

PC 4 : Vecteurs aléatoires à densités - lois conditionnelles

1 Calcul de lois et conditionnement

Exercice 1. Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$. Pour tout $p \in \mathbb{R}$, justifier que e^{pX} est intégrable et calculer $\mathbb{E}(e^{pX})$.

Exercice 2. Soit X et Y deux variables aléatoires réelles indépendantes et $S = X + Y$. Lorsque X et Y suivent la loi exponentielle de paramètre $\lambda > 0$, déterminer la densité conditionnelle de X sachant $S = s$. En déduire $\mathbb{E}(X|S)$. Reprendre cette question lorsque X et Y suivent la loi uniforme sur $[0, 1]$.

Exercice 3. Pour $n \geq 2$, soient U_1, \dots, U_n des variables aléatoires uniformes sur $[0, 1]$ indépendantes et $(X, Y) = (\min_{1 \leq i \leq n} U_i, \max_{1 \leq i \leq n} U_i)$. Pour $0 \leq x \leq y \leq 1$, calculer $\mathbb{P}(x < X, Y \leq y)$. En déduire la densité conditionnelle de X sachant $Y = y$ puis $\mathbb{E}[X|Y]$.

2 Complément de cours : droites de régression

Exercice 4 (cas d'égalité dans l'inégalité de Cauchy-Schwarz). Soit X et Y deux v.a. réelles dans L^2 telle que

$$\mathbb{E}[XY]^2 = \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

qu'en déduisez-vous sur X et Y ? Soit Z une v.a. réelle dans L^2 telle que

$$\mathbb{E}[Z^2] = \mathbb{E}[Z]^2$$

qu'en déduisez-vous sur Z ?

Exercice 5 (Régression linéaire en dimension 1). Soit X, Y deux v.a. réelles dans L^2 . Montrer que

$$\min_{a,b} \mathbb{E}[(Y - aX - b)^2] = \text{Var}(Y)(1 - \rho^2(X, Y)) = \mathbb{E}[(Y - \hat{a}X - \hat{b})^2]$$

avec

$$\hat{a} := \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad \hat{b} = \mathbb{E}[Y] - \hat{a}\mathbb{E}[X].$$

Exercice 6 (Régression linéaire en dimension n). Soit X_1, \dots, X_n, Y des v.a. réelles dans L^2 . Montrer que

$$\min_{a_j, 0 \leq j \leq n} \mathbb{E}[(Y - a_0 - \sum_{j=1}^n a_j X_j)^2] = \text{Var}(Y)(1 - R^2) = \mathbb{E}[(Y - \hat{a}_0 - \sum_{j=1}^n \hat{a}_j X_j)^2]$$

avec

$$\hat{a}_j := (\mathbf{C}^{-1}\mathbf{c})_j, \quad 1 \leq j \leq n, \quad \hat{a}_0 := \mathbb{E}[Y] - \sum_{j=1}^n \hat{a}_j \mathbb{E}[X_j],$$

où $\mathbf{c} = (\text{Cov}(X_j, Y))_{1 \leq j \leq n}$ est le vecteur de covariance entre Y et $\mathbf{X} = (X_j)_{1 \leq j \leq n}$,

$\mathbf{C} = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n}$ est la matrice de covariance des X_j , et R^2 est le coefficient de détermination

$$R^2 = \frac{\mathbf{c}^T \mathbf{C}^{-1} \mathbf{c}}{\text{Var}(Y)}.$$

3 Indépendance, corrélation & conditionnement

Exercice 7. (UN CONTRE-EXEMPLE CLASSIQUE)

Construire un couple de v.a. (X, Y) tel que : X et Y sont 2 v.a. gaussiennes, $\text{Cov}(X, Y) = 0$ et X n'est pas indépendante de Y .

Exercice 8. Soit (X_1, X_2, X_3) un vecteur aléatoire de matrice de variance covariance

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 5 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

1. Calculer la variance de $X_3 - \alpha_1 X_1 - \alpha_2 X_2$ pour $\alpha_1, \alpha_2 \in \mathbb{R}$.
2. En déduire qu'il existe une constante c telle que $X_3 = X_1 + X_2 + c$ p.s.
3. Plus généralement, on considère un vecteur aléatoire Y à valeurs dans \mathbb{R}^n de matrice de variance covariance Γ .
 - (a) Montrer que pour tout $u \in \mathbb{R}^n$, $\text{Var}(\sum_{i=1}^n u_i Y_i) = u^\top \Gamma u$.
 - (b) En déduire que Γ est non-inversible si et seulement si l'une des composantes de Y est presque sûrement égale à une fonction affine des autres composantes de Y .
4. Si Y est un vecteur aléatoire de matrice de variance covariance non-inversible, peut-il avoir une densité ? Le vecteur (X_1, X_2, X_3) a-t-il une densité ?

Exercice 9. Soit (X, Y) un couple de variables aléatoires réelles à densité sur \mathbb{R}^2 . On suppose que X et Y sont indépendantes.

1. Montrer que

$$\mathbb{E}[X | Y] = \mathbb{E}[X].$$

2. Plus généralement, montrer que $\mathbb{E}[h(X, Y) | X] = \Phi(X)$, avec

$$\Phi(x) = \mathbb{E}[h(x, Y)].$$

Exercice 10 (LOIS UNIFORMES).

1. Soit (X, Y) un couple de variables aléatoires réelles. Montrer que (X, Y) suit la loi uniforme sur le carré $[0, 1]^2$ si et seulement si X et Y sont indépendantes et de loi uniforme sur $[0, 1]$.
2. On coupe un bâton au hasard en trois morceaux en utilisant deux v.a. indépendantes et uniformes sur $[0, 1]$ pour déterminer les points de coupe. Vérifier que les longueurs des trois morceaux ainsi obtenus sont des v.a. de même loi. Sont-elles indépendantes ? Quelle est la probabilité de pouvoir former un triangle avec ces trois morceaux ?
3. Deux amis se donnent rendez vous entre 12h et 13h, et arrivent indépendamment uniformément entre ces deux horaires. Calculer le temps moyen d'attente du premier arrivé.

4 Simulation

Exercice 11 (SUR L'ALGORITHME DU REJET).

1. On considère une fonction \tilde{g} , positive, et intégrable sur \mathbb{R} . Montrer que l'algorithme
 - Tirer X suivant la densité $a\tilde{g}$, où a est une constante de renormalisation ;
 - Tirer U suivant une loi uniforme sur $[0, \tilde{g}(X)]$;permet de tirer (X, U) suivant une loi uniforme sur l'ensemble

$$\mathcal{A} = \{(x, u) \in \mathbb{R} \times \mathbb{R}^+ : 0 \leq u \leq \tilde{g}(x)\}.$$

2. Réciproquement, si (X, U) suit une loi uniforme sur \mathcal{A} quelle est la loi de X ?

3. On considère maintenant deux densités f et g telles que $f(x) \leq cg(x)$ pour tout x . Montrer que l'algorithme « Tirer uniformément (X, U) sur \mathcal{A} (avec $\tilde{g} = cg$) jusqu'à ce que la marginale U soit intérieure à $f(X)$ » donne un vecteur (X, U) de loi uniforme sur

$$\mathcal{B} = \{(x, u) \in \mathbb{R} \times \mathbb{R}^+ : 0 \leq u \leq f(x)\}.$$

En déduire une interprétation graphique de l'algorithme du rejet.

Exercice 12 (SIMULATION DE LA LOI GAMMA). On rappelle que pour $a, \lambda > 0$, la loi Gamma $\Gamma(a, \lambda)$ est à densité, donnée par

$$f_{\Gamma(a, \lambda)} : z \mapsto \frac{\lambda^a z^{a-1}}{\Gamma(a)} e^{-\lambda z} \mathbf{1}_{\{z > 0\}}.$$

On suppose dans la suite que $a > 1$ et on note

$$g_a(z) = z^{a-1} e^{-z} \mathbf{1}_{\{z > 0\}} \quad \text{et} \quad \mathcal{D}_a = \{(x, y) \in \mathbb{R}_+^2 : x \leq \sqrt{g_a(y/x)}\}.$$

1. Calculer $\sup_{z > 0} g_a(z)$ et $\sup_{z > 0} z^2 g_a(z)$. En déduire que $\mathcal{D}_a \subset [0, x_a] \times [0, y_a]$, où $x_a = \left(\frac{a-1}{e}\right)^{\frac{a-1}{2}}$ et $y_a = \left(\frac{a+1}{e}\right)^{\frac{a+1}{2}}$.
2. Soit $(X, Y) \sim \mathcal{U}(\mathcal{D}_a)$ un couple uniformément distribué sur \mathcal{D}_a , i.e. (X, Y) possède la densité $\frac{1}{|\mathcal{D}_a|} \mathbf{1}_{\{0 \leq y\}} \mathbf{1}_{\{0 \leq x \leq \sqrt{g_a(y/x)}\}}$, où $|\mathcal{D}_a|$ désigne la surface de \mathcal{D}_a . Quelle est la loi de $W = \frac{Y}{X}$? En déduire que $|\mathcal{D}_a| = \frac{\Gamma(a)}{2}$. Conclure que $Z = \frac{W}{\lambda} \sim \Gamma(a, \lambda)$.
3. Comment simuler suivant les lois $\mathcal{U}(\mathcal{D}_a)$ et $\Gamma(a, \lambda)$?
4. On vient de voir que pour simuler suivant la loi $\Gamma(a, 1)$, il n'y a pas besoin de connaître la constante $\Gamma(a)$ qui permet de normaliser g_a pour obtenir la densité $f_{\Gamma(a, 1)}$. Est-ce que remplacer g_a par cg_a où $c > 0$ dans la méthode ci-dessus change son efficacité?

Exercice 13 (BORNE DE LETAC POUR LA MÉTHODE DU REJET). Soit p une densité de probabilité sur l'intervalle $[0, 1]$ suivant laquelle on souhaite simuler en utilisant un algorithme de rejet construit à l'aide d'une suite $((U_i, X_i))_{i \geq 1}$ de vecteurs aléatoires i.i.d. où les U_i sont uniformément réparties sur $[0, 1]$. Plus précisément, on suppose qu'il existe un ensemble d'acceptation \mathcal{A} tel que $\mathbb{P}((U_1, X_1) \in \mathcal{A}) > 0$ et que la loi conditionnelle de U_1 sachant $(U_1, X_1) \in \mathcal{A}$ possède la densité p . On note $N = \min\{i \geq 1 : (U_i, X_i) \in \mathcal{A}\}$ et B un sous-ensemble borélien de $[0, 1]$.

1. Quelle est la loi de N ? Et celle de U_N ?
2. Montrer que pour $n \in \mathbb{N}^*$, $\mathbb{P}(U_n \in B, N \geq n) = \mathbb{P}(U_n \in B) \mathbb{P}(N \geq n)$.
3. En déduire que $\mathbb{P}(U_N \in B) \leq \mathbb{P}(U_1 \in B) \mathbb{E}(N)$.
4. Conclure que $\mathbb{E}(N) \geq \sup\{\rho \geq 0 : \int_0^1 \mathbf{1}_{\{p(u) \geq \rho\}} du > 0\}$.

5 Exercice corrigé

Exercice 14 (COUPLES DE VARIABLES ALÉATOIRES, DENSITÉ ET INDÉPENDANCE).

1. Déterminer la constante c pour que la fonction $f(x, y) = c(x^2 + y^2) \exp(-\frac{x^2 + y^2}{2})$ soit une densité sur \mathbb{R}^2 . Si (X, Y) est un couple qui suit cette densité, déterminer ses lois marginales, ainsi que la covariance de X et Y . Les variables X et Y sont-elles indépendantes?
2. Soit (X, Y) un couple de v.a.r. de loi uniforme sur le disque $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. Montrer que les v.a.r. X et Y ont même loi et calculer leur densité. Sont-elles indépendantes?
3. Soit (X, Y) un vecteur aléatoire à densité. Montrer que $\mathbb{P}(X = Y) = 0$. Si X est une v.a.r, le vecteur (X, X) a-t-il une densité?

Solution. 1. En appliquant le théorème de Fubini à la fonction continue et positive f , on trouve :

$$\begin{aligned}\int_{\mathbb{R}^2} (x^2 + y^2) \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} (x^2 + y^2) \exp\left(-\frac{x^2 + y^2}{2}\right) dx \right) dy \\ &= 2\pi \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx + y^2 \int_{\mathbb{R}} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \right) \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= 2\pi \int_{\mathbb{R}} (1 + y^2) \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \\ &= 4\pi,\end{aligned}$$

où on a utilisé le fait que $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ est une densité dont le moment d'ordre 2 vaut 1. On en déduit la valeur de la constante de renormalisation : $c = 1/(4\pi)$.

On sait que si (X, Y) est un vecteur aléatoire de densité f sur \mathbb{R}^2 , alors X et Y admettent également des densités f_X et f_Y données par les formules

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx.$$

Par symétrie de f , il est clair qu'ici $f_X = f_Y$ (autrement dit X et Y ont la même loi). Calculons f_X : pour tout $x \in \mathbb{R}$,

$$f_X(x) = \frac{1}{4\pi} \int_{\mathbb{R}} (x^2 + y^2) \exp\left(-\frac{x^2 + y^2}{2}\right) dy = \frac{1}{2\sqrt{2\pi}} (x^2 + 1) e^{-x^2/2}.$$

Ainsi

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx = \frac{1}{2\sqrt{2\pi}} \int_{\mathbb{R}} x(x^2 + 1) e^{-x^2/2} dx = 0$$

par imparité. Calculons $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY)$:

$$\text{Cov}(X, Y) = \frac{1}{4\pi} \int_{\mathbb{R}^2} xy(x^2 + y^2) \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy = 0,$$

toujours par imparité. Finalement, pour tout $(x, y) \in \mathbb{R}^2$,

$$f_X(x)f_Y(y) = \frac{1}{8\pi} (x^2 + 1)(y^2 + 1) e^{-(x^2 + y^2)/2}$$

et donc $f(x, y) \neq f_X(x)f_Y(y)$ et on en déduit que X et Y ne sont pas indépendantes.

2. Notons $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$. La densité d'un vecteur aléatoire (X, Y) uniformément distribué sur D est $f_{(X,Y)} = \frac{1}{\pi} \mathbb{1}_D$. Déterminons les lois marginales : si $|x| \leq 1$,

$$f_X(x) = \frac{1}{\pi} \int_{\mathbb{R}} \mathbb{1}_{x^2 + y^2 \leq 1} dy = \frac{1}{\pi} \int_{\mathbb{R}} \mathbb{1}_{|y| \leq \sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1-x^2},$$

et si $|x| > 1$, on voit que $f_X(x) = 0$. Par symétrie, $f_Y = f_X$. La loi commune de ces deux variables est appelée « loi du demi-cercle ».

3. En notant f une densité du couple (X, Y) , on a en appliquant la formule de transfert

$$\begin{aligned}\mathbb{P}(X = Y) &= \int_{\mathbb{R}^2} \mathbb{1}_{\{(x,y) \in \mathbb{R}^2 : x=y\}} f(x, y) dx dy \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{(x,y) \in \mathbb{R}^2 : x=y\}} f(x, x) dx dy \\ &= \int_{\mathbb{R}^2} f(x, x) \left(\int_{\mathbb{R}} \mathbb{1}_{\{(x,y) \in \mathbb{R}^2 : x=y\}} dy \right) dx \\ &= \int_{\mathbb{R}^2} f(x, x) \left(\int_{\{x\}} 1 dy \right) dx = 0,\end{aligned}$$

où la troisième égalité vient du théorème de Fubini, et la dernière du fait que le singleton $\{x\}$ est de mesure nulle. Pour toute variable aléatoire X , on a $\mathbb{P}(X = X) = 1$ et, donc d'après ce qui précède, le vecteur (X, X) n'a pas de densité sur \mathbb{R}^2 .