Vecteurs aléatoires — partie 2

À l'intention des étudiants de STT-1500



Jérôme Soucy

Département de mathématiques et de statistique
Université Laval

Hiver 2025

Résumé

1. Espérance d’une fonction de variables aléatoires
2. Somme de v.a. et f.g.m.
3. Covariance et corrélation
4. Espérance par conditionnement

1

Espérance d’une fonction de variables aléatoires

Théorème de changement de variable

Théorème de changement de variable

Théorème. Si (X, Y) est un vecteur aléatoire, et si g est une fonction réelle dont le domaine contient l'ensemble des valeurs possibles du couple (X, Y), alors

$$\mathbb{E}(g(X,Y)) = \begin{cases} \sum_{(x,y)\in\mathbb{R}^2} g(x,y)p(x,y) & \text{dans le cas discret,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy & \text{dans le cas abs. continu.} \end{cases}$$

La fonction p désigne la fonction de masse conjointe des variables aléatoires X et Y dans le cas discret, et la fonction f désigne la fonction de densité conjointe des variables aléatoires X et Y dans le cas absolument continu.

Conséquences sur l’espérance

Conséquences sur l'espérance

Théorème. Si $X_1, X_2, \dots, X_n$ sont des variables aléatoires avec espérance finie, alors

$$\mathbb{E}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathbb{E}(X_k).$$

De plus, **si les variables $X_1, X_2, \dots, X_n$ sont indépendantes**, alors

$$\mathbb{E}\left(\prod_{k=1}^{n} X_k\right) = \prod_{k=1}^{n} \mathbb{E}(X_k).$$

Conséquences sur l’espérance — suite

Conséquences sur l'espérance — suite

Théorème. Si X et Y sont des variables aléatoires indépendantes, et si les espérances $\mathbb{E}(h(X))$ et $\mathbb{E}(g(Y))$ sont finies, alors

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

2

Somme de v.a. et f.g.m.

Somme de v.a. et f.g.m.

Somme de v.a. et f.g.m.

Nous avons déjà vu que la fonction génératrice des moments d'une v.a. X est définie par

$$M_X(t) = \mathbb{E}(e^{tX}) = \begin{cases} \sum_{x \in \mathbb{R}} e^{tx} p(x) & \text{dans le cas discret,} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{dans le cas abs. continu.} \end{cases}$$

Si on connait la f.g.m. d'une somme de variables aléatoires indépendantes, on peut parfois connaître la distribution de cette somme si on peut identifier la loi de distribution qui possède la f.g.m. trouvée. La f.g.m. de cette somme peut être calculée facilement si on connaît les f.g.m. des deux variables dont on veut en étudier la somme.

Résultat

Proposition. Si $X_1, X_2, \dots, X_n$ sont des variables aléatoires indépendantes avec respectivement $M_{X_1}, M_{X_2}, \dots, M_{X_n}$ comme fonction génératrice des moments, alors la fonction génératrice des moments de la variable aléatoire

$$S = \sum_{k=1}^{n} X_k$$

est donnée par

$$M_S(t) = \prod_{k=1}^{n} M_{X_k}(t).$$

La loi de Poisson

Supposons que $X_1, X_2, \dots, X_n$ sont des v.a. i.i.d. (indépendantes identiquement distribuées) suivant une loi de Poisson de paramètre ν. Trouvez la loi de la somme des X_i.

La loi exponentielle

Supposons que $X_1, X_2, \dots, X_n$ sont des v.a. indépendantes dont la loi est l'exponentielle de paramètre λ. Trouvez la fonction génératrice de la v.a. $S = \sum_{k=1}^{n} X_k$.

3

Covariance et corrélation

Définition

Définition

Définition. Soit X et Y des variables aléatoires. La covariance de X et Y, notée $\mathrm{Cov}(X, Y)$, est définie par

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

Résultats

Résultats

Proposition. Soit X et Y des v.a., a, b, c et d des nombres réels. Alors

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y).$$

Proposition. Soit X et Y des v.a.. Un calcul direct montre que

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

Résultats — cas des variables indépendantes

Résultats — cas des variables indépendantes

Proposition. Soit X et Y des variables aléatoires indépendantes. Alors

$$\mathrm{Cov}(X, Y) = 0.$$



$\mathrm{Var}(X) = \mathrm{Cov}(X, X)$

Covariance et variance

Proposition. Soit X et Y des variables aléatoires. Alors

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y).$$

Plus généralement, si $X_1, X_2, \dots, X_n$ sont des variables aléatoires avec variance finie, alors

$$\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{1 \le i < j \le n}^{n} \mathrm{Cov}\big(X_i, X_j\big).$$

Ainsi, si les $X_1, X_2, \dots, X_n$ sont des variables aléatoires **indépendantes**, on peut conclure que la variance d'une somme est la somme des variances.

Coefficient de corrélation

Coefficient de corrélation

Définition. Soit X et Y des variables aléatoires telles que leur variance est strictement positive. On définit le coefficient de corrélation de X et Y, noté $\rho_{X,Y}$, par

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

Le coefficient de corrélation, tout comme la covariance, est une mesure du degré d'association linéaire de deux variables. L'avantage du coefficient de corrélation par rapport à la covariance est qu'il est normalisé : Il se situe toujours entre -1 et 1, et il n'a pas d'unité.

Un exemple

Supposons que X soit une variable aléatoire uniforme sur l'intervalle $]0, 1[$. On défini Y comme étant la variable aléatoire $1 - X$. Calculez le coefficient de corrélation de X et Y.

Solution

On déduit que Y est aussi une uniforme sur l'intervalle $]0, 1[$ car elles ont la même densité.

> L'espérance et la variance d'une v.a. uniforme sur (a, b) sont respectivement $\frac{a+b}{2}$ et $\frac{(b-a)^2}{12}$.

D'une part, on trouve que $\mathbb{E}(X) = \mathbb{E}(Y) = 1/2$ et $\text{Var}(X) = \text{Var}(Y) = 1/12$.

(suite)

D'autre part, on trouve que

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\\ &= \mathbb{E}(X(1-X)) - 1/2 \cdot 1/2\\ &= \mathbb{E}(X) - \mathbb{E}(X^2) - 1/4\\ &= -1/12.\end{aligned}$$

Ainsi,

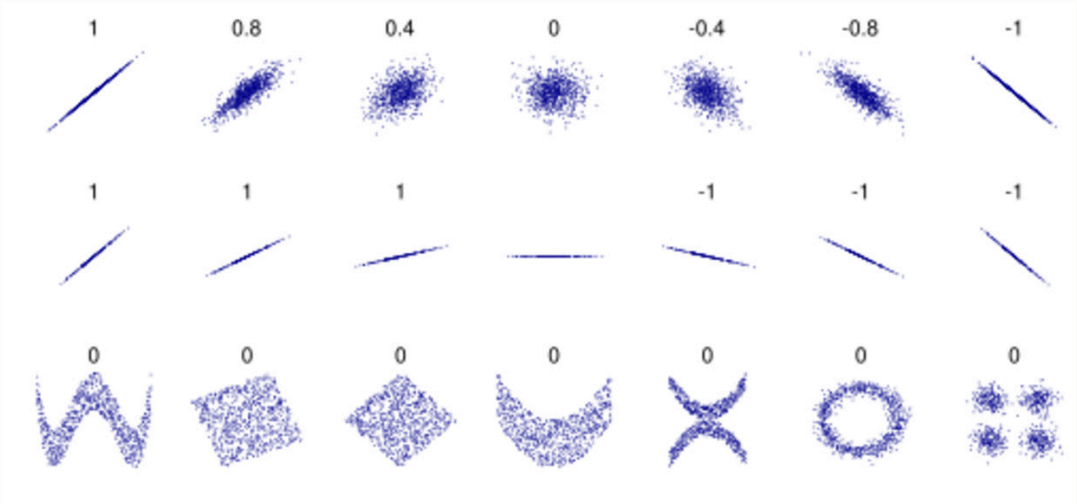
$$\begin{aligned}\rho_{X,Y} &= \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}\\ &= \frac{-1/12}{\sqrt{1/12} \cdot \sqrt{1/12}}\\ &= -1.\end{aligned}$$

Coefficient de corrélation — suite

Coefficient de corrélation — suite

Théorème. Soit X et Y des variables aléatoires telles que $0 < \sigma_X < \infty$ et $0 < \sigma_Y < \infty$. Alors

1. $-1 \leq \rho_{X,Y} \leq 1$;
2. $\rho_{X,Y} = 1$ si et seulement si il existe des constantes a et b, avec $a > 0$, telles que $\mathbb{P}(Y = aX + b) = 1$;
3. $\rho_{X,Y} = -1$ si et seulement si il existe des constantes a et b, avec $a < 0$, telles que $\mathbb{P}(Y = aX + b) = 1$;
4. Si a, b, c et d sont des constantes, et si $a \neq 0$ et $c \neq 0$ sont de même signe, alors $\rho_{aX+b,cY+d} = \rho_{X,Y}$;
5. Si a, b, c et d sont des constantes, et si $a \neq 0$ et $c \neq 0$ sont de signe opposés, alors $\rho_{aX+b,cY+d} = -\rho_{X,Y}$;



https://commons.wikimedia.org/w/index.php?title=File:Correlation_examples2.svg&oldid=696631151

Un exemple

On lance une pièce de monnaie 3 fois. On note X le nombre de fois où pile a été obtenu et on note Y la longueur de la plus longue sous-suite composée uniquement du résultat pile. Par exemple, pour la suite PPF on a que $X = 2$ et $Y = 2$ alors que pour PFP on a que $X = 2$ et $Y = 1$. Calculez le coefficient de corrélation de X et Y.

Nous avons que

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Calculons chacun des termes apparaissant dans cette expression.

Calcul de σ_X

On constate que $X \sim$ binomiale $\left(3, \frac{1}{2}\right)$.

> Rappel : la variance d'une v.a. de loi binomiale (n, p) est $np(1-p)$.

Ainsi,

$$\sigma_X = \sqrt{3 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \frac{\sqrt{3}}{2}.$$

Calcul de σ_Y

Rappel : la variance d'une variable aléatoire Y est $\mathbb{E}(Y^2) - \mathbb{E}(Y)^2$.

Valeurs de k	$\mathbb{P}(Y = k)$
0	$1/8$
1	$4/8$
2	$2/8$
3	$1/8$

Valeurs de k	$\mathbb{P}(Y^2 = k)$
0	$1/8$
1	$4/8$
4	$2/8$
9	$1/8$

$$\mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = 1 \cdot 4/8 + 4 \cdot 2/8 + 9 \cdot 1/8 - \Big(\underbrace{1 \cdot 4/8 + 2 \cdot 2/8 + 3 \cdot 1/8}_{\mathbb{E}(Y) = 11/8}\Big)^2 = \frac{47}{64},$$

d'où $\sigma_Y = \sqrt{47}/8$.

Calcul de $\mathrm{Cov}(X, Y)$

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

Les valeurs possibles pour X et Y sont $0, 1, 2$ et 3.
Les valeurs possibles pour XY sont $0, 1, 4$ et 9 (6 n'est pas une valeur possible).

Valeurs de k	$\mathbb{P}(XY = k)$
0	$1/8$
1	$3/8$
2	$1/8$
4	$2/8$
9	$1/8$

$$\mathrm{Cov}(X, Y) = \underbrace{1 \cdot 3/8 + 2 \cdot 1/8 + 4 \cdot 2/8 + 9 \cdot 1/8}_{\mathbb{E}(XY)} - \underbrace{3/2}_{\mathbb{E}(X)} \cdot \underbrace{11/8}_{\mathbb{E}(Y)} = 11/16.$$

Conclusion

Nous avons que

$$\begin{aligned}\rho_{X,Y} &= \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \\ &= \frac{{}^{11}\!/_{16}}{\frac{\sqrt{3}}{2} \cdot \frac{\sqrt{47}}{8}} \\ &= \frac{11\sqrt{141}}{141} \\ &\approx 0{,}9264.\end{aligned}$$

4

Espérance par conditionnement

Espérance par conditionnement

On veut calculer l'espérance d'une v.a. Y. Le problème est tel qu'on connaît la densité de probabilité $f(x)$ (ou la fonction de masse $p(x)$) d'une variable aléatoire X et aussi l'espérance de Y sachant la valeur prise par la variable aléatoire X. Dans ce cas, nous avons que

$$\mathbb{E}(Y) = \begin{cases} \sum_{x \in \mathbb{R}} \mathbb{E}(Y|X=x)p(x) & \text{dans le cas discret,} \\ \int_{-\infty}^{\infty} \mathbb{E}(Y|X=x)f(x)dx & \text{dans le cas abs. continu.} \end{cases}$$

Remarques

Remarques.

- Ce document est rédigé avec Beamer, une classe de LaTeX.
- Les images proviennent de Wikipédia et un lien a été placé sous ces dernières.

Pour signaler une erreur dans ce document, veuillez écrire à jerome.soucy@mat.ulaval.ca.