Valérie Girardin
Nikolaos Limnios

LICENCE 3
ÉCOLES D'INGÉNIEURS
CAPES & AGRÉGATION
MATHÉMATIQUES

Probabilités

et introduction à la statistique

- Cours complet
- Exercices d'application corrigés
- Problèmes de synthèse

Vuibert

Table des matières

CHAPITRE 3

Vecteurs aléatoires

Nous avons considéré dans le chapitre précédent des variables aléatoires réelles, c'est-à-dire des applications mesurables à valeurs dans $\mathbb{R}$. Nous présentons dans ce chapitre l'étude simultanée de plusieurs variables aléatoires réelles, c'est-à-dire celle des vecteurs aléatoires réels.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Toute fonction borélienne

$$\begin{array}{rcl} X = (X_1, \ldots, X_d)' \colon (\Omega, \mathcal{F}, \mathbb{P}) & \longrightarrow & (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \\ \omega & \longrightarrow & (X_1(\omega), \ldots, X_d(\omega))' \end{array}$$

est appelée d-vecteur aléatoire réel si $d > 1$. Les variables aléatoires réelles correspondent au cas $d = 1$.

Il est facile de voir que X est un vecteur aléatoire si et seulement si toutes les applications coordonnées, X_j, pour $j = 1, \ldots, d$, sont des variables aléatoires. Étudier X revient à étudier simultanément les d phénomènes ou variables $(X_1, \ldots, X_d)$ liés à une même expérience aléatoire. Lorsqu'il n'y a pas d'ambiguïté, nous noterons indifféremment $(X_1, \ldots, X_d)$ ou $(X_1, \ldots, X_d)'$.

Exemple 3.1 Mesure simultanée de caractéristiques physiques comme le poids, la taille, etc., au sein d'une population. $\triangleleft$

Les vecteurs aléatoires réels ont beaucoup de propriétés similaires à celles des variables aléatoires réelles. Nous leur étendons les notions vues dans le chapitre 2. Nous y ajoutons celles liées aux relations des variables entre elles comme la covariance, l'indépendance, l'entropie, les statistiques d'ordre... Nous insistons sur le calcul effectif de la loi de fonctions de plusieurs variables aléatoires et sur le cas des vecteurs gaussiens.

3.1 Relations entre variables aléatoires

Nous commençons par des relations entre variables aléatoires, nécessaires à l'étude des vecteurs et intéressantes en elles-mêmes dans les applications.

3.1.1 Covariance

La définition de la variance pour les vecteurs nécessite celle de la covariance de variables aléatoires.

Définition 3.2 Soient X et Y deux variables aléatoires réelles de carré intégrable. Leur covariance est définie par

$$\mathbb{C}\mathrm{ov}\,(X,Y) = \mathbb{E}\,[(X-\mathbb{E}\,X)(Y-\mathbb{E}\,Y)] = \mathbb{E}\,(XY) - (\mathbb{E}\,X)(\mathbb{E}\,Y).$$

Leur coefficient de corrélation (linéaire) est défini par

$$\rho_{X,Y} = \frac{\mathbb{C}\mathrm{ov}\,(X,Y)}{\sqrt{\mathbb{V}\mathrm{ar}\,X}\sqrt{\mathbb{V}\mathrm{ar}\,Y}}.$$

△

Deux variables dont la covariance est nulle sont dites non-corrélées. L'espérance de leur produit est clairement égal au produit de leurs espérances.

▼ Propriétés de la covariance et du coefficient de corrélation

1. $\mathbb{C}\mathrm{ov}\,(X,X) = \mathbb{V}\mathrm{ar}\,X$ et $\mathbb{C}\mathrm{ov}\,(X,Y) = \mathbb{C}\mathrm{ov}\,(Y,X)$.
2. $\mathbb{C}\mathrm{ov}\,(aX+bY,Z) = a\,\mathbb{C}\mathrm{ov}\,(X,Z)+b\,\mathbb{C}\mathrm{ov}\,(Y,Z)$, pour $(a,b)\in\mathbb{R}^2$. Ainsi, $|\rho_{X,Y}| = |\rho_{aX+b,cY+d}|$, pour tous $a\in\mathbb{R}^*,\ c\in\mathbb{R}^*,\ b\in\mathbb{R},\ d\in\mathbb{R}$, et le coefficient de corrélation est indépendant de l'unité et de l'origine.
3. $\mathbb{C}\mathrm{ov}\,(X,Y)^2 \le \mathbb{V}\mathrm{ar}\,X\cdot\mathbb{V}\mathrm{ar}\,Y$, par l'inégalité de Cauchy-Schwarz et donc $\rho_{X,Y}\in[-1,1]$. ▲

Le coefficient de corrélation de deux variables mesure leur degré de linéarité.

Proposition 3.3 Si X et Y sont deux variables aléatoires réelles de carré intégrable,
$|\rho_{X,Y}| = 1$ si et seulement si $Y = aX+b$, avec $(a,b)\in\mathbb{R}^*\times\mathbb{R}$.

Démonstration.- Nous pouvons supposer que $\mathbb{E}\,X = 0$. Si $Y = aX+b$, alors $\mathbb{E}\,Y = b$ et $\mathbb{C}\mathrm{ov}\,(X,Y) = a\mathbb{E}\,(X^2) = a\mathbb{V}\mathrm{ar}\,X$, et donc $|\rho_{X,Y}| = 1$.
Réciproquement, nous pouvons supposer aussi que $\mathbb{E}\,Y = 0$. Ainsi $|\rho_{X,Y}| = 1$ implique $[\mathbb{E}\,(XY)]^2 = \mathbb{E}\,(X^2)\mathbb{E}\,(Y^2)$. Par conséquent,

$$\mathbb{V}\mathrm{ar}\left[Y - \frac{\mathbb{E}\,(XY)}{\mathbb{E}\,(X^2)}X\right] = \mathbb{V}\mathrm{ar}\,(Y) - a^2\mathbb{V}\mathrm{ar}\,(X) = 0,$$

c'est-à-dire que cette variable est constante p.s., d'où le résultat, avec précisément $a = \mathbb{C}\mathrm{ov}\,(X,Y)/\mathbb{V}\mathrm{ar}\,X$. □

Pour plus de détails, voir le paragraphe dédié à la droite de régression dans le chapitre 5.

3.1.2 Indépendance

L'indépendance de tribus et d'événements a été présentée dans le chapitre 1. La notion d'indépendance de variables aléatoires repose sur celle d'indépendance des tribus qu'elles engendrent.

Définition 3.4 Soient $X_i \colon (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\Omega_i, \mathcal{F}_i)$, pour $i = 1, \ldots, d$, des variables aléatoires.

Elles sont dites indépendantes si les tribus engendrées $\sigma(X_1), \ldots, \sigma(X_d)$ sont indépendantes, soit

$$\mathbb{P}(X_1 \in B_1, \ldots, X_d \in B_d) = \prod_{i=1}^{d} \mathbb{P}(X_i \in B_i),$$

pour tous $B_1 \in \mathcal{F}_1, \ldots, B_d \in \mathcal{F}_d$. △

Si $X_1, \ldots, X_d$ sont indépendantes alors $g_1(X_1), \ldots, g_d(X_d)$ sont aussi indépendantes, pour toutes fonctions boréliennes $g_1, \ldots, g_d$.

Des événements de $\mathcal{F}$ sont indépendants si et seulement si leurs fonctions indicatrices sont des variables aléatoires indépendantes. Et d sous-tribus $\mathcal{F}_1, \ldots, \mathcal{F}_d$ de $\mathcal{F}$ sont indépendantes si et seulement si les d variables aléatoires coordonnées $c_i : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\Omega_i, \mathcal{F}_i)$ le sont.

Si $\sigma(X_1, \ldots, X_d)$ et une sous-tribu $\mathcal{G}$ de $\mathcal{F}$ sont indépendantes, nous dirons que $X_1, \ldots, X_d$ et $\mathcal{G}$ sont indépendantes. Si $X_1, \ldots, X_d$ sont indépendantes et de même loi P, nous dirons qu'elles sont i.i.d. (indépendantes et identiquement distribuées) de loi P.

Des variables aléatoires sont dites indépendantes conditionnellement à un événement B (non négligeable) si elles sont indépendantes pour la probabilité $\mathbb{P}(\cdot \mid B)$.

Dans le cas discret, l'indépendance de deux variables X et Y de valeurs respectives $\{x_i : i \in I\}$ et $\{y_j : j \in J\}$ s'écrit

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j), \quad i \in I, j \in J.$$

Exemple 3.5 Deux personnes ont rendez-vous. Chacune arrive au hasard entre minuit et une heure du matin. Calculons la probabilité que la première arrivée attende l'autre au moins 1/4 d'heure.

Soient X et Y les instants d'arrivées des deux personnes. On calcule

$$\begin{aligned}
&\mathbb{P}(X + 15 < Y) + \mathbb{P}(Y + 15 < X) = \\
&= \quad 2\mathbb{P}(X < Y - 15) = 2\int_{15}^{60} \int_{0}^{y-15} \frac{1}{(60)^2} dxdy \\
&= \quad \frac{2}{(60)^2} \int_{15}^{60} (y - 15) dy = \frac{(45)^2}{(60)^2} = 9/16,
\end{aligned}$$

puisque X et Y sont i.i.d. de loi uniforme sur $[0, 60]$. ◁

Nous allons voir différents critères d'indépendance. Pour commencer, la notion de corrélation est liée à celle de dépendance.

Proposition 3.6 Des variables aléatoires indépendantes et de carré intégrable sont non corrélées.

Démonstration.- Pour des variables discrètes.
Posons $X = \sum_{i\in I} x_i \mathbb{1}_{A_i}$ où $A_i = (X = x_i)$ et $Y = \sum_{j\in J} y_j \mathbb{1}_{B_j}$ où $B_j = (Y = y_j)$. Nous avons donc

$$XY = \sum_{(i,j)\in I\times J} x_i y_j \mathbb{1}_{A_i} \mathbb{1}_{B_j} = \sum_{(i,j)\in I\times J} x_i y_j \mathbb{1}_{A_i\cap B_j}.$$

Comme l'espérance de X et celle de Y sont finies, nous avons

$$\begin{aligned}
\mathbb{E}\,(XY) &= \sum_{(i,j)\in I\times J} x_i y_j \mathbb{P}(A_i\cap B_j) \stackrel{(1)}{=} \sum_{i\in I}\sum_{j\in J} x_i y_j \mathbb{P}(A_i)\mathbb{P}(B_j) \\
&= \Big[\sum_{i\in I} x_i \mathbb{P}(A_i)\Big]\Big[\sum_{j\in J} y_j \mathbb{P}(B_j)\Big] = (\mathbb{E}\,X)(\mathbb{E}\,Y).
\end{aligned}$$

(1) par indépendance de X et Y. □

Par contre, deux variables non corrélées ne sont pas nécessairement indépendantes, comme le montre le contre-exemple suivant.

Exemple 3.7 Soit $Y \sim \mathcal{U}(0,1)$. Par l'exemple 2.57, on connaît les moments de Y. Si $X = Y(Y+a)$ où $a \in \mathbb{R}$, on a

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\,(X,Y) &= \mathbb{E}\,(Y^3) + a\mathbb{E}\,(Y^2) - (\mathbb{E}\,Y)\mathbb{E}\,(Y^2) - a(\mathbb{E}\,Y)^2 \\
&= \frac{1}{4} + \frac{a}{3} - \frac{1}{6} - \frac{a}{4} = \frac{1+a}{12}.
\end{aligned}$$

Si $a = -1$, ces variables sont non corrélées et elles ne sont évidemment pas indépendantes puisque $X = Y(Y-1)$ est une fonction de Y, mesurable pour la tribu engendrée par Y. ◁

Nous pouvons définir maintenant la loi conditionnelle pour des variables discrètes.

Définition 3.8 Soient X et T deux variables aléatoires discrètes de valeurs respectives $\{x_i : i \in I\}$ et $\{t_j : j \in J\}$. La loi conditionnelle de X sachant $(T = t_j)$ est donnée par $\{\mathbb{P}(X = x_i \mid T = t_j) : i \in I\}$. △

Exemple 3.9 *Suite de l'exemple 1.28.-* Soient X et T les fonctions indicatrices respectives des événements A et B. Elles suivent chacune une loi de Bernoulli de paramètre 1/2. La loi conditionnelle de X sachant T est donnée par

$$\mathbb{P}(X = i \mid T = j) = \frac{\mathbb{P}(X = i, T = j)}{\mathbb{P}(T = j)}, \quad i,j = 0,1.$$

Par exemple, $(X=0, T=0)$ ="le résultat est impair et supérieur à 3"="le résultat est 5" donc $\mathbb{P}(X=0 \mid T=0)=\mathbb{P}(\overline{B} \mid \overline{A})=1/3$. De même, on calcule $\mathbb{P}(X=1 \mid T=1)=1/3$ et $\mathbb{P}(X=1 \mid T=0)=\mathbb{P}(X=0 \mid T=1)=2/3$. $\triangleleft$

Pour des variables discrètes, on peut faire ici directement le lien entre indépendance et loi conditionnelle à partir de leur définition.

Proposition 3.10 Deux variables aléatoires X et Y discrètes de valeurs respectives $\{x_i : i \in I\}$ et $\{y_j : j \in J\}$ sont indépendantes si et seulement si la loi conditionnelle de X sachant $(Y=y_j)$ est égale à la loi de X pour tout j.

La proposition peut être énoncée symétriquement en considérant les lois conditionnelles de Y.

L'indépendance dépendant étroitement de la probabilité choisie, elle n'implique aucune indépendance conditionnelle en général, comme le montre le contre-exemple suivant.

Exemple 3.11 Soient X_1 et X_2 i.i.d. de loi $\mathcal{B}(1/2)$ à valeurs dans $\{-1,1\}$. On a

$$\begin{aligned}\mathbb{P}(X_1=1 \mid X_1+X_2=0) &= \frac{\mathbb{P}(X_1=1, X_1+X_2=0)}{\mathbb{P}(X_1+X_2=0)} \\ &= \frac{\mathbb{P}(X_1=1)\mathbb{P}(X_2=-1)}{\mathbb{P}(X_1+X_2=0)}=\frac{1/2.1/2}{1/2}=\frac{1}{2}.\end{aligned}$$

De même, par symétrie, $\mathbb{P}(X_2=1 \mid X_1+X_2=0)=1/2$, mais $\mathbb{P}(X_1=1, X_2=1 \mid X_1+X_2=0)=0$, donc X_1 et X_2 ne sont pas indépendantes conditionnellement à X_1+X_2.

Notons que $\mathbb{P}(X_1=-1 \mid X_1+X_2=-2)=1$. $\triangleleft$

3.1.3 Relation d'ordre stochastique

Les relations d'ordre stochastiques sont en particulier utilisées pour étudier la fiabilité ou la durée de vie de systèmes. Nous ne définirons ici que la plus simple d'entre elles.

Définition 3.12 Soient X et Y deux variables aléatoires réelles. La variable X est dite stochastiquement plus grande que la variable Y, et nous noterons $X \overset{st}{\geq} Y$, si

$$\mathbb{P}(X>x) \geq \mathbb{P}(Y>x), \quad x \in \mathbb{R}. \tag{3.1}$$

Autrement dit, $X \overset{st}{\geq} Y$ si la fiabilité de X est supérieure à celle de Y, ou encore $R_X(t) \geq R_Y(t)$, pour tout $t \in \mathbb{R}$, voir la définition 2.87 page 70.

Exemple 3.13 Soient $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\mu)$. Si $\lambda \leq \mu$, alors $X \overset{st}{\geq} Y$. Par contre, si $X \sim \mathcal{N}(0, \sigma^2)$ et $Y \sim \mathcal{N}(0, \tau^2)$, avec $\sigma \geq \tau$, alors $X \overset{st}{\geq} Y$. $\triangleleft$

Exemple 3.14 *Systèmes en série et en parallèle.*- Considérons deux systèmes comprenant deux composants, voir la figure 3.1.

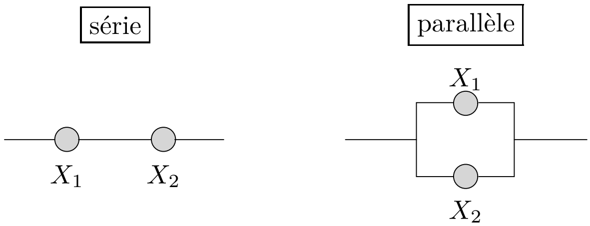

FIGURE 3.1. Systèmes à deux composants, en série et en parallèle.

Soient T_1 et T_2 les durées de vie des composants supposées indépendantes, de fonctions de répartition F_1 et F_2 respectivement.

Si les deux composants sont montés en série, la durée de vie du système est $T = \min(T_1, T_2)$, de fonction de répartition $F_T(t) = 1 - [1 - F_1(t)][1 - F_2(t)]$.

Si les deux composants sont montés en parallèle, la durée de vie du système est $T' = \max(T_1, T_2)$, de fonction de répartition $F_{T'}(t) = F_1(t)F_2(t)$.

Or $F_1(t) + F_2(t) - 2F_1(t)F_2(t) = F_1(t)[1 - F_2(t)] + F_2(t)[1 - F_1(t)]$, dont on déduit que $T' \overset{st}{\geq} T$. Autrement dit, $\max(T_1, T_2) \overset{st}{\geq} \min(T_1, T_2)$. $\triangleleft$

Proposition 3.15 Soient X et Y deux variables aléatoires réelles. Si $X \overset{st}{\geq} Y$ alors $\mathbb{E}\, X \geq \mathbb{E}\, Y$.

Démonstration.- Supposons d'abord que X et Y sont positives. Par intégration des deux membres de l'inégalité (3.1) et en utilisant l'expression alternative de l'espérance (2.7) page 76 pour $p = 1$, nous obtenons $\mathbb{E}\, X \geq \mathbb{E}\, Y$. Pour des variables quelconques, si $X \overset{st}{\geq} Y$ alors $X^+ \overset{st}{\geq} Y^+$ et $Y^- \overset{st}{\geq} X^-$, donc $\mathbb{E}\, X^+ \geq \mathbb{E}\, Y^+$ et $\mathbb{E}\, Y^- \geq \mathbb{E}\, X^-$ d'où $\mathbb{E}\, X \geq \mathbb{E}\, Y$. $\square$

Cette condition nécessaire peut être transformée en condition suffisante en considérant des espérances plus générales.

Théorème 3.16 Soient X et Y deux variables aléatoires réelles. On a $X \overset{st}{\geq} Y$ si et seulement si $\mathbb{E}\,[h(X)] \geq \mathbb{E}\,[h(Y)]$ pour toute fonction borélienne h croissante.

Démonstration.- Soit h^{-1} l'inverse généralisée de h, définie par $h^{-1}(y) = \inf\{t \in \mathbb{R} \colon h(t) > y\}$. Nous avons

$$\mathbb{P}[h(X) > x] = \mathbb{P}[X > h^{-1}(x)] \geq \mathbb{P}[Y > h^{-1}(x)] = \mathbb{P}[h(Y) > x], \quad x \in \mathbb{R},$$

donc $h(X) \overset{st}{\geq} h(Y)$, d'où $\mathbb{E}[h(X)] \geq \mathbb{E}[h(Y)]$ en appliquant la proposition 3.15.
La réciproque est immédiate en utilisant la fonction croissante $\mathbb{1}_{[x,+\infty[}$. $\square$

Par contre, $X \overset{st}{\geq} Y$ n'implique clairement pas que $X(\omega) \geq Y(\omega)$ pour tout $\omega \in \Omega$.

3.1.4 Entropie

Nous avons déjà vu différents paramètres qui renseignent sur la loi d'une variable aléatoire ou sur les relations entre deux variables et qui ont une interprétation physique, comme l'espérance, la variance ou la covariance. L'entropie, l'entropie jointe et l'information de Kullback-Leibler que nous présentons maintenant renseignent sur le degré d'incertitude du phénomène représenté par la variable, globalement ou relativement à un autre.

Définition 3.17 Soient X et Y des variables discrètes de valeurs respectives $\{x_i \,:\, i \in I\}$ et $\{y_j \,:\, j \in J\}$.

L'entropie (de Shannon) de X est l'entropie de sa loi, soit

$$\mathbb{S}(X) = -\sum_{i\in I} \mathbb{P}(X = x_i) \log \mathbb{P}(X = x_i).$$

L'entropie (jointe) de (X, Y) est

$$\mathbb{S}(X, Y) = -\sum_{i\in I}\sum_{j\in J} \mathbb{P}(X = x_i, Y = y_j) \log \mathbb{P}(X = x_i, Y = y_j).$$

L'entropie conditionnelle de X par rapport à Y est

$$\mathbb{S}(X \mid Y) = \mathbb{S}(X, Y) - \mathbb{S}(Y).$$

Si $\{x_i \,:\, i \in I\} \subset \{y_j \,:\, j \in J\}$, l'information de Kullback-Leibler de X par rapport à Y est celle de leurs lois, soit

$$\mathbb{K}(X \mid Y) = \sum_{i\in I} \mathbb{P}(X = x_i) \log \frac{\mathbb{P}(X = x_i)}{\mathbb{P}(Y = x_i)}.$$

$\triangle$

On peut aussi écrire

$$\mathbb{S}(X \mid Y) = -\sum_{i\in I}\sum_{j\in J} \mathbb{P}(X = x_i, Y = y_j) \log \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)}, \qquad (3.2)$$

ou obtenir $\mathbb{S}(X \mid Y)$ comme une moyenne, sous la forme

$$\mathbb{S}(X \mid Y) = \sum_{j\in J} \mathbb{P}(Y = y_j) \mathbb{K}(P_j \mid \mathbb{P}_Y),$$

où P_j est la loi conditionnelle de X sachant $(Y = y_j)$ et $\mathbb{K}(P_j \mid \mathbb{P}_Y)$ l'information de Kullback-Leibler de P_j par rapport à $\mathbb{P}_Y$. On voit ainsi que l'entropie conditionnelle entre deux variables discrètes est une quantité positive.

Exemple 3.18 L'entropie d'une variable aléatoire de loi $\mathcal{G}(p)$ est

$$\begin{aligned}
-\sum_{k\geq 1} pq^{k-1}\log(pq^{k-1}) &= -p\log p\sum_{k\geq 1} q^{k-1} - p\log q\sum_{k\geq 1}(k-1)q^{k-1}\\
&= -\log p - \frac{q}{p}\log q,
\end{aligned}$$

avec $q = 1 - p$. $\triangleleft$

▼ Propriétés de l'entropie

1. L'entropie d'une variable qui prend un nombre N fini de valeurs est maximale (et égale à $\log N$) si celle-ci suit une loi uniforme, comme nous l'avons vu dans le chapitre 1.
2. $\mathbb{S}(X \mid Y) = 0$ si et seulement si $X = g(Y)$. En considérant l'entropie conditionnelle sous la forme (3.2), si elle est nulle, nécessairement $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(Y = y_j)$ pour tous i et j, c'est-à-dire que X est une fonction de Y. Et la réciproque est claire.
3. $\mathbb{S}(X \mid Y) \leq \mathbb{S}(X)$, avec égalité si et seulement si les variables sont indépendantes. En effet, on peut écrire

$$\begin{aligned}
\mathbb{S}(X) &= -\sum_{i\in I}\sum_{j\in J}\mathbb{P}(X = x_i, Y = y_j)\log\mathbb{P}(X = x_i)\\
&= \sum_{i\in I}\sum_{j\in J}\mathbb{P}(X = x_i, Y = y_j)\log\frac{\mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)}{\mathbb{P}(X = x_i, Y = y_j)}
\end{aligned}$$

donc, comme $\log x \leq x - 1$ pour tout réel $x > 0$, on a

$$\mathbb{S}(X) \leq \sum_{i\in I}\sum_{j\in J}\mathbb{P}(X = x_i, Y = y_j)\left(\frac{\mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)}{\mathbb{P}(X = x_i, Y = y_j)} - 1\right) = 0,$$

et l'égalité est équivalente à $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$, pour tous $i \in I$ et $j \in J$, c'est-à-dire à l'indépendance des variables.
4. $\mathbb{S}(X, Y) \leq \mathbb{S}(X) + \mathbb{S}(Y)$, avec égalité si et seulement si les variables sont indépendantes, d'après la définition et le point 3. ▲

Les propriétés 2. et 3. s'interprètent comme suit : l'information apportée sur X par la connaissance de Y diminue l'incertitude sur X. À la limite (si X est fonction de Y), cette incertitude devient nulle.

Exemple 3.19 *Interprétation en termes de communication.*- Un système est constitué d'une source, d'un canal de transmission et d'un récepteur. Une partie de ce qui est produit par la source (le message) est modifié pendant la transmission au récepteur par la présence de bruit dans le canal de transmission.

L'entropie est un moyen d'évaluation du taux de transmission de l'information par le canal. Si la variable X représente la source et la variable Y le récepteur, $\mathbb{S}(X)$ est l'entropie de la source, $\mathbb{S}(Y)$ celle du récepteur et $\mathbb{S}(X,Y)$ celle du système global. L'entropie conditionnelle $\mathbb{S}(Y \mid X)$ représente l'entropie du récepteur connaissant le message émis par la source, c'est une mesure du bruit. Enfin $\mathbb{S}(X \mid Y)$ est celle de la source connaissant le message reçu. Elle mesure ce que l'on peut espérer apprendre du message transmis à partir du message reçu. Une mesure de la capacité de transmission du canal est donnée par $C = \max[\mathbb{S}(X) - \mathbb{S}(X \mid Y)]$, où le maximum est pris sur toutes les lois possibles pour X et Y.

En l'absence de bruit, on a $\mathbb{P}(X = x_i, Y = y_j) = 0$ si $i \neq j$, donc $\mathbb{S}(X \mid Y) = \mathbb{S}(Y \mid X) = 0$ et $\mathbb{S}(X,Y) = \mathbb{S}(X) = \mathbb{S}(Y)$. L'incertitude du récepteur est identique à celle de la source. La perte d'information par le canal de transmission est minimum et $C = \max \mathbb{S}(X) = \log N$.

Par contre, lorsqu'il n'existe pas de corrélation entre le message transmis et le message reçu (c'est-à-dire si tout x_i peut être transformé en un y_j quelconque avec la même probabilité), alors $\mathbb{P}(X = x_i, Y = y_j) = p_i$ pour tout j, avec $\sum_{i\in I} p_i = 1/N$. On en déduit que $\mathbb{S}(X \mid Y) = \mathbb{S}(X)$ et que $\mathbb{S}(Y \mid X) = \mathbb{S}(Y)$. Le système ne transmet aucune information, la perte d'information par le canal de transmission est maximale et $C = 0$. $\lhd$

L'entropie de Shannon peut être étendue à des variables à densité. L'interprétation en termes d'information est plus difficile.

Définition 3.20 Soient X et Y des variables aléatoires réelles de densités respectives f_X et f_Y strictement positives.

L'entropie de X est celle de sa loi, soit

$$\mathbb{S}(X) = \mathbb{E}\,[-\log f_X(X)] = -\int_{\mathbb{R}} f_X(u) \log f_x(u) du.$$

L'information de Kullback-Leibler de X par rapport à Y est celle de leurs lois, soit

$$\mathbb{K}(X \mid Y) = \int_{\mathbb{R}} f_X(u) \log \frac{f_X(u)}{f_Y(u)} du.$$

$\triangle$

Exemple 3.21 L'entropie d'une variable X de loi $\mathcal{E}(\lambda)$ est

$$\int_{\mathbb{R}_+} \lambda e^{-\lambda x} \log(\lambda e^{-\lambda x}) dx = \log\lambda \int_{\mathbb{R}_+} \lambda e^{-\lambda x} dx - \lambda \int_{\mathbb{R}_+} x\lambda e^{-\lambda x} dx = \log\lambda + 1,$$

puisque $\int_{\mathbb{R}_+} f_X(x) dx = 1$ et que $\mathbb{E}\, X = 1/\lambda$. $\lhd$

Proposition 3.22 L'information de Kullback-Leibler est positive, et elle est nulle si et seulement si les variables ont la même loi.

Démonstration.- Comme $\int_{\mathbb{R}} f_Y(u)du = \int_{\mathbb{R}} f_X(u)du$, nous pouvons écrire

$$\mathbb{K}(X \mid Y) = \int_{\mathbb{R}} \frac{f_X(u)}{f_Y(u)} \log\left[\frac{f_X(u)}{f_Y(u)}\right] f_Y(u)du + \int_{\mathbb{R}} \left[1 - \frac{f_X(u)}{f_Y(u)}\right] f_Y(u)du.$$

Et la conclusion en découle puisque $x \log x \geq 1 - x$ avec égalité seulement si $x = 1$. □

L'entropie $\mathbb{S}(X)$ peut être négative ou infinie pour des variables à densité. La notion de maximum n'est donc pas envisageable directement ; par contre, on peut étudier sa maximisation sous contraintes, c'est-à-dire les méthodes de maximum d'entropie.

Exemple 3.23 Le maximum de $\mathbb{S}(X)$ pour $a \leq X \leq b$ où $a < b$ sont des réels fixés est obtenu pour une variable de loi $\mathcal{U}(a, b)$ et est égal à $\log(b - a)$.

Le maximum de $\mathbb{S}(X)$ pour $X \geq 0$ et $\mathbb{E}\, X = a$ où $a \neq 0$ est un réel fixé est obtenu pour une variable de loi $\mathcal{E}(1/a)$ et est égal à $\log(ea)$.

Le maximum de $\mathbb{S}(X)$ pour $\mathbb{E}\, X = 0$ et $\mathbb{E}\,(X^2) = \sigma^2$ où $\sigma^2 \in \mathbb{R}_+^*$ est fixé est obtenu pour une variable de loi $\mathcal{N}(0, \sigma^2)$ et est égal à $\log(\sqrt{2\pi e}\sigma)$.

La démonstration de ces résultats est basée sur la méthode des multiplicateurs de Lagrange. Leur généralisation à des vecteurs est traitée dans l'exercice 3.2. ◁

3.2 Caractéristiques des vecteurs aléatoires

On ne peut en général pas déterminer la loi de la somme ou du produit de variables $X_1, \ldots, X_d$ non indépendantes si l'on connaît seulement la loi marginale de chaque variable. Plus généralement, il faut connaître la loi du vecteur $(X_1, \ldots, X_d)$, variable aléatoire à valeurs dans $\mathbb{R}^d$, pour calculer la loi de l'une de ses fonctions.

3.2.1 Produit d'espaces de probabilité

Pour étudier des événements ou des variables aléatoires liés à plusieurs expériences aléatoires se produisant simultanément, nous devons d'abord construire un espace probabilisé qui décrit l'ensemble de ces expériences. Présentons cette construction d'abord pour des espaces de probabilité dénombrables.

Soient deux espaces de probabilité $(\Omega_1, \mathcal{P}(\Omega_1), \mathbb{P}_1)$ et $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_2)$ finis ou dénombrables décrivant deux expériences aléatoires particulières et un observateur qui observe les deux expériences en même temps. L'issue d'une telle expérience se présente comme un couple de points (ω_1, ω_2) avec $\omega_1 \in \Omega_1$ et $\omega_2 \in \Omega_2$. Autrement dit, l'espace fondamental de cette expérience est l'espace produit cartésien $\Omega_1 \times \Omega_2$, noté Ω et l'ensemble des événements est $\mathcal{P}(\Omega) = \mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2)$. Sur cet espace nous pouvons définir une probabilité produit $\mathbb{P}$, notée $\mathbb{P}_1 \otimes \mathbb{P}_2$ et appelée produit tensoriel de $\mathbb{P}_1$ et $\mathbb{P}_2$, par les relations

$$\mathbb{P}(\{\omega\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}), \quad \omega = (\omega_1, \omega_2) \in \Omega,$$

et

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}), \quad A \subset \Omega.$$

Tout événement lié à la première expérience peut se représenter soit comme un sous-ensemble A_1 de Ω_1, soit comme le sous-ensemble $A_1 \times \Omega_2$ de Ω et

$$\mathbb{P}(A_1 \times \Omega_2) = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in \Omega_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) = \mathbb{P}_1(A_1).$$

De plus,

$$\begin{aligned}\mathbb{P}(A_1 \times A_2) &= \mathbb{P}[(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)] = \sum_{\omega_1 \in A_1} \sum_{\omega_2 \in A_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) \\ &= \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) = \mathbb{P}(A_1 \times \Omega_2)\mathbb{P}(\Omega_1 \times A_2).\end{aligned}$$

Le produit de n espaces probabilisés quelconques se définit de manière similaire.

Définition 3.24 Soient $(\Omega_1, \mathcal{F}_1), \ldots, (\Omega_n, \mathcal{F}_n)$ des espaces mesurables.
Le produit $\prod_{i=1}^n (\Omega_i, \mathcal{F}_i)$ est défini comme $(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{F}_i)$, où :
l'ensemble $\prod_{i=1}^n \Omega_i = \{(\omega_1, \ldots, \omega_n) \,:\, \omega_i \in \Omega_i\}$ est le produit cartésien,
la tribu $\otimes_{i=1}^n \mathcal{F}_i = \sigma(\{A_1 \times \cdots \times A_n \,:\, A_i \in \mathcal{F}_i\})$ est la tribu engendrée par les pavés. △

C'est la plus petite tribu rendant les projections sur les coordonnées mesurables.

Nous noterons $\prod_{i=1}^n (\Omega, \mathcal{F}) = (\Omega^n, \mathcal{F}^{\otimes n}) = (\Omega, \mathcal{F})^n$ avec en particulier :
$\mathcal{P}(\Omega) \times \mathcal{P}(\Omega) = \mathcal{P}(\Omega \times \Omega)$ pour tout Ω fini ;
$(\mathbb{R}, \mathcal{B}(\mathbb{R}))^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, soit $\mathcal{B}(\mathbb{R})^{\otimes n} = \mathcal{B}(\mathbb{R}^n)$.

Nous admettrons le résultat suivant qui définit une probabilité sur cet espace mesurable.

Définition-Théorème 3.25 Soient $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1), \ldots, (\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, des espaces de probabilité. Il existe une unique probabilité $\mathbb{P}$, appelée probabilité produit, définie sur $\prod_{i=1}^n (\Omega_i, \mathcal{F}_i)$ et vérifiant

$$\mathbb{P}(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i),$$

pour tous $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$. △

sont indépendantes de lois respectives $\sigma^2\chi^2(r_1)$,…,$\sigma^2\chi^2(r_p)$.

Démonstration.- Nous pouvons toujours supposer que $M = 0$.
1. Notons $(e_1, \ldots, e_d)$ une base orthonormale de $\mathbb{R}^d$. Les coordonnées de X dans cette base s'écrivent $Y = (< X, e_1 >, \ldots, < X, e_d >)'$ et nous avons

$$\varphi_Y(t) = \mathbb{E}\Big(\exp i \sum_{j=1}^{d} t_j < X, e_j >\Big) = \mathbb{E}\Big(\exp i < X, \sum_{j=1}^{d} t_j e_j >\Big).$$

La fonction caractéristique de X est

$$\varphi_X(t) = \mathbb{E}\,(\exp i < t, X >) = \exp\Big(-\frac{1}{2}\sigma^2 \sum_{j=1}^{d} t_j^2\Big),$$

donc $\varphi_Y(t) = \exp(-\sigma^2 \sum_{j=1}^{d} t_j^2/2)$, c'est-à-dire que $Y \sim X$.
2. Si A est la matrice d'une isométrie de $\mathbb{R}^d$, ses lignes (et ses colonnes) sont orthonormées et nous avons $AA' = I_d$, d'où la conclusion par 1..
3. Nous savons qu'il existe une base orthonormale de $\mathbb{R}^d$, soit $(e_1, \ldots, e_d)$, telle que E_j soit engendré par $(e_{j_1}, \ldots, e_{j_{r_1}})$.
Dans cette base, $\Pi_{E_j}(X) = \sum_{k=r_{j-1}+1}^{r_j} < X, e_{j_k} > e_{j_k}$, de norme euclidienne $\|\Pi_{E_j}(X)\|^2 = \sum_{k=r_{j-1}+1}^{r_j} < X, e_{j_d} >^2$. En utilisant le vecteur aléatoire Y, nous obtenons le résultat. □

3.5 Exercices et compléments

3.5.1 Enoncés

Exercice 3.1 *Problème des rencontres (suite des exercices 1.3 et 2.3)*.- Soit X_i la variable indicatrice de l'événement "Le sportif i a repris son propre sac".

Exprimer S_n en fonction de $X_1, \ldots, X_n$ et en déduire l'espérance et la variance de S_n.

Solution de l'exercice 3.1

On a $S_n = X_1 + \cdots + X_n$. De plus, $\mathbb{P}(X_i = 1) = 1/n = 1 - \mathbb{P}(X_i = 0)$, donc $\mathbb{E}\,X_i = 1/n$ et $\mathbb{E}\,S_n = 1$. D'autre part, $\mathbb{V}\mathrm{ar}\,X_i = (n-1)/n^2$ et l'on a $\mathbb{P}(X_iX_j = 1) = 1/n(n-1) = 1 - \mathbb{P}(X_iX_j = 0)$ donc

$$\mathbb{C}\mathrm{ov}\,(X_i, X_j) = \frac{1}{n(n-1)} - \frac{1}{n}\cdot\frac{1}{n} = \frac{1}{n^2(n-1)}$$

et $\mathbb{V}\mathrm{ar}\,S_n = \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}\,X_i - \sum_{i\neq j} \mathbb{C}\mathrm{ov}\,(X_i, X_j) = 1$.

Exercice 3.2 *Maximum d'entropie sous contraintes.*- L'entropie de Shannon d'un vecteur aléatoire X de densité f est donnée par la définition 3.17.

1. Calculer l'entropie d'un vecteur gaussien $\widetilde{X} \sim \mathcal{N}_n(0, \Gamma)$ non dégénéré.

2. Montrer que le maximum d'entropie est obtenu parmi les vecteurs aléatoires n-dimensionnels centrés de matrice de covariance Γ donnée pour le vecteur gaussien $\widetilde{X}$.

Solution de l'exercice 3.2

1. La densité de $\widetilde{X}$ est d'après la proposition 3.77,

$$f_{\widetilde{X}}(x) = (2\pi)^{-n/2}(\det\Gamma)^{-1/2}\exp[-\frac{1}{2}x'\Gamma^{-1}x], \quad x \in \mathbb{R}^n.$$

Son entropie est donc

$$\mathbb{S}(\widetilde{X}) = \frac{1}{2}\Big[n\log(2\pi) + \log(\det\Gamma) + \int_{\mathbb{R}^n} x'\Gamma^{-1}x f_{\widetilde{X}}(x)dx\Big].$$

Si l'on note $\mathrm{Tr}M = \sum_{i=1}^n M_{ii}$ la trace d'une matrice n-dimensionnelle, on a

$$\begin{aligned}\int_{\mathbb{R}^n} x'\Gamma^{-1}x f_{\widetilde{X}}(x)dx &= \mathbb{E}\,(\widetilde{X}'\Gamma^{-1}\widetilde{X}) = \mathbb{E}\,[\mathrm{Tr}(\widetilde{X}'\widetilde{X}\Gamma^{-1})] \\ &= \mathrm{Tr}[\mathbb{E}\,(\widetilde{X}'\widetilde{X})\Gamma^{-1}] = \mathrm{Tr}(\Gamma\Gamma^{-1}) = n.\end{aligned}$$

Par conséquent, $\mathbb{S}(\widetilde{X}) = \log[\sqrt{(2\pi e)^n\det\Gamma}]$.

2. On peut écrire

$$\mathbb{S}(X) = \int_{\mathbb{R}^n} f(x)\log\frac{f_{\widetilde{X}}(x)}{f(x)}dx - \int_{\mathbb{R}^n} f(X)\log f_{\widetilde{X}}(x)dx = (1) + (2).$$

Or $\log x \leq x - 1$ pour tout x positif, donc

$$0 \leq (1) \leq \int_{\mathbb{R}^n} f(x)\Big[\frac{f_{\widetilde{X}}(x)}{f(x)} - 1\Big]dx = 0,$$

et

$$(2) = \frac{1}{2}\Big[\log(2\pi) + \log(\det\Gamma) + \int_{\mathbb{R}^n} x'\Gamma^{-1}xf(x)dx\Big].$$

Comme dans 1., puisque l'on suppose que X vérifie $\mathbb{E}\,(X'X) = \Gamma$, on a

$$\int_{\mathbb{R}^n} x'\Gamma^{-1}xf(x)dx = \mathrm{Tr}(\Gamma\Gamma^{-1}) = n,$$

Ainsi, on a (2)= $\mathbb{S}(\widetilde{X})$ et la conclusion en découle.

Exercice 3.3 *Entropie mutuelle.*-

A. Soient X et Y deux variables aléatoires discrètes définies sur le même espace de probabilité et de valeurs respectives $\{x_i : i \in I\}$ et $\{y_j : j \in J\}$. Leur entropie (ou information) mutuelle est définie par

$$\mathcal{S}(X,Y) = \sum_{i\in I}\sum_{j\in J} \mathbb{P}(X = x_i, Y = y_j) \log \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)}.$$

Exprimer $\mathcal{S}(X,Y)$ en fonction des entropies jointe, marginales et conditionnelles de X et Y. En déduire que c'est une quantité positive. A quelle condition sur X et Y est-elle nulle ?

B. Soient X et Y deux variables aléatoires définies sur le même espace de probabilité, de densité jointe f et de densités marginales respectives f_X et f_Y.

1. Définir l'équivalent de $\mathcal{S}$ pour ces variables continues.

2. a. Utiliser la convexité de la fonction logarithme pour montrer que l'entropie mutuelle $\mathcal{S}(X,Y)$ est positive.

b. Utiliser le passage du discret au continu pour montrer qu'elle est finie.

Solution de l'exercice 3.3

A. On a $\mathcal{S}(X,Y) = \mathbb{S}(X) - \mathbb{S}(X \mid Y)$. Par définition, $\mathbb{S}(X \mid Y) = \mathbb{S}(X,Y) - \mathbb{S}(Y)$, donc $\mathcal{S}(X \mid Y) = \mathbb{S}(X) + \mathbb{S}(Y) - \mathbb{S}(X,Y)$ par la propriété 2. de l'entropie.

On sait que $\mathbb{S}(X \mid Y) \leq \mathbb{S}(X)$, donc $\mathcal{S}(X,Y)$ est bien positive, et nulle si et seulement si X et Y sont indépendantes.

B. 1. Si f_X et f_Y sont non nulles et de même support $E \in \mathcal{B}(\mathbb{R})$, on définit

$$\mathcal{S}(X,Y) = \iint_{E\times E} f(x,y) \log \frac{f(x,y)}{f_X(x)f_Y(y)} dxdy.$$

2. a. On peut écrire

$$\mathcal{S}(X,Y) = -\iint_{E\times E} f(x,y) \log \frac{f_X(x)f_Y(y)}{f(x,y)} dxdy.$$

Or $\log x \leq x - 1$ pour x positif, donc

$$\begin{aligned}
\mathcal{S}(X,Y) &\geq -\iint_{E\times E} f(x,y) \left[\frac{f_X(x)f_Y(y)}{f(x,y)} - 1\right] dxdy \\
&= -\iint_{E\times E} f_X(x)f_Y(y)dxdy + \iint_{E\times E} f(x,y)dxdy \\
&= -1 + 1 = 0.
\end{aligned}$$

b. Lorsque Δx et Δy tendent vers 0,

$$\log \frac{\mathbb{P}(X \in \Delta x, Y \in \Delta y)}{\mathbb{P}(X \in \Delta x)\mathbb{P}(Y \in \Delta y)} \longrightarrow \log \frac{f(x,y)}{f_X(x)f_Y(y)}$$

et reste finie contrairement à chacun des termes qui la composent. Ainsi, l'information mutuelle $\mathcal{S}(X, Y)$ est finie alors que les entropies $\mathbb{S}(X)$, $\mathbb{S}(Y)$ et $\mathbb{S}(X, Y)$ ne le sont pas nécessairement si les variables sont continues.

Exercice 3.4 *Le théorème du scrutin.-*

1. Une urne contient a boules portant le numéro 0 et b boules portant le numéro 2, avec $a > b$. On tire toutes les boules une par une et sans remise. Soit X_i la variable aléatoire égale au numéro de la boule tirée au i-ième tirage, pour $i = 1, \dots, n$, où $n = a + b$.

a. Calculer l'espérance de $S_m = \sum_{i=1}^{m} X_i$ pour $1 \leq m \leq n$.

b. Montrer que pour tout $0 \leq b' \leq b$,

$$\begin{aligned}&\mathbb{P}(S_m < m,\ m = 1, \dots, n \mid S_{2b} = 2b') = \\ &\quad = \ \mathbb{P}(S_m < m,\ m = 1, \dots, 2b \mid S_{2b} = 2b').\end{aligned}$$

c. En déduire par récurrence que

$$\mathbb{P}(S_m < m,\ m = 1, \dots, n \mid S_n = 2b) = 1 - 2b/n. \tag{3.10}$$

2. Un scrutin oppose deux candidats A et B et voit la victoire de A. Déduire de ce qui précède la probabilité que A ait été en tête tout au long du dépouillement des bulletins de vote.

Une illustration est donnée dans la figure 3.4. Notons que cela constitue un exemple des marches aléatoires qui seront définies dans le chapitre 4.

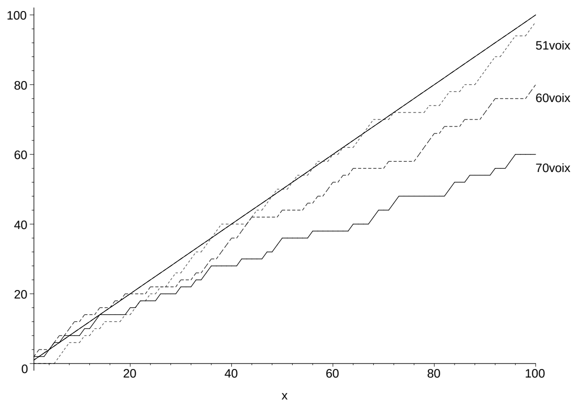

FIGURE 3.4. Illustration du théorème du scrutin.

Solution de l'exercice 3.4

1. a. On a $\mathbb{E}\, S_n = 2b$ car $\mathbb{P}(S_n = 2b) = 1$. De plus, les variables X_i sont échangeables (soit $(X_1, \dots, X_n) \sim (X_{(1)}, \dots, X_{(n)})$), donc $\mathbb{E}\, X_i = \mathbb{E}\, S_n/n = 2b/n$. Par conséquent, $\mathbb{E}\, S_m = 2bm/n$.

b. On sait que $S_n = 2b$, donc si $S_{2b} = 2b'$, nécessairement $S_m < m$ pour tout m tel que $n \geq m \geq 2b+1$, d'où l'égalité cherchée.

c. Pour que (3.10) soit vérifiée au rang 1, on doit avoir $a = 1$ et $b = 0$, et donc $\mathbb{P}(X_1 < 1) = \mathbb{P}(X_1 = 0) = 1$.

Supposons (3.10) vraie pour tout $m < n$. On a d'une part

$$\begin{aligned}&\mathbb{P}(S_m < m,\ m = 1, \ldots, n) = \\ &\quad = \sum_{b'=0}^{b} \mathbb{P}(S_m < m,\ m = 1, \ldots, n \mid S_{2b} = 2b')\mathbb{P}(S_{2b} = 2b').\end{aligned}$$

et d'autre part, puisque $\mathbb{P}(S_n = 2b) = 1$,

$$\mathbb{P}(S_n < m, m = 1, \ldots, n) = \mathbb{P}(S_n < m, m = 1, \ldots, n \mid S_n = 2b).$$

Par hypothèse de récurrence appliquée à $n = 2b$, on a

$$\mathbb{P}(S_m < m,\ m = 1, \ldots, 2b \mid S_{2b} = 2b') = 1 - b'/b.$$

Ainsi,

$$\begin{aligned}\mathbb{P}(S_m < m, m = 1, \ldots, n) &\overset{(1)}{=} \sum_{b'=0}^{b} (1 - b'/b)\mathbb{P}(S_{2b} = 2b') = 1 - \mathbb{E}\, S_{2b}/2b \\ &\overset{(2)}{=} 1 - \frac{2b}{n}.\end{aligned}$$

(1) par 1.b. et (2) par 1.a. et la démonstration est achevée.

2. En considérant les bulletins comme les boules de l'urne et le dépouillement comme le tirage de ces boules, la probabilité cherchée est $(a-b)/(a+b)$ où a est le nombre de voix obtenu par A et b celui obtenu par B.

Exercice 3.5 *Loi d'un couple.*- Considérons la fonction $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, définie par

$$F(x, y) = 1 - e^{-x} - e^{-y} + (e^x + e^y - 1)^{-1} \mathbb{1}_{\mathbb{R}_+^2}(x, y).$$

1. Montrer que F est la fonction de répartition d'un couple de variables aléatoires, soit (X, Y).

2. Donner la densité de probabilité $f(x, y)$ de (X, Y). Les variables sont-elles indépendantes ?

3. Calculer les fonctions de répartition marginales. Quelle est la loi de X et quelle est celle de Y ?

4. Soit $g : \mathbb{R}^2 \to \mathbb{R}^2$ définie par $g(x, y) = (e^{-x+y}, y + x)$. Déterminer la loi de $(U, V) = g(X, Y)$.

Solution de l'exercice 3.5

1. La fonction F doit vérifier les points 1. à 4. des propriétés des fonctions de répartition. De façon évidente, F est continue à droite pour chacun de ses arguments, $F(x,y)$ est nulle pour x ou y négatif et $F(x,y)$ tend vers 1 lorsque x et y tendent simultanément vers $+\infty$.

Montrons que F est croissante pour chacun de ses arguments. On a

$$\frac{\partial}{\partial x}F(x,y) = \frac{e^x(e^x+e^y-1)^2 - e^x}{(e^x+e^y-1)^2}.$$

Or $x \geq 0$ et $y \geq 0$ donc $e^x \geq 1$, et $e^y \geq 1$. Par conséquent, $(e^x+e^y-1)^2 \geq 1$ et F est croissante en x. Puisque F est symétrique en x et y, la propriété est vérifiée.

Il reste à montrer que $\Delta_{x_1x_2}\Delta_{y_1y_2}F(x,y) \geq 0$ pour $x_1 \leq x_2$ et $y_1 \leq y_2$. Or

$$\begin{aligned}
&\Delta_{x_1x_2}\Delta_{y_1y_2}F(x,y) = \\
&\quad = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \\
&\quad = -\frac{e^{x_2}-e^{x_1}}{(e^{x_2}+e^{y_2}-1)(e^{x_1}+e^{y_2}-1)} + \frac{e^{x_2}-e^{x_1}}{(e^{x_2}+e^{y_1}-1)(e^{x_1}+e^{y_1}-1)}.
\end{aligned}$$

On sait que $e^{x_2}-e^{x_1} \geq 0$ et

$$(e^{x_2}+e^{y_1}-1)(e^{x_1}+e^{y_1}-1) \leq (e^{x_2}+e^{y_2}-1)(e^{x_1}+e^{y_2}-1).$$

Par conséquent, F est bien une fonction de répartition bidimensionnelle.

2. La densité de probabilité est

$$f(x,y) = \frac{\partial^2}{\partial x \partial y}F(x,y) = \frac{2e^x e^y}{(e^x+e^y-1)^3} \neq f_1(x)f_2(y).$$

La densité jointe n'est pas le produit des densités marginales et d'après le théorème 3.41, les variables ne sont donc pas indépendantes.

3. On a

$$F_X(x) = \lim_{y\to+\infty} F(x,y) = 1 - e^{-x}, \quad x \geq 0.$$

Par symétrie, $F_Y(y) = 1 - e^{-y}$ pour $y \geq 0$. Donc X et Y suivent la même loi exponentielle de paramètre 1.

4. Posons $u = e^{-x+y}$ et $v = x+y$ ou encore $x = (v - \log u)/2$ et $y = (v + \log u)/2$, avec $u > 0$. Le jacobien du changement de variable est

$$\det\begin{pmatrix} -e^{-x+y} & 1 \\ e^{-x+y} & 1 \end{pmatrix} = -2e^{-x+y} = -2u,$$

donc la densité du couple (U,V) est

$$f_{(U,V)}(u,v) = [u(e^{(v-\log u)/2} + e^{(v+\log u)/2} - 1)^3]^{-1} e^{2v} \mathbb{1}_{D_2}(u,v),$$

où $D_2 = \{(u,v) \in \mathbb{R}_+^2 : e^{-v} < u < e^v\}$.

Valérie Girardin & Nikolaos Limnios

Probabilités

et introduction à la statistique

Rédigé principalement à l'attention des étudiants en 3e année de Licence de mathématiques et en écoles d'ingénieurs, cet ouvrage présente l'ensemble du programme de probabilités avec un cours complet, de nombreux exercices corrigés ainsi que des problèmes de synthèse.

D'une lecture aisée, ce manuel aborde, de manière rigoureuse, les fondements théoriques des probabilités et de leurs applications. Son contenu servira également de base de révision aux concours du CAPES (dans le cadre du Master MEEF) et de l'Agrégation de mathématiques.

Sommaire

À la fin de chaque chapitre, on trouvera des exercices suivis de leurs corrigés

Agrégée de mathématiques, **Valérie Girardin** est maître de conférences à l'Université de Caen Basse-Normandie, Laboratoire de Mathématiques Nicolas Oresme, et habilitée à diriger les recherches.

Nikolaos Limnios est professeur à l'Université de Technologie de Compiègne (UTC), Laboratoire de Mathématiques Appliquées.

ISBN 978-2-311-40014-4

9 782311 400144

www.VUIBERT.FR