E\left(g\left(X\right)\right)$$

avec X concentrée sur I. Si g est strictement convexe, l'inégalité est stricte (à moins que X soit constante, i.e. dégénérée).

Exemple 18 *Soit $X \sim Poi(\lambda)$. Alors (Jensen) $\lambda^3 < E\left(X^3\right)$ puisque la fonction $g(x) = x^3, x > 0$, est strictement convexe. Note : $E\left(X^3\right) = \lambda^3 + 3\lambda^2 + \lambda$.*

4 Covariance. Coefficient de corrélation

Étant donné un couple de variables aléatoires (X, Y), leur covariance est définie (sous réserve de convergence absolue) par

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

C'est une mesure de la variabilité du couple (X, Y) par rapport au point moyen $(\mu_X, \mu_Y) = (E(X), E(Y))$. Il est immédiat de vérifier que

$$Cov(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

Théorème 10 *Inégalité de Cauchy-Schwarz. Si les variables X et Y sont de variances finies, alors*

$$|Cov(X, Y)| \leq \sigma_X \sigma_Y$$

De plus, si $\sigma_X > 0$, on a égalité si et seulement si

$$Y - \mu_Y = c(X - \mu_X)$$

(colinéaires) où la constante c est donnée par

$$c = \frac{Cov(X, Y)}{\sigma_X^2}$$

En d'autres termes,

$$Y = \frac{Cov(X, Y)}{\sigma_X^2}(X - \mu_X) + \mu_Y$$

Preuve. *Calculer $Var(Y - tX)$ et choisir t de manière judicieuse.* ■

Les propriétés suivantes sont faciles à montrer.

- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(a + bX + cY, Z) = bCov(X, Z) + cCov(Y, Z)$.
- $Cov(a + X, Y) = Cov(X, a + Y) = Cov(X, Y)$ pour toute constante a.
- $Cov(bX, Y) = Cov(X, bY) = bCov(X, Y)$ pour toute constante b.
- $VAR(a + bX + cY) = VAR(bX + cY) = b^2 VAR(X) + 2bcCov(X, Y) + c^2 VAR(Y)$.
- Si X et Y sont indépendantes, alors $Cov(X, Y) = 0$. On dit que X et Y sont non corrélées.
- Si X et Y sont indépendantes (ou non corrélées), alors $VAR(X + Y) = VAR(X) + VAR(Y)$.
- $|E(XY)| \leq \sqrt{E(X^2) \cdot E(Y^2)}$ (autre forme de l'inégalité de Cauchy-Schwarz).

Remarque 7 *Deux variables non corrélées ne sont pas nécessairement indépendantes. Cependant, dans le cas important où les deux variables sont conjointement normalement distribuées, ceci est vrai.*

Exemple 19 *Vous lancez une pièce de monnaie équilbrée 2 fois. Vous gagnez 1 dollar pour chaque pile obtenu, mais vous en perdez 1 pour chaque face obtenue. Soit X votre gain et soit $Y = X^2$. On vérifie facilement que X et Y ne sont pas corrélées :*

$$E(X) = -2 \times \frac{1}{4} + 0 \times \frac{1}{2} + 2 \times \frac{1}{4} = 0$$

$$E(XY) = E\left(X^3\right) = -8 \times \frac{1}{4} + 0 \times \frac{1}{2} + 8 \times \frac{1}{4} = 0$$

$$E(X)E(Y) = 0$$

On a donc $Cov(X, Y) = 0$. Pourtant, ces deux variables sont clairement dépendantes (relation fonctionnelle). Par exemple,

$$P(X = 0, Y = 0) = P(X = 0) = \frac{1}{2}$$

$$P(X = 0)P(Y = 0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Exemple 20 *Soit* $X \sim N(0,1)$ *et* $Y = X^2$. *On a*

$$Cov(X,Y) = E(XY) - E(X)E(Y) = E(XY) = E\left(X^3\right) = 0$$

Par ailleurs, on a

$$P(Y > 1) = P\left(X^2 > 1\right) = 2\Phi(-1) > 0$$

de toute évidence, mais

$$P(Y > 1\,|{-1} \leq X \leq 1) = P\left(X^2 > 1\,|{-1} \leq X \leq 1\right) = 0$$

ce qui montre que X *et* Y *sont dépendantes.*

4.1 Coefficient de corrélation

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = E\left[\frac{X-\mu_X}{\sigma_X}\frac{Y-\mu_Y}{\sigma_Y}\right]$$

On a les propriétés immédiates suivantes.

- $-1 \leq \rho \leq 1$
- $\rho = 0 \Leftrightarrow X$ et Y sont non corrélées.
- $\rho = \pm 1 \Rightarrow Y = aX + b = \mu_Y + \frac{Cov(X,Y)}{\sigma_X^2}(X - \mu_X)$ pp $(P(Y = aX + b) = 1)$. Voir l'inégalité de Cauchy-Schwarz.

4.2 Exercice

Exercice 5 *Soit* $X_1, \ldots, X_n$ *iid (indépendantes et identiquement distribuées) et posons* $E(X_i) = \mu$, $Var(X_i) = \sigma^2$. *Soit* $Y = c_1X_1 + \ldots + c_nX_n$, *où les* c_i *sont des constantes.*

1. *Trouver la moyenne et la variance de* Y.
2. *Pour quelles valeurs des constantes la variance de* Y *est-elle minimale ?*

5 Matrice de variance-covariance*

Un vecteur aléatoire $X = (X_1, \ldots, X_n)$ s'écrit, en notation matricielle, comme une matrice colonne :

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

On définit l'espérance de X comme le vecteur des espérances individuelles $\mu_i = E(X_i)$:

$$\mu_X = (\mu_1, \ldots, \mu_n) = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

La matrice de variance covariance de X est la matrice

$$Cov(X) = \sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & & & \sigma_n^2 \end{bmatrix}$$

où $\sigma_{ij} = Cov(X_i, X_j)$. En particulier, $\sigma_{ii} = \sigma_i^2 = Var(X_i)$. Noter que $\sum$ est une matrice symétrique puisque $Cov(X_i, X_j) = Cov(X_j, X_i)$. Si les X_i sont indépendants, cette matrice est diagonale. Si, de plus, elles sont identiquement distribuées, alors $\sum = \sigma^2 I_n$ (où I_n est la matrice identité d'ordre n).

L'écriture matricielle simplifie souvent les notations et permet d'écrire des résultats sous forme succinte. Par exemple, si $X = (X_1, \ldots, X_n)$ et $a = (a_1, \ldots, a_n)$, on a

$$a^T X = [a_1 \ldots a_n] \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = a_1 X_1 + \ldots + a_n X_n$$

et donc

$$E\left(a^T X\right) = a^T \mu_X$$

et (vérifier)

$$Var\left(a^T X\right) = a^T Cov\,(X)\,a$$

Ici, le T en exposant indique la matrice transposée.

5.1 Exercice

Exercice 6 *Soit donné le couple $X = (X_1, X_2)$ de loi conjointe*

$X_1 \setminus X_2$	1	2	3
-1	$1/6$	0	$1/6$
1	$2/6$	$1/6$	$1/6$

Calculer la matrice de variance-covariance.

Exercice 7 *Soit $X = (X_1, X_2)$ de loi conjointement normale. Trouver la matrice de variance-covariance correspondante.*

6 Espérance conditionnelle

Rappelons que l'espérance conditionnelle de Y en X est donnée par

$$E\,(Y \mid X = x) = \begin{cases} \sum_j y_j p_{Y|X}\,(y_j \mid x) & \text{si } Y \text{ est discrète} \\ \int_{\mathbb{R}} y f_{Y|X}\,(y \mid x)\,dy & \text{si } Y \text{ est continue} \end{cases}$$

Plus généralement, pour une fonction h, on a

$$E\,(h\,(Y) \mid X = x) = \begin{cases} \sum_j h\,(y_j)\,p_{Y|X}\,(y_j \mid x) & \text{si } Y \text{ est discrète} \\ \int_{\mathbb{R}} h\,(y)\,f_{Y|X}\,(y \mid x)\,dy & \text{si } Y \text{ est continue} \end{cases}$$

On a le résultat important suivant.

Théorème 11

$$E\,(Y) = E\,[E\,(Y \mid X)] \quad \textit{(espérance itérée)}$$

$$VAR\,(Y) = VAR\,[E\,(Y \mid X)] + E\,[VAR\,(Y \mid X)]$$

Plus généralement,

$$E\,(h\,(X, Y)) = E\,[E\,(h\,(Y \mid X) \mid X)]$$

En particulier,

$$E\,(h\,(X)\,Y \mid x) = h\,(x)\,E\,(Y \mid x)$$

Exemple 21 *Somme aléatoire. Soit $T = \sum_{i=1}^{N} X_i$ où N est une variable aléatoire discrète concentrée sur les entiers positifs (avec espérance finie) et les X_i sont indépendants, identiquement distribués et indépendants de N (avec espérance finie $E\,(X)$). Alors,*

$$E\,(T) = E\,[E\,(T \mid N)] = E\,(N\,(E\,(X))) = E\,(N)\,E\,(X)$$

Si on ajoute l'hypothèse que les variances existent, alors (exercice)

$$VAR\,(T) = E\,(X)^2\,VAR\,(N) + E\,(N)\,VAR\,(X)$$

6.1 Estimation. Prédiction

Le meilleur prédicteur $h(X)$ de Y en X (qui minimise l'erreur quadratique moyenne

$$EQM = E\left[(Y - h(X))^2\right]$$

est donné par

$$h(X) = E(Y \mid X)$$

En particulier, le meilleur prédicteur linéaire $(h(X) = a + bX)$ est

$$h(X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

Dans le cas ou X et Y sont normalement distribuées, le meilleur prédicteur est le meilleur prédicteur linéaire :

$$\boxed{E(Y \mid X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)}$$

7 Fonction génératrice des moments (FGM)

La fonction génératrice des moments d'une variable aléatoire X est donnée par (sous réserve de convergence absolue dans un voisinage de $t = 0$) :

$$M_X(t) = E\left(e^{tX}\right) = \begin{cases} \sum_i e^{tx_i} p_X(x_i) & \text{si } X \text{ est discrète} \\ \int_{\mathbb{R}} e^{tx} f_X(x)\, dx & \text{si } X \text{ est continue} \end{cases}$$

Citons quelques unes des propriétés de cette fonction.

- Si X et Y sont indépendantes,
$$M_{X+Y}(t) = M_X(t)\, M_Y(t)$$
- $$M_{a+bX}(t) = e^{at} M_X(bt)$$
pour toutes constantes a et b.
- $$M_X(t) = 1 + \sum_{k\geq 1} \frac{E\left(X^k\right)}{k!} t^k$$
sous réserve de convergence (sinon, on se contentera d'un développement limité d'ordre fini).
- Pour tout k entier ≥ 0,
$$\mu_k = M_X^{(k)}(0) = E\left(X^k\right)$$
C'est le <u>moment d'ordre k</u> de la variable X.
- Soit X et Y admettant pour FGM $M_X(t)$ et $M_Y(t)$. Alors $F_X(u) = F_Y(u)$ pour tout $u \in \mathbb{R}$ si et seulement si $M_X(t) = M_Y(t)$ pour tout t dans un voisinage $]-h; h[$ $(h > 0)$ de $t = 0$.

Exemple 22 *Si $X \sim Poi(\lambda)$. Alors*

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(e^t \lambda\right)^k}{k!} = e^{-\lambda\left(1+e^t\right)}$$

Exemple 23 *Si* $X \sim Exp(\lambda)$. *Alors*

$$M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{-x(\lambda-t)}dx = \frac{\lambda}{\lambda-t} \qquad si\ t<\lambda$$

On a aussi (pour $|t|<\lambda$*)*

$$M_X(t) = \frac{\lambda}{\lambda-t} = \frac{1}{1-t/\lambda} = \sum_{k=0}^\infty (t/\lambda)^k = \sum_{k=0}^\infty \left(\frac{k!}{\lambda^k}\right)\frac{t^k}{k!}$$

On en déduit

$$E\left(X^k\right) = \frac{k!}{\lambda^k}$$

En particulier,

$$E(X) = \frac{1}{\lambda} \text{ et } E\left(X^2\right) = \frac{2}{\lambda^2}$$

Exemple 24 *Si* $X \sim N(0,1)$. *Alors*

$$\begin{aligned} M_X(t) &= \int_{-\infty}^\infty e^{tx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-x^2/2+tx}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}\left(x^2-2tx\right)}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}\left(x^2-2tx+t^2-t^2\right)}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}\left((x-t)^2-t^2\right)}dx = e^{t^2/2}\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}(x-t)^2}dx}_{=1} \\ &= e^{t^2/2} \end{aligned}$$

On a

$$M_X'(t) = \frac{d}{dt}e^{t^2/2} = te^{\frac{1}{2}t^2} \Rightarrow M_X'(0) = E(X) = 0$$

et

$$M_X''(t) = \frac{d^2}{dt^2}e^{t^2/2} = e^{\frac{1}{2}t^2}\left(t^2+1\right) \Rightarrow M_X''(0) = E\left(X^2\right) = Var(X) = 1$$

Noter qu'on peut trouver tous les moments de $E(X^n)$ *directement à partir du développement en série de McLaurin de* $M_X(t) = e^{t^2/2}$:

$$e^{t^2/2} = 1 + \frac{1}{2}t^2 + \frac{1}{8}t^4 + \frac{1}{48}t^6 + \frac{1}{384}t^8 + \ldots = \sum_{n=0}^\infty \frac{(2n)!}{n!2^n}\frac{t^{2n}}{(2n)!}$$

$$\Rightarrow \boxed{\left\{\begin{array}{l} E\left(X^{2n}\right) = \dfrac{(2n)!}{n!2^n} \\ E\left(X^{2n+1}\right) = 0 \end{array}\right. ,n=0,1,\ldots}$$

On peut facilement déduire la FGM d'une variable $Y \sim N\left(\mu,\sigma^2\right)$ *à partir des propriétés ci-dessus. Puisque* $X = \frac{Y-\mu}{\sigma} \sim N(0,1)$, *on a* $Y = \sigma X + \mu$ *et*

$$\begin{aligned} M_Y(t) &= M_{\sigma X+\mu}(t) = e^{\mu t}M_X(\sigma t) = e^{\mu t}e^{(\sigma t)^2/2} \\ &= e^{\mu t+\sigma^2t^2/2} \end{aligned}$$

Remarque 8 *La fonction* $\varphi_X(t) = M_X(it)$ $(i=\sqrt{-1})$ *est toujours définie et s'appelle fonction caractéristique de* X. *Elle est préférable à la FGM (utilisée dans des cours plus avancés). Exemples : Cauchy, Riemann.*

7.1 Fonctions génératrices (FG)

Si la variable X est discrète concentrée sur les entiers non négatifs, il est parfois préférable d'utiliser un autre outil, celui de la fonction génératrice. Elle est définie par

$$G_X(z) = E\left(z^X\right) = \sum_{k\geq 0} z^k p_X(k)$$

Noter que

$$G_X\left(e^t\right) = M_X(t)$$

et que

$$p_X(k) = \frac{1}{k!} G_X^{(k)}(0)$$

De plus,

$$\frac{d^k}{dz^k} G_X(z)\left.\right|_{z=1} = E\left(X(X-1)\cdots(X-k+1)\right), \quad k = 1, 2, \ldots$$

si ces derniers existent. En particulier,

$$\begin{aligned}
&G_X'(1) = E(X) \\
&G_X''(1) = E(X(X-1)) = E\left(X^2\right) - E(X) \Rightarrow E\left(X^2\right) = G_X''(1) + G_X'(1)
\end{aligned}$$

Exemple 25 *Si $X \sim Poi(\lambda)$. Alors*

$$G_X(z) = E\left(z^X\right) = \sum_{k=0}^{\infty} z^k \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(z-1)}$$

On a

$$\begin{aligned}
&\frac{d}{dz} e^{\lambda(z-1)} = \lambda e^{\lambda(z-1)} \Rightarrow E(X) = \lambda \\
&\frac{d^2}{dz^2} e^{\lambda(z-1)} = \lambda^2 e^{\lambda(z-1)} \Rightarrow E\left(X^2\right) - E(X) = \lambda^2 \\
&et\ Var(X) = \lambda
\end{aligned}$$

:

8 Méthode delta

Il arrive souvent qu'on ne connaisse que les deux premiers moments d'une variable aléatoire (donc pas sa loi). Soit donc donné une variable aléatoire X avec $E(X) = \mu_X$ et $VAR(X) = \sigma_X^2$. Soit $Y = g(X)$, où g est une fonction (suffisamment lisse) donnée. On s'intéresse à $E(Y) = \mu_Y$ et $VAR(Y) = \sigma_Y^2$. Faisons un développement limité en série de Taylor d'ordre un autour de μ_X :

$$g(x) = g(\mu_X) + (x - \mu_X) g'(\mu_X) + erreur$$

Ce qui donne, approximativement,

$$Y \approxeq g(\mu_X) + (X - \mu_X) g'(\mu_X)$$

Prenant l'espérance et la variance, on obtient :

$$\begin{aligned}
\mu_Y &\approxeq g(\mu_X) \\
\sigma_Y^2 &\approxeq \sigma_X^2 g'(\mu_X)^2
\end{aligned}$$

Si on fait un développement limité d'ordre deux, on obtient :

$$\mu_Y \approxeq g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$$

Remarque 9 *La méthode delta peut aussi être utilisé dans le cas d'un couple de variables aléatoires* (X,Y) *(ou plus) :* $Z = g(X,Y)$*. Le développement limité se fera autour du point* (μ_X, μ_Y) *:*

$$Z = g(X,Y) \approxeq g(\mu_X, \mu_Y) + (X - \mu_X)\frac{\partial g}{\partial x}(\mu_X, \mu_Y) + (Y - \mu_Y)\frac{\partial g}{\partial y}(\mu_X, \mu_Y) + erreur$$

Exemple 26 *Soit* $X > 0$ *et* $g(x) = \ln x$*. Alors*

$$g'(x) = \frac{1}{x}$$
$$g''(x) = -\frac{1}{x^2}$$

Ce qui donne

$$\mu_Y = E(\ln X) \approxeq \ln \mu_X - \frac{1}{2}\frac{\sigma_X^2}{\mu_X^2}$$
$$\sigma_Y^2 = VAR(\ln X) \approxeq \frac{\sigma_X^2}{\mu_X^2}$$

Comme application numérique, prenons $X \sim U\,]0;1[$*. On a* $E(X) = \frac{1}{2}$, $\sigma_X^2 = \frac{1}{12}$*. Alors,*

$$E(\ln X) \approxeq -\ln 2 - \frac{1}{6} = -0.859\,81$$
$$VAR(\ln X) \approxeq \frac{1}{3}$$

Le calcul exact donne :

$$E(\ln X) = \int_0^1 \ln x dx = -1$$
$$VAR(\ln X) = \int_0^1 \ln^2 x dx - 1 = 2 - 1 = 1$$

:On voit que l'approximation n'est pas fameuse.

Exemple 27 *Reprenons l'exemple précédent* ($Y = \ln X$)*, mais cette fois-ci* $X \sim U\,]1;2[$*. On a* $E(X) = \frac{3}{2}$, $\sigma_X^2 = \frac{1}{12}$ *et*

$$E(\ln X) \approxeq \ln\frac{3}{2} - \frac{1}{2 \times 12 \times \frac{9}{4}} = 0.386\,95$$
$$VAR(\ln X) \approxeq \frac{1}{12} \times \frac{4}{9} = \frac{1}{27} = 0.037037$$

Les valeurs exactes sont :

$$E(\ln X) = \int_1^2 \ln x dx = 2\ln 2 - 1 = 0.386\,29$$
$$VAR(\ln X) = \int_1^2 \ln^2 x dx - (2\ln 2 - 1)^2 = 2(\ln 2 - 1)^2 - (2\ln 2 - 1)^2 = 0.03\,909\,4$$

L'approximation est bien meilleure. Quel commentaire cela vous inspire-t-il ?