

TD7. Vecteurs aléatoires.

1. Soient $\alpha, \beta \in]0, 1[$ deux réels. Pour tout $(i, j) \in \mathbb{N}^2$, on pose $p_{i,j} = \alpha\beta(1 - \alpha)^i(1 - \beta)^j$.
- Montrer qu'en posant $\mathbb{P}(\{(i, j)\}) = p_{i,j}$ pour tout $(i, j) \in \mathbb{N}^2$, on définit une mesure de probabilités sur \mathbb{N}^2 muni de la tribu $\mathcal{P}(\mathbb{N}^2)$.
 Pour tout $(i, j) \in \mathbb{N}^2$, on pose $X((i, j)) = i$ et $Y((i, j)) = j$.
 - Déterminer la loi de X et la loi de Y .
 - Calculer $\mathbb{P}(X < Y)$, $\mathbb{P}(X = Y)$ et $\mathbb{P}(X > Y)$.

Solution de l'exercice 1.

- Il s'agit de vérifier que $\forall (i, j) \in \mathbb{N}^2, p_{i,j} \geq 0$ et $\sum_{(i,j) \in \mathbb{N}^2} p_{i,j} = 1$. La positivité des $p_{i,j}$ est évidente d'après leur définition. Comme $\alpha, \beta \in]0, 1[$, on peut utiliser la somme d'une série géométrique : $\sum_{i \in \mathbb{N}} x^i = 1/(1 - x)$ avec $x = 1 - \alpha$ et $x = 1 - \beta$, on obtient :

$$\sum_{(i,j) \in \mathbb{N}^2} p_{i,j} = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \alpha\beta(1 - \alpha)^i(1 - \beta)^j = \alpha\beta \left(\sum_{i \in \mathbb{N}} (1 - \alpha)^i \right) \left(\sum_{j \in \mathbb{N}} (1 - \beta)^j \right) = 1.$$

- Remarquons que la loi jointe du vecteur (X, Y) est $p_{i,j}$. En effet,

$$\mathbb{P}[X = i, Y = j] = \mathbb{P}[\{(i', j') \in \mathbb{N}^2 : X((i', j')) = i, Y((i', j')) = j\}] = \mathbb{P}[\{(i, j)\}] = p_{i,j}.$$

D'après le cours (formule des probabilités totales), la loi marginale de X est donnée par :

$$\forall i \in \mathbb{N}, \mathbb{P}[X = i] = \sum_{j \in \mathbb{N}^*} \mathbb{P}[X = i, Y = j] = \alpha(1 - \alpha)^i \sum_{j \in \mathbb{N}} \beta(1 - \beta)^j = \alpha(1 - \alpha)^i.$$

Ainsi, X suit une loi géométrique de paramètre $1 - \alpha$. De même, $\mathbb{P}(Y = j) = \beta(1 - \beta)^j$, autrement dit Y suit une loi géométrique de paramètre $1 - \beta$.

-

$$\mathbb{P}[X < Y] = \mathbb{E}[\mathbb{1}_{\{X < Y\}}] = \sum_{(i,j) \in \mathbb{N}^2} \mathbb{1}_{\{i < j\}} \mathbb{P}[X = i, Y = j] = \sum_{\substack{(i,j) \in \mathbb{N}^2 \\ i < j}} p_{i,j} = \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} p_{i,j}.$$

Comme pour tout $i \in \mathbb{N}$, $\sum_{j=i+1}^{\infty} \beta(1-\beta)^j = (1-\beta)^{i+1}$, cela donne

$$\begin{aligned} \mathbb{P}[X < Y] &= \sum_{i=0}^{\infty} \alpha(1-\alpha)^i(1-\beta)^{i+1} = \sum_{i=0}^{\infty} \alpha(1-\beta)((1-\alpha)(1-\beta))^i \\ &= \frac{\alpha(1-\beta)}{1-(1-\alpha)(1-\beta)} = \frac{\alpha - \alpha\beta}{\alpha + \beta - \alpha\beta}. \end{aligned}$$

De même (en échangeant les rôles de X et Y , donc de α et β), on obtient

$$\mathbb{P}[X > Y] = \frac{\beta - \alpha\beta}{\alpha + \beta - \alpha\beta}.$$

On peut calculer $\mathbb{P}[X = Y]$ directement en sommant les $p_{i,i}$, ou alors on peut utiliser le fait que les parties $\{X = Y\}$, $\{X < Y\}$ et $\{X > Y\}$ forment une partition de \mathbb{N}^2 et les résultats précédents :

$$\mathbb{P}[X = Y] = 1 - \mathbb{P}[X > Y] - \mathbb{P}[X < Y] = \frac{\alpha + \beta - \alpha\beta - (\beta - \alpha\beta) - (\alpha - \alpha\beta)}{\alpha + \beta - \alpha\beta} = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}.$$

2. On note N la variable aléatoire comptant le nombre d'œufs qu'un insecte donné pond. On suppose que N suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}^+$. On suppose également que chaque œuf donne naissance à un nouvel insecte avec probabilité p , indépendamment de l'éclosion des autres œufs. On considère alors une famille $(X_i)_{i \geq 1}$ de variables aléatoires de loi de Bernoulli de paramètre $p \in [0, 1]$. On suppose que les variables aléatoires (N, X_1, X_2, \dots) sont indépendantes et on note D le nombre de descendants de l'insecte.

- Ecrire D en fonction des variables aléatoires N et X_i
- Pour tout $(n, d) \in \mathbb{N}^2$, calculer $\mathbb{P}(D = d | N = n)$.
- En déduire la loi de D et la loi de la variable aléatoire de \mathbb{N}^2 $Z = (D, N)$.
- Retrouver la loi de D en calculant la fonction génératrice de cette variable aléatoire.

Solution de l'exercice 2.

- On a $D = \sum_{i=1}^N X_i$ (l'indice de la somme est lui même aléatoire)
- On a :

$$\mathbb{P}[D = d | N = n] = \mathbb{P}\left[\sum_{i=1}^N X_i = d | N = n\right] = \mathbb{P}\left[\sum_{i=1}^n X_i = d | N = n\right] = \mathbb{P}\left[\sum_{i=1}^n X_i = d\right],$$

par indépendance de (N, X_1, X_2, \dots) .

La variable aléatoire $\sum_{i=1}^n X_i$ somme de n Bernoullis indépendantes de paramètre p , suit une loi binomiale de paramètres n et p , d'où

$$\mathbb{P}[D = d | N = n] = \binom{n}{d} p^d (1-p)^{n-d},$$

pour $d \leq n$, et on a bien sûr $\mathbb{P}(D = d | N = n) = 0$ si $d > n$ (le nombre de descendants ne peut être supérieur au nombre d'œufs).

- c) La variable aléatoire $Z = (D, N)$ est à valeurs de \mathbb{N}^2 . Sa loi est donnée par les valeurs des probabilités $\mathbb{P}[\{D = d\} \cap \{N = n\}]$, $d \leq n$. On a :

$$\begin{aligned}\mathbb{P}[\{D = d\} \cap \{N = n\}] &= \mathbb{P}[D = d|N = n] \mathbb{P}[N = n] \\ &= e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{d} p^d (1-p)^{n-d} \mathbb{1}_{\{d \leq n\}}.\end{aligned}$$

Pour obtenir la loi de D , on écrit avec la formule des probabilités totales :

$$\begin{aligned}\mathbb{P}[D = d] &= \sum_{n \geq 0} \mathbb{P}[\{D = d\} \cap \{N = n\}] = \sum_{n \geq d} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{d} p^d (1-p)^{n-d} \\ &= \frac{(\lambda p)^d}{d!} e^{-\lambda} \sum_{n \geq d} \frac{\lambda^{n-d} (1-p)^{n-d}}{(n-d)!} = e^{-p\lambda} \frac{(p\lambda)^d}{d!}\end{aligned}$$

On reconnaît la loi de Poisson de paramètre $p\lambda$.

- d) On peut retrouver ce résultat en utilisant une fonction génératrice. On se rappelle que la fonction génératrice d'une loi de Bernoulli de paramètre p est $1 - p + ps$ et celle d'une loi de Poisson de paramètre μ est $e^{\mu(s-1)}$. On écrit :

$$\begin{aligned}G_D(s) &= \mathbb{E}[s^D] = \mathbb{E}\left[s^{\sum_{i=1}^N X_i}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} s^{\sum_{i=1}^N X_i} \mathbb{1}_{(N=n)}\right] \\ &\stackrel{(*)}{=} \sum_{n=0}^{\infty} \mathbb{E}\left[s^{\sum_{i=1}^n X_i} \mathbb{1}_{(N=n)}\right] = \sum_{n=0}^{\infty} \mathbb{E}\left[s^{\sum_{i=1}^n X_i}\right] \mathbb{E}\left[\mathbb{1}_{(N=n)}\right] \\ &= \sum_{n=0}^{\infty} \prod_{i=1}^n \mathbb{E}[s^{X_i}] \mathbb{P}(N = n) = \sum_{n=0}^{\infty} (1 - p + ps)^n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{p\lambda(s-1)}\end{aligned}$$

On reconnaît la fonction génératrice d'une loi de Poisson de paramètre $p\lambda$, et comme la fonction génératrice caractérise la loi, D suit $\mathcal{P}(p\lambda)$. Les arguments utilisés dans le calcul ci-dessus sont l'indépendance des variables aléatoires (qui induit la multiplication des espérances), le fait que l'espérance de l'indicatrice d'un événement soit la probabilité de celui-ci, et une interversion série/espérance en (*) justifiée par le théorème de convergence monotone (les variables aléatoires qui interviennent dans la série sont positives).

3. Soit (X, Y) un vecteur aléatoire sur \mathbb{R}^2 dont la loi admet la densité

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

Déterminer les lois de X , Y , $X + Y$, $X^2 + Y^2$.

Solution de l'exercice 3. La loi de X est absolument continue par rapport à la mesure de Lebesgue et admet la densité

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y) dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

La loi de X est donc la loi normale centrée réduite. La loi de Y est égale à celle de X .

Pour calculer la loi de $X + Y$, considérons une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée. Calculons $\mathbb{E}[g(X + Y)]$:

$$\mathbb{E}[g(X + Y)] = \frac{1}{2\pi} \int_{\mathbb{R}^2} g(x + y) e^{-\frac{x^2+y^2}{2}} dx dy.$$

Faisons le changement de variable $(u, v) = (x + y, x - y)$, c'est-à-dire $(x, y) = \frac{1}{2}(u + v, u - v)$. Le jacobien $\frac{D(x,y)}{D(u,v)}$ est donné par

$$\det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \frac{1}{4} \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = -\frac{1}{2}.$$

On a $x^2 + y^2 = \frac{1}{2}(u^2 + v^2)$, donc

$$\mathbb{E}[g(X + Y)] = \frac{1}{2\pi} \int_{\mathbb{R}^2} g(u) e^{-\frac{u^2+v^2}{4}} \left| \frac{D(x,y)}{D(u,v)} \right| du dv = \frac{1}{4\pi} \int_{\mathbb{R}^2} g(u) e^{-\frac{u^2+v^2}{4}} du dv.$$

On peut intégrer par rapport à v en utilisant la relation $\int_{\mathbb{R}} e^{-\frac{v^2}{4}} dv = 2\sqrt{\pi}$, et on trouve

$$\mathbb{E}[g(X + Y)] = \frac{1}{2\sqrt{\pi}} \int_{\mathbb{R}} g(u) e^{-\frac{u^2}{4}} du,$$

donc $X + Y$ suit la loi normale centrée de variance 2 : on écrit $X + Y \sim \mathcal{N}(0, 2)$. On observe qu'on a également $X - Y \sim \mathcal{N}(0, 2)$.

Déterminons enfin la loi de $X^2 + Y^2$. On procède de la même manière que pour déterminer la loi de $X + Y$. On fait cependant un autre changement de variables : on passe en coordonnée polaires, écrivant $(x, y) = (r \cos \theta, r \sin \theta)$. L'application qui à (r, θ) associe (x, y) est un C^1 -difféomorphisme de $\mathbb{R}_+^* \times [0, 2\pi[$ dans $\mathbb{R}^2 \setminus \{(0, 0)\}$, dont le jacobien vaut

$$\left| \frac{D(x,y)}{D(r,\theta)} \right| = \left| \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \right| = \left| \det \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} \right| = r.$$

Ainsi,

$$\begin{aligned} \mathbb{E}[g(X^2 + Y^2)] &= \frac{1}{2\pi} \int_{\mathbb{R}^2} g(x^2 + y^2) e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \frac{1}{2\pi} \int_{\mathbb{R}_+^*} dr \int_0^{2\pi} d\theta g(r^2) r e^{-\frac{r^2}{2}} \\ &= \int_0^{+\infty} g(r^2) e^{-\frac{r^2}{2}} r dr \\ &= \frac{1}{2} \int_0^{+\infty} g(t) e^{-\frac{t}{2}} dt, \end{aligned}$$

où nous sommes passés de l'avant-dernière ligne à la dernière en faisant le changement de variable $t = r^2$. On reconnaît la densité de la loi exponentielle de paramètre $\frac{1}{2}$. Ainsi, $X^2 + Y^2 \sim \mathcal{E}(\frac{1}{2})$.

4. Soit (X, Y) un vecteur aléatoire de densité $f_{(X,Y)}(x, y) = \mathbb{1}_{[0,1]^2}(x, y)$. Déterminer les lois de X , Y et $Z = XY$.

Solution de l'exercice 4. Puisque le vecteur aléatoire (X, Y) admet une densité, chacune des variables X et Y admettent une densité, qu'on peut calculer respectivement par les formules $f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy$ et $f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx$. On trouve

$$f_X(x) = \int_{-\infty}^{\infty} \mathbb{1}_{[0,1]^2}(x, y) dy = \int_{-\infty}^{\infty} \mathbb{1}_{[0,1]}(x) \mathbb{1}_{[0,1]}(y) dy = \mathbb{1}_{[0,1]}(x) \int_0^1 dy = \mathbb{1}_{[0,1]}(x).$$

Ainsi, X suit la loi uniforme sur l'intervalle $[0, 1]$. Le même calcul en échangeant x et y montre que Y suit également la loi uniforme sur l'intervalle $[0, 1]$.

Pour calculer la loi de Z , considérons une fonction mesurable bornée $g : \mathbb{R} \rightarrow \mathbb{R}$ et calculons, grâce au théorème de transfert, l'espérance de $g(Z)$. On a

$$\begin{aligned} \mathbb{E}[g(Z)] &= \mathbb{E}[g(XY)] \\ &= \int_{\mathbb{R}^2} g(xy) f_{(X,Y)}(x, y) dx dy \\ &= \int_{[0,1]^2} g(xy) dx dy. \end{aligned}$$

Pour calculer cette intégrale, on effectue un changement de variables. On pose $t = xy$ et on cherche une autre variable u qui dépende de x et de y et telle que l'application $(x, y) \mapsto (t(x, y), u(x, y))$ soit bijective, de classe C^1 et admette une réciproque de classe C^1 . Posons $u(x, y) = x$. On peut retrouver x et y à partir de t et u puisque $x = u$ et $y = \frac{t}{u}$. Notons que nous devons nous restreindre à $(x, y) \in]0, 1] \times]0, 1]$ pour assurer que u ne soit pas nul, mais cela nous suffit puisque c'est, à un ensemble négligeable près, le domaine sur lequel nous intégrons.

Déterminons le domaine dans lequel (t, u) varie lorsque (x, y) décrit $]0, 1]^2$ (considérer ce domaine encore un peu plus petit ne change rien à l'intégrale et simplifie le calcul du domaine de (t, u)). On a $u = x \in]0, 1[$ et $t = xy \in]0, x[=]0, u[$. Ainsi, $(t, u) \in D = \{(a, b) \in]0, 1]^2 : a < b\}$. De plus, pour tout (t, u) dans le domaine D , le couple $(x(t, u), y(t, u)) = (u, \frac{t}{u})$ appartient à $]0, 1]^2$.

Calculons, pour tout $(t, u) \in D$, le jacobien de la transformation $(t, u) \mapsto (x(t, u), y(t, u))$. Il vaut

$$\frac{D(x, y)}{D(t, u)} = \det \begin{pmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial u} \end{pmatrix} = \det \begin{pmatrix} 0 & 1 \\ \frac{1}{u} & -\frac{t}{u^2} \end{pmatrix} = -\frac{1}{u}.$$

Nous pouvons maintenant effectuer le changement de variables dans l'intégrale, ce qui nous donne

$$\begin{aligned}
 \mathbb{E}[g(Z)] &= \int_{[0,1]^2} g(xy) \, dx dy \\
 &= \int_D g(t) \left| \frac{D(x,y)}{D(t,u)} \right| dt du \\
 &= \int_D g(t) \frac{1}{u} dt du \\
 &= \int_0^1 \int_0^1 g(t) \frac{1}{u} \mathbb{1}_{t < u} dt du \\
 &= \int_0^1 g(t) \left(\int_0^1 \frac{1}{u} \mathbb{1}_{t < u} du \right) dt \\
 &= \int_0^1 g(t) \left(\int_t^1 \frac{1}{u} du \right) dt \\
 &= \int_0^1 g(t) (-\log t) dt \\
 &= \int_{-\infty}^{+\infty} g(t) \underbrace{(-\log t) \mathbb{1}_{[0,1]}(t)}_{\text{Densité de la loi de } Z} dt.
 \end{aligned}$$

La variable aléatoire $Z = XY$ a donc une loi absolument continue par rapport à la mesure de Lebesgue, de densité $f_Z(t) = (-\log t) \mathbb{1}_{[0,1]}(t)$.

5.

- a) Soient $\theta > 0$ un nombre réel et $k, l \geq 1$ deux nombres entiers. Déterminer l'unique réel c tel que la fonction

$$f_{(X,Y)}(x, y) = cx^{k-1}y^{l-1}e^{-\theta(x+y)} \mathbb{1}_{\mathbb{R}_+ \times \mathbb{R}_+}(x, y)$$

soit la densité d'une mesure de probabilités sur \mathbb{R}^2 . Déterminer la loi de X . Cette loi s'appelle la loi Gamma de paramètres θ et k : on écrit $X \sim \Gamma(\theta, k)$. Quelle est cette loi lorsque $k = 1$?

- b) Soit (X, Y) un vecteur aléatoire dont la loi admet la densité décrite à la question précédente. Déterminer la loi de $X + Y$.

Solution de l'exercice 5.

- a) S'il existe, ce réel c est l'unique réel tel que la fonction $f_{(X,Y)}$ soit positive et telle que son intégrale sur \mathbb{R}^2 par rapport à la mesure de Lebesgue vaille 1. N'importe quel c positif rend la fonction $f_{(X,Y)}$ positive. Pour que son intégrale vaille 1, il faut que

$$1 = \int_{(\mathbb{R}^+)^2} c x^{k-1} y^{l-1} e^{-\theta(x+y)} dx dy = c \int_0^{+\infty} x^{k-1} e^{-\theta x} dx \int_0^{+\infty} y^{l-1} e^{-\theta y} dy.$$

On vérifie aisément, par récurrence sur $n \geq 0$ et en utilisant une intégration par parties pour passer d'un rang au suivant, que

$$\forall n \geq 0, \int_0^{+\infty} x^n e^{-x} dx = n!.$$

Par un changement de variables linéaires, on en déduit que les intégrales que nous devons calculer valent respectivement $(k-1)!\theta^{-k}$ et $(l-1)!\theta^{-l}$. Ainsi, l'unique valeur de c qui convient est $c = \frac{\theta^{k+l}}{(k-1)!(l-1)!}$.

La loi de X est absolument continue par rapport à la mesure de Lebesgue, de densité

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) dy = \frac{\theta^k}{(k-1)!} x^{k-1} e^{-\theta x} \mathbb{1}_{\mathbb{R}^+}(x).$$

Pour $k = 1$, on reconnaît la loi exponentielle de paramètre θ .

- b) On fait, comme à l'exercice précédent, le changement de variables $(x, y) = \frac{1}{2}(u + v, u - v)$, de sorte que $x + y = u$ et $\left| \frac{D(x,y)}{D(u,v)} \right| = \frac{1}{2}$. Ainsi, pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée, on a

$$\begin{aligned} \mathbb{E}[g(X + Y)] &= \frac{\theta^{k+l}}{(k-1)!(l-1)!} \int_{\mathbb{R}^2} g(x+y) x^{k-1} y^{l-1} e^{-\theta(x+y)} \mathbb{1}_{\{x \geq 0\}} \mathbb{1}_{\{y \geq 0\}} dx dy \\ &= \frac{\theta^{k+l}}{(k-1)!(l-1)! 2^{k+l-1}} \int_{\mathbb{R}^2} g(u) (u+v)^{k-1} (u-v)^{l-1} e^{-\theta u} \mathbb{1}_{\{u+v \geq 0\}} \mathbb{1}_{\{u-v \geq 0\}} du dv \\ &= \frac{\theta^{k+l}}{(k-1)!(l-1)! 2^{k+l-1}} \int_0^{+\infty} g(u) e^{-\theta u} \left(\int_{-u}^u (u+v)^{k-1} (u-v)^{l-1} dv \right) du, \end{aligned}$$

en utilisant le fait que les conditions $u+v \geq 0$ et $u-v \geq 0$ équivalent aux conditions $u \geq 0$ et $|v| \leq u$. Une suite de $l-1$ intégrations par parties permet d'établir l'égalité

$$\begin{aligned} \int_{-u}^u (u+v)^{k-1} (u-v)^{l-1} dv &= \frac{(l-1)!(k-1)!}{(k+l-2)!} \int_{-u}^u (v+u)^{k+l-2} dv \\ &= \frac{(l-1)!(k-1)!}{(k+l-2)!} \int_0^{2u} v^{k+l-2} dv \\ &= \frac{(l-1)!(k-1)!}{(k+l-1)!} 2^{k+l-1} u^{k+l-1}. \end{aligned}$$

Finalement,

$$\begin{aligned}\mathbb{E}[g(X + Y)] &= \frac{\theta^{k+l}}{(k-1)!(l-1)!2^{k+l-1}} \int_0^{+\infty} g(u)e^{-\theta u} \frac{(l-1)!(k-1)!}{(k+l-1)!} 2^{k+l-1} u^{k+l-1} du \\ &= \frac{\theta^{k+l}}{(k+l-1)!} \int_0^{+\infty} g(u)u^{k+l-1} e^{-\theta u} du.\end{aligned}$$

Ainsi, $X + Y$ suit une loi Gamma de paramètres θ et $k + l$.

6. Soient p et q deux réels compris entre 0 et 1. Soit (X, Y) un vecteur aléatoire tel que X suive la loi de Bernoulli de paramètre p et Y la loi de Bernoulli de paramètre q . Montrer qu'on a

$$\max(p + q - 1, 0) - pq \leq \text{Cov}(X, Y) \leq \min(p, q) - pq$$

et que les deux bornes de cette égalité peuvent être atteintes.

Solution de l'exercice 6. On a $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - pq$. Il nous faut donc encadrer $\mathbb{E}[XY]$.

D'une part, $XY \leq X$ donc $\mathbb{E}[XY] \leq \mathbb{E}[X] = p$. De même, $\mathbb{E}[XY] \leq q$. Ainsi, $\mathbb{E}[XY] \leq \min(p, q)$.

D'autre part, $\mathbb{E}[XY] = \mathbb{P}(X = 1 \text{ et } Y = 1)$. Or pour tous événements A et B , l'égalité $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ et l'inégalité $\mathbb{P}(A \cup B) \leq 1$ entraînent $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$. Comme de plus $\mathbb{P}(A \cap B) \geq 0$, on a $\mathbb{P}(A \cap B) \geq \max(\mathbb{P}(A) + \mathbb{P}(B) - 1, 0)$. Avec $A = \{X = 1\}$ et $B = \{Y = 1\}$, on trouve $\mathbb{E}[XY] = \mathbb{P}(A \cap B) \geq \max(p + q - 1, 0)$. On a ainsi établi l'inégalité voulue.

Montrons que les deux bornes peuvent être atteintes pour tous p et q . Supposons $p \leq q$. Si le couple (X, Y) a la loi

$$\mathbb{P}(X = 1, Y = 1) = p, \mathbb{P}(X = 0, Y = 1) = q - p, \mathbb{P}(X = 0, Y = 0) = 1 - q,$$

alors X et Y suivent respectivement des lois de Bernoulli de paramètres p et q et $\text{Cov}(X, Y) = \min(p, q) - pq$.

Pour la borne inférieure, distinguons deux cas suivant le signe de $p + q - 1$. Supposons tout d'abord $p + q < 1$. Alors si le couple (X, Y) a la loi

$$\mathbb{P}(X = 1, Y = 0) = p, \mathbb{P}(X = 0, Y = 1) = q, \mathbb{P}(X = 0, Y = 0) = 1 - p - q,$$

les lois de X et Y sont les bonnes et $\text{Cov}(X, Y) = -pq$.

Enfin, si $p + q \geq 1$ et si le couple (X, Y) a la loi

$$\mathbb{P}(X = 1, Y = 0) = 1 - q, \mathbb{P}(X = 0, Y = 1) = 1 - p, \mathbb{P}(X = 1, Y = 1) = p + q - 1,$$

alors les lois de X et Y sont les bonnes et $\text{Cov}(X, Y) = p + q - 1 - pq$.

7. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de dimension n dont chaque composante est de carré intégrable. On appelle *matrice de variance-covariance* (ou simplement *matrice de covariance*) de X la matrice

$$D = (\text{Cov}(X_i, X_j))_{i,j=1\dots n}.$$

Montrer que cette matrice est symétrique et positive, au sens où pour tout vecteur ligne $A = (a_1, \dots, a_n)$ de n réels, on a l'inégalité

$$AD^tA = \sum_{i,j=1}^n a_i D_{ij} a_j \geq 0.$$

Solution de l'exercice 7. Soient a_1, \dots, a_n des réels. Calculons la somme dont nous voulons montrer qu'elle est positive, en écrivant la définition de la matrice D puis de la covariance, puis la linéarité de l'espérance

$$\begin{aligned} \sum_{i,j=1}^n a_i D_{ij} a_j &= \sum_{i,j=1}^n a_i \text{Cov}(X_i, X_j) a_j \\ &= \sum_{i,j=1}^n a_i \mathbb{E}[(X_i - E[X_i])(X_j - E[X_j])] a_j \\ &= \sum_{i,j=1}^n \mathbb{E}[a_i (X_i - E[X_i]) a_j (X_j - E[X_j])] \\ &= \mathbb{E} \left[\sum_{i,j=1}^n a_i (X_i - E[X_i]) a_j (X_j - E[X_j]) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n a_i (X_i - E[X_i]) \sum_{j=1}^n a_j (X_j - E[X_j]) \right] \\ &= \mathbb{E} \left[\left(\sum_{i=1}^n a_i (X_i - E[X_i]) \right)^2 \right]. \end{aligned}$$

La quantité qui nous intéresse est donc l'espérance d'une variable aléatoire positive et intégrable : c'est donc un nombre réel positif.