3

PROBABILITÉS

Exercice 3-1

Une urne contient N boules numérotées de 1 à N. On tire, au hasard et sans remise, n boules de l'urne. On note X_i la variable aléatoire associée au numéro de la boule obtenue au i-ème tirage, ceci pour $i = 1, 2, \dots, n$.

1. a) Déterminer la loi de la variable aléatoire X_i. Préciser son espérance et sa variance.

b) Déterminer la loi du n-uplet $(X_1, X_2, \dots, X_n)$.

A l'issue du n-ième tirage les numéros obtenus sont classés par ordre croissant. On note Y_j la variable aléatoire associée au j-ième numéro, classé par ordre croissant, obtenu parmi les n numéros.
Ainsi Y_1 est la variable aléatoire $\min(X_1, X_2, \dots, X_n)$, Y_2 est le nombre aléatoire venant juste après Y_1, par ordre croissant, etc. et Y_n est la variable aléatoire $\max(X_1, X_2, \dots, X_n)$.

2. a) Déterminer la loi du n-uplet $(Y_1, Y_2, \dots, Y_n)$.

b) Déterminer la loi de Y_j , pour $j = 1, 2, \dots, n$.

c) Déterminer la loi du couple (Y_j, Y_k), pour $1 \leq j < k \leq n$.

On conserve les notations précédentes, en se restreignant au cas où $n = 3$.

3. a) Préciser la loi du couple (Y_1, Y_3). En déduire la loi de $D = Y_3 - Y_1$. Calculer l'espérance de D.

b) Déterminer la loi de Y_2. Que peut-on remarquer ?

c) Déterminer deux estimateurs sans biais de N, l'un défini comme fonction de D, l'autre comme fonction de Y_2.

Solution :

1. a) La variable aléatoire X_i prend ses valeurs dans l'ensemble $\{1, 2, \dots, N\}$, ou $X_i(\Omega) = \{1, 2, \dots, N\}$.

L'événement $(X_i = k)$ correspond à $\cap_{j=1}^{i-1}(X_j \neq k) \cap (X_i = k)$. Par la formule des probabilités composées, il vient :

$$P(X_i = k) = \frac{N-1}{N} \times \frac{N-2}{N-1} \times \cdots \times \frac{N-i+1}{N-i+2} \times \frac{1}{N-i+1} = \frac{1}{N}$$

Ainsi, X_i suit une loi uniforme sur $\{1, 2, \dots, N\}$. On a alors :

$$E(X_i) = \frac{N+1}{2}, \qquad V(X_i) = \frac{N^2-1}{12}$$

b) Pour tout $(i_1, i_2, \dots, i_n)$ distinct de $(\{1, 2, \dots, N\})^n$, on a :

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \frac{1}{A_N^n}$$

En effet, il y a A_N^n n-uplets possibles, et 1 seul favorable à la réalisation de cet événement (le décompte se fait sans répétition, puisque les (i_k) sont distincts).

2. a) Pour tout $(j_1, j_2, \dots, j_n)$ de $(\{1, 2, \dots, N\})^n$, avec $1 \leq j_1 < j_2 < \cdots < j_n \leq N$, on a :

$$P(Y_1 = j_1, Y_2 = j_2, \dots, Y_n = j_n) = \frac{1}{C_N^n}$$

En effet, il y a C_N^n n-uplets possibles, et 1 seul favorable à la réalisation de cet événement (le décompte se fait sans ordre et sans répétition, puisque les (j_k) sont distincts).

b) Pour tout $j \in \{1, 2, \dots, n\}$, on a $Y_j(\Omega) = \{j, \dots, N-n+j\}$. Pour tout $k \in Y_j(\Omega)$:

$$P(Y_j = k) = \frac{C_{k-1}^{j-1} C_1^1 C_{N-k}^{n-j}}{C_N^n}$$

En effet, l'événement $(Y_j = k)$ correspond à l'événement « les variables aléatoires $Y_1, \dots, Y_{j-1}$ occupent un $(j-1)$-uplet sans ordre et sans répétition avant k, $Y_j = k$, les variables aléatoires $Y_{j+1}, \dots, Y_n$ occupent un $(n-j)$-uplet sans ordre et sans répétition après k ».

c) Pour les mêmes raisons que ci-dessus, pour $1 \leq j < k \leq n$,

$$(Y_j, Y_k)(\Omega) = \{(m,p) \in (\{1, \dots, N\})^2 \mid j \leq m \leq p - k + j \leq N - n + j\}$$

et

$$P(Y_j = m, Y_k = p) = \frac{C_{m-1}^{j-1} C_1^1 C_{p-m-1}^{k-j-1} C_1^1 C_{N-p}^{n-k}}{C_N^n}$$

3. a) En reprenant la question précédente :

$$\begin{cases} (Y_1, Y_3)(\Omega) = \{(m,p) \in (\{1, \ldots, N\})^2 \mid 1 \leq m \leq p-2 \leq N-2\} \\ P(Y_1 = m, Y_3 = p) = \dfrac{C_{p-m-1}^1}{C_N^3} \end{cases}$$

La variable aléatoire D est à valeurs dans $\{2, \ldots, N-1\}$, et :

$$\begin{aligned} P(Y_3 - Y_1 = k) &= \sum_{j=1}^{N-k} P[(Y_3 = j+k) \cap (Y_1 = j)] = \sum_{j=1}^{N-k} \frac{C_{k-1}^1}{C_N^3} \\ &= \frac{(N-k)(k-1)}{C_N^3} \end{aligned}$$

Le calcul de l'espérance de D se fait alors aisément :

$$E(D) = \sum_{k=2}^{N-1} \frac{k(N-k)(k-1)}{C_N^3} = \frac{1}{C_N^3} \sum_{k=2}^{N-1} ((N+1)k^2 - Nk - k^3) = \frac{N+1}{2}$$

b) Déterminons la loi de Y_2. Par les questions précédentes, on a :

$$Y_2(\Omega) = \{2, \ldots, N-1\}, \quad \text{et } P(Y_2 = k) = \frac{(k-1)(N-k)}{C_N^3}$$

On remarque que D et Y_2 suivent la même loi d'où $E(Y_2) = \dfrac{N+1}{2}$.

c) $D' = 2D - 1$ et $Y' = 2Y_2 - 1$ sont deux estimateurs sans biais de N. Ils ont de plus même variance ; on ne peut donc préférer l'un à l'autre.

Exercice 3-2

Soit $(X_k)_{k \in \mathbb{N}^*}$, une suite de variables aléatoires indépendantes suivant toutes la loi de Bernoulli de paramètre p (avec $p \in]0,1[$) et définies sur le même espace probabilisé $(\Omega, \mathcal{B}, P)$.

On pose, pour $k \in \mathbb{N}^*$, $Y_k = X_k X_{k+1}$.

1. Donner la loi de Y_k, ainsi que l'espérance et la variance de Y_k.

2. Soient i et j deux entiers naturels distincts.
Discuter, suivant les valeurs de i et de j, de l'indépendance de Y_i et de Y_j.

3. On pose, pour $n \in \mathbb{N}^*$, $Z_n = \dfrac{Y_1 + \cdots + Y_n}{n}$. Montrer que :

$$\forall\, \varepsilon > 0,\ \lim_{n\to\infty} P\big(|Z_n - p^2| \geq \varepsilon\big) = 0$$

Solution :

1. Pour tout $k \in \mathbb{N}^*$, Y_k suit une loi de Bernoulli de paramètre $p(Y_k = 1)$. Or

$$(Y_k = 1) = (X_k = 1) \cap (X_{k+1} = 1).$$

Par l'indépendance des (X_k), $p(Y_k = 1) = p^2$. Finalement :

$$E(Y_k) = p^2, \qquad V(Y_k) = p^2(1 - p^2)$$

2. Si $j \neq i-1$ ou $j \neq i+1$, les variables aléatoires Y_i et Y_j sont indépendantes commes fonctions de variables aléatoires indépendantes.
Par contre, $P(Y_i Y_{i+1} = 1) = P(X_i X_{i+1}^2 X_{i+2} = 1) = p^3$,
alors que $P(Y_i = 1)P(Y_{i+1} = 1) = p^2$. Les variables Y_i et Y_{i+1} ne sont pas indépendantes.

3. Calculons l'espérance et la variance de Z_n.

$$E(Z_n) = \frac{1}{n}\sum_{k=1}^{n} E(Y_k) = p^2$$

et

$$\begin{aligned}
V(Z_n) &= \frac{1}{n^2}\left(\sum_{k=1}^{n} V(Y_k) + 2\sum_{1\leq i<j\leq n} \mathrm{cov}(Y_i, Y_j)\right)\\
&\frac{1}{n^2}\left(np^2(1-p^2) + 2\sum_{1\leq i<j\leq n} \mathrm{cov}(Y_i, Y_j)\right)\\
&= \frac{1}{n^2}\left(np^2(1-p^2) + 2\sum_{i=1}^{n-1} E(Y_i Y_{i+1}) - E(Y_i)E(Y_{i+1})\right)\\
&= \frac{1}{n^2}\big(np^2(1-p^2) + 2(n-1)(p^3 - p^4)\big)
\end{aligned}$$

Il reste à appliquer l'inégalité de Bienaymé-Chebicheff à Z_n, soit :

$$p(|Z_n - E(Z_n)| \geq \varepsilon) \leq \frac{V(Z_n)}{\varepsilon^2}$$

avec

$$\lim_{n\to\infty} V(Z_n) = \lim_{n\to\infty} \frac{1}{n^2}\big(np^2(1-p^2) + 2(n-1)(p^3 - p^4)\big) = 0$$

Exercice 3-3

Soit X une variable aléatoire réelle à densité, de loi uniforme sur l'intervalle $]0,1]$, et b un nombre réel strictement positif.

1. On pose $Y = -b\ln(X)$. Déterminer une densité de Y, ainsi que son espérance et sa variance.

Soient Y_1 et Y_2 deux variables aléatoires réelles indépendantes de même loi que Y.
On pose : $U = Y_1 + Y_2, V = Y_1/Y_2$ (quotient de Y_1 par Y_2), $S = \ln(V)$.

2. a) Déterminer une densité de U.

b) Déterminer une densité de S (on déterminera au préalable une densité de $T = -\ln Y_2$).

c) En déduire une densité de V, puis une densité de la variable aléatoire W définie par $W = \dfrac{1}{1+V}$.
Préciser l'espérance et la variance de W.
d) En exprimant W en fonction de Y_1 et de Y_2, montrer que la valeur de l'espérance de W était prévisible.

e) Calculer $E(UW) - E(U)E(W)$. Quelle remarque peut-on faire ?

Solution :

1. On a $Y(\Omega) = \mathbb{R}^{+*}$, et pour tout $y > 0$:

$$P(Y < y) = P(X > e^{-\frac{y}{b}}) = 1 - e^{-\frac{y}{b}}$$

Aussi la densité de Y est définie par :

$$f_Y(y) = \begin{cases} \dfrac{1}{b}e^{-y/b} & \text{si } y > 0 \\ 0 & \text{sinon} \end{cases}$$

On remarque que Y suit une loi $\Gamma(b,1)$, donc que $E(Y) = b, V(Y) = b^2$.

2. a) Les variables Y_1 et Y_2 suivant toutes deux une loi $\Gamma(b,1)$ et étant indépendantes, le cours nous dit que $U = Y_1 + Y_2$ suit une loi $\Gamma(b,2)$.

b) Soit $T = -\ln Y_2$. Alors $T(\Omega) = \mathbb{R}$ et pour tout t réel :

$$P(T < t) = P(-b\ln X > e^{-t}) = P(X < \exp(-\frac{e^{-t}}{b})) = \exp(-\frac{e^{-t}}{b})$$

D'où la densité de T définie sur $\mathbb{R}$ par :

$$f_T(t) = \exp(-\frac{e^{-t}}{b}) \cdot \frac{1}{b}e^{-t}$$

Un calcul similaire donne qu'une densité de $\ln Y_1$ est définie sur $\mathbb{R}$ par :

$$f_{\ln Y_1}(t) = \exp(-\frac{e^t}{b}) \cdot \frac{1}{b}e^t$$

Posons $S = \ln Y_1 - \ln Y_2$. Alors $S(\Omega) = \mathbb{R}$ et pour tout $s \in \mathbb{R}$:

$$f_S(s) = \int_{\mathbb{R}} f_{\ln Y_1} f_{-\ln Y_1}(s-t)dt$$

soit

$$f_S(s) = \int_{\mathbb{R}} \exp(-\frac{e^{-t}}{b}) \cdot \frac{1}{b}e^{-t}\exp(-\frac{e^{s-t}}{b}) \cdot \frac{1}{b}e^{s-t}dt$$

Posons $u = \varphi(t) = \dfrac{1}{b}e^{-t}$ qui est une bijection de classe C^1 de $\mathbb{R}$ sur $\mathbb{R}^+$. On obtient :

$$f_S(s) = e^s \int_0^{+\infty} ue^{-u(1+e^s)}du$$

et, pour tout s réel :

$$f_S(s) = \frac{e^s}{(1+e^s)^2}\Gamma(2) = \frac{e^s}{(1+e^s)^2}$$

c) On a $V = \exp(S), V(\Omega) = \mathbb{R}^+$ et une densité de V est définie pour $v \geq 0$ par $f_V(v) = \dfrac{1}{(1+v)^2}$.

On sait que $W(\Omega) =]0,1]$ et pour tout $0 < w \leq 1$:

$$P(W < w) = P(V > \frac{1}{w} - 1) = 1 - F_V\left(\frac{1}{w} - 1\right)$$

On obtient ainsi, pour la densité de W :

$$f_W(w) = -f_V\left(\frac{1}{w} - 1\right)\left(-\frac{1}{w^2}\right) = 1$$

ce qui donne que W suit une loi uniforme sur $]0,1]$. Ainsi $E(W) = \dfrac{1}{2}, V(W) = \dfrac{1}{12}$.

d) On sait que $W = \dfrac{Y_2}{Y_1+Y_2} = 1 - \dfrac{Y_1}{Y_1+Y_2} = 1 - W'$. Comme Y_1 et Y_2 jouent des rôles symétriques, W et W' suivent la même loi, et $E(W) = E(W') = 1/2$.

e) On a $UW = Y_2$. Donc $E(UW) - E(U)E(W) = 0$. Si on peut étendre les théorèmes connus des variables aléatoires discrètes aux variables continues, alors $\rho_{U,W} = 0$. Cela signifie-t-il que U et W sont indépendantes ?

Exercice 3-4

A. Soit k un entier naturel strictement positif. On considère une suite $(x_n)_{n>0}$, de nombres réels appartenant au segment $[0,k[$.
On pose pour tout n entier strictement positif :

$$s_n = \sum_{j=1}^{n} x_j, \qquad r_n = f(s_n) = s_n - \lfloor s_n \rfloor$$

où $\lfloor x \rfloor$ représente la partie entière de x.

1. Montrer que la suite (r_n) est bornée.

2. Montrer que $r_{n+1} = f(f(s_n) + x_{n+1})$, pour tout $n > 0$.

B. Soient X_1 et X_2 deux variables aléatoires réelles continues indépendantes, définies sur le même espace probabilisé $(\Omega, \mathcal{B}, P)$ et toutes deux de loi uniforme sur l'intervalle $[0,k[$.

1. Déterminer une densité de la variable aléatoire S_2, définie par $S_2 = X_1 + X_2$.

2. On définit la variable aléatoire R_2 par : $R_2 = S_2 - \lfloor S_2 \rfloor$.
Déterminer la fonction de répartition de R_2 en fonction de celle de S_2 et en déduire une densité de R_2, dans les cas :
   a) $k = 1$,
   b) $k = 2$.

3. Soient $(X_n)_{n>0}$, une suite de variables aléatoires indépendantes définies sur $(\Omega, \mathcal{B}, P)$, toutes de loi uniforme sur l'intervalle $[0,1[$. On définit pour tout $n > 0$:

$$S_n = X_1 + X_2 + \cdots + X_n, \quad R_n = S_n - \lfloor S_n \rfloor$$

Déterminer la loi de R_n (on pourra utiliser les résultats de la question A.2 et un raisonnement par récurrence).

Solution :

A. 1. Pour tout n, s_n est un nombre réel et pour tout n, $0 \leq r_n < 1$. (r_n s'appelle la partie fractionnaire de s_n).

2. Calculons, pour tout $n \in \mathbb{N}^*$:

$$f(s_n) + x_{n+1} = s_n - \lfloor s_n \rfloor + x_{n+1} = s_{n+1} - \lfloor s_n \rfloor$$

et

$$\begin{aligned} f(f(s_n)+x_{n+1}) &= s_{n+1} - \lfloor s_n \rfloor - \lfloor (s_{n+1} - \lfloor s_n \rfloor) \rfloor \\ &= s_{n+1} - \lfloor s_n \rfloor - \lfloor s_{n+1} \rfloor + \lfloor s_n \rfloor \\ &= r_{n+1} \end{aligned}$$

B. 1. On sait que $S_2(\Omega) = [0, 2k[$ et que pour tout x réel, la densité f de S_2 est définie par :

$$f(x) = \int_{\mathbb{R}} f_{X_1}(t) f_{X_2}(x-t)dt$$

Donc

- $0 \leq x \leq k,\ f(x) = \displaystyle\int_0^x \frac{dt}{k^2} = \frac{x}{k^2}$
- $k \leq x < 2k,\ f(x) = \displaystyle\int_{x-k}^k \frac{dt}{k^2} = \frac{2k-x}{k^2}$.

Finalement :

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x}{k^2} & \text{si } 0 \leq x \leq k \\ \dfrac{2k-x}{k^2} & \text{si } k \leq x < 2k \\ 0 & \text{si } x \geq 2k \end{cases}$$

2. a) Pour $k = 1$, il vient :

$$R_2 = \begin{cases} S_2 & \text{si } 0 \leq S_2 < 1 \\ S_2 - 1 & \text{si } 1 \leq S_2 < 2 \end{cases}$$

Ainsi, pour $t \in [0, 1[$:

$P(0 \leq R_2 < t)$

$= P\left\{[(0 \leq S_2 < t) \text{ et } (0 \leq S_2 < 1)] \text{ ou } [(0 \leq S_2 - 1 < t) \text{ et } (1 \leq S_2 < 2)]\right\}$

et, par disjonction et inclusion :

$$P(0 \leq R_2 < t) = P(0 \leq S_2 < t) + P(1 \leq S_2 < 1+t)$$

La fonction de répartition de R_2 est définie par :

$$F_{R_2}(t) = \int_0^t x dx + \int_1^{1+t} (2-x) dx = t, \text{ pour } t \in [0, 1[$$

soit :

$$F_{R_2}(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t < 1 \\ 1 & \text{si } t \geq 1 \end{cases}$$

ce qui signifie que R_2 suit une loi uniforme sur $[0, 1[$.

Pour $k = 2$, on a :

$$R_2 = \begin{cases} S_2 & \text{si } 0 \leq S_2 < 1 \\ S_2 - 1 & \text{si } 1 \leq S_2 < 2 \\ S_2 - 2 & \text{si } 2 \leq S_2 < 3 \\ S_2 - 3 & \text{si } 3 \leq S_2 < 4 \end{cases}$$

et de même que précédemment, pour $t \in [0, 1[$, par disjonction et inclusion :

$$\begin{aligned} P(0 \leq R_2 < t) &= P(0 \leq S_2 < t) + P(1 \leq S_2 < 1 + t) + P(2 \leq S_2 < 2 + t) \\ &\quad + P(3 \leq S_2 < 3 + t) \\ &= \int_0^t \frac{xdx}{4} + \int_1^{1+t} \frac{xdx}{4} + \int_2^{2+t} \left(1 - \frac{x}{4}\right) dx + \int_3^{3+t} \left(1 - \frac{x}{4}\right) dx \\ &= t \end{aligned}$$

De même que précédemment, R_2 suit une loi uniforme sur $[0, 1[$.

3. Peut-on envisager de montrer par récurrence que R_n suit une loi uniforme sur $[0, 1[$?

- c'est vrai pour $n = 1, 2$.
- supposons que R_n suive une loi uniforme sur $[0, 1[$. Alors $R_{n+1} = f(S_{n+1}) = f(f(S_n) + X_{n+1}) = f(R_n + X_{n+1})$. Or R_n et X_{n+1} suivent chacune une loi uniforme sur $[0, 1]$ et sont indépendantes. Par la question précédente, R_{n+1} suit une loi uniforme sur $[0, 1[$.

Exercice 3-5

Soit $n \in \mathbb{N}, n \geq 2$.
n véhicules numérotés de 1 à n sont engagés dans un rallye automobile. Ils finissent une étape entre minuit et une heure du matin.
Pour tout $i \in \{1, \ldots, n\}$, on note U_i la variable aléatoire égale à l'heure d'arrivée du véhicule i. On suppose que les variables (U_i) sont indépendantes et suivent une loi uniforme sur $[0, 1]$.

Pour $t \in [0, 1]$, on note N_t la variable aléatoire représentant le nombre de véhicules arrivés au plus tard à l'instant t.

Enfin, pour tout $i \in \{1, \ldots, n\}$, on note T_i la variable aléatoire représentant l'heure d'arrivée du véhicule classé en i-ième position.

1. Déterminer la loi de N_t.

2. Soit $i \in \{1, \ldots, n\}$. Déterminer, sous forme d'une somme, la fonction de répartition de T_i, puis une densité et l'espérance de T_i.

Solution :

1. Chaque variable aléatoire U_i suit une loi uniforme sur $[0,1]$. Ainsi pour tout t réel, $P(U_i \leq t) = t$.
Chaque voiture a la probabilité t d'arriver avant l'instant t, les heures d'arrivée sont indépendantes et N_t représentant le nombre de voitures arrivées avant l'instant t, suit une loi binomiale $\mathcal{B}(n,t)$.

2. Pour tout $i \in \{1,\dots,n\}$, notons F_i la fonction de répartition de la variable aléatoire T_i.
Pour tout $t \in]0,1[$,

$$F_i(t) = P(T_i \leq t) = P(N_t \geq i) = \sum_{k=i}^{n} C_n^k t^k (1-t)^{n-k}$$

Ainsi, pour tout t réel :

$$F_i(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \displaystyle\sum_{k=i}^{n} C_n^k t^k (1-t)^{n-k} & \text{si } t \in]0,1[\\ 1 & \text{si } t \geq 1 \end{cases}$$

La fonction F_i vérifie les propriétés des fonctions de répartition des variables à densité (F est continue, dérivable sauf peut-être en 0 et 1, de limite 0 en $-\infty$ et 1 en $+\infty$).

Pour tout $0 < t < 1$:

$$\begin{aligned} F_i'(t) &= \sum_{k=i}^{n} C_n^k k t^{k-1} (1-t)^{n-k} - \sum_{k=i}^{n-1} C_n^k (n-k) t^k (1-t)^{n-k-1} \\ &= \sum_{k=i-1}^{n-1} C_n^{k+1} (k+1) t^k (1-t)^{n-k-1} - \sum_{k=i}^{n-1} C_n^k (n-k) t^k (1-t)^{n-k-1} \\ &= C_n^i i t^{i-1} (1-t)^{n-i} + \sum_{k=i}^{n-1} \left((k+1) C_n^{k+1} - (n-k) C_n^k \right) t^k (1-t)^{n-k-1} \\ &= C_n^i i t^{i-1} (1-t)^{n-i} \end{aligned}$$

et $E(T_i) = \displaystyle\int_0^1 i C_n^i t^i (1-t)^{n-i}$. Une intégration par parties donne que :

$$\begin{aligned} J_i &= \int_0^1 t^i (1-t)^{n-i} dt = \frac{n-i}{i+1} \int_0^1 t^{i+1} (1-t)^{n-i-1} dt \\ &= \frac{n-i}{i+1} J_{i+1} = \cdots = \frac{(n-i)!}{(i+1)(i+2)\dots(n+1)} \end{aligned}$$

et

$$E(T_i) = \frac{iC_n^i}{(i+1)C_{n+1}^i} = \frac{i}{n+1}$$

Exercice 3-6

Une urne contient des boules blanches en proportion p et des boules rouges en proportion q, toutes indiscernables au toucher, avec $q = 1-p$ et $0 < p < 1$. On procède à une suite de n tirages au hasard d'une boule de l'urne, dont on note la couleur avant de la replacer dans l'urne, n étant un entier naturel supérieur ou égal à 3, fixé a priori. On constitue ainsi des séries unicolores, au gré des couleurs obtenues dans la suite des n tirages. Par exemple, pour $n = 9$, en notant B l'obtention d'une boule blanche et R l'obtention d'une boule rouge, la suite de résultats $B, B, R, B, R, R, R, B, B$ fournit 5 séries unicolores : BB, puis R, puis B, puis RRR, et enfin BB.

Soit X la variable aléatoire qui compte le nombre de séries unicolores obtenues à l'issue des n tirages.

1. Calculer :

   a) la probabilité $P[X = m]$.

   b) la probabilité $P[X = M]$,

m et M désignant respectivement les valeurs minimales et maximales que peut prendre X.

On désigne par Y_i la variable aléatoire valant 1 lorsqu'une boule blanche est obtenue au $(i-1)$-ième tirage et une boule rouge est obtenue au i-ème tirage, valant 0 sinon.

Symétriquement, on désigne par Z_j la variable aléatoire valant 1 lorsqu'une boule rouge est obtenue au $(j-1)$-ième et une boule blanche est obtenue au j-ème tirage, valant 0 sinon.

2. Exprimer X en fonction des Y_i et des Z_j. En déduire l'espérance de X.

Tout au long de la suite des n lancers, on gagne un Euro à chaque fois que l'obtention d'une boule blanche est immédiatement suivie de l'obtention d'une boule rouge ; inversement, on perd un Euro à chaque fois que l'obtention d'une boule rouge est immédiatement suivie de l'obtention d'une boule blanche. Soit G_k la variable aléatoire représentant le gain global en Euro(s), (positif ou négatif), obtenu à l'issue du k-ième tirage.

3. a) Déterminer les lois de G_1, G_2, G_3.

b) Montrer que les variables $G_k, k > 1$, sont toutes de même loi.

c) Préciser l'espérance et la variance de G_k.

d) En utilisant la question précédente, calculer la variance de X.

Solution :

1. Le nombre de séries unicolores est compris entre 1 et n. Donc $X(\Omega) \in \{1, \ldots, n\}$.
L'événement $(X = 1)$ correspond a des séries unicolores de longueur n, soit de boules blanches, soit de boules rouges. Aussi $P(X = 1) = p^n + q^n$.
L'événement $(X = n)$ correspond à un changement de couleur à chaque tirage. Aussi
• si $n = 2k$ est pair, l'événement $(X = n)$ correspond au tirage $(BRBR \ldots BR)$ ou $(RBRB \ldots RB)$ et $P(X = n) = 2(pq)^k$.
• si $n = 2k + 1$ est impair, l'événement $(X = n)$ correspond au tirage $(BRBR \ldots BRB)$ ou $(RBRB \ldots RBR)$ et $P(X = n) = p^{k+1}q^k + p^k q^{k+1} = (pq)^k$.

2. Les variables aléatoires (Y_i) et (Z_j) représentent les ruptures de couleurs. En effet l'événement $(Y_i + Z_i = 1)$ correspond à un changement de couleur. Aussi $X = 1 + \sum_{i=2}^{n} Y_i + \sum_{j=2}^{n} Z_j$. Les variables aléatoires (Y_i) et (Z_j) suivant toutes des lois de Bernoulli de paramètre pq, on a $E(X) = 1 + 2(n - 1)pq$.

3. a) La variable aléatoire G_1 est certaine : $G_1(\Omega) = \{0\}, E(G_1) = 0, V(G_1) = 0$.
La variable aléatoire G_2 prend ses valeurs dans $\{-1, 0, 1\}$ et :

$$P(G_2 = 1) = pq, \quad P(G_2 = -1) = pq, \quad P(G_2 = 0) = 1 - 2pq$$

La variable aléatoire G_3 prend ses valeurs dans le même ensemble.
• $(G_3 = 1) = (B, R, R) \cup (B, B, R) \Rightarrow P(G_3 = 1) = pq^2 + p^2q = pq$
• $(G_3 = -1) = (R, B, B) \cup (R, R, B) \Rightarrow P(G_3 = -1) = pq^2 + p^2q = pq$
• $(G_3 = 0) = (B, R, B) \cup (R, B, R) \cup (B, B, B) \cup (R, R, R) \Rightarrow P(G_3 = 0) = 1 - 2pq$.

b) Soit $k > 1$. La variable aléatoire G_k est à valeurs dans $\{-1, 0, 1\}$. En effet l'événement $(|G_k| \geq 2)$ est impossible puisque perdre (ou gagner) un second Euro nécessite d'avoir gagné (ou perdu) le premier.
• l'événement $(G_k = 1)$ signifie avoir obtenu une boule blanche au premier tirage, une boule rouge au k-ième et des résultats quelconques entre ces deux

tirages. Aussi :

$$P(G_k = 1) = pq \sum_{i=0}^{k-2} C_{k-2}^i p^i q^{k-2-i} = pq$$

• l'événement $(G_k = -1)$ signifie avoir obtenu une boule rouge au premier tirage, une boule blanche au k-ième et des résultats quelconques entre ces deux tirages. Aussi :

$$P(G_k = -1) = pq \sum_{i=0}^{k-2} C_{k-2}^i p^i q^{k-2-i} = pq$$

• ainsi $P(G_k = 0) = 1 - 2pq$.

c) Par des calculs évidents, il vient, pour $k > 1, E(G_k) = 0$ et $V(G_k) = 2pq$.

d) On peut écrire $G_n = \sum_{i=2}^{n} Y_i - \sum_{j=2}^{n} Z_j = Y - Z$. Ainsi :

$$\begin{cases} V(G_n) = V(Y) + V(Z) - 2\mathrm{cov}(Y, Z) = 2pq \\ V(X) = V(Y) + V(Z) + 2\mathrm{cov}(Y, Z) \end{cases}$$

Or Y et Z suivent la même loi en inversant les rôles de B et R. Aussi :

$$V(X) = 4V(Y) - 2pq$$

Mais :

$$V(Y) = \sum_{i=2}^{n} V(Y_i) + 2 \sum_{2 \le i < j \le n} \mathrm{cov}(Y_i, Y_j) = (n-1)pq(1-pq) - 2(n-2)(pq)^2$$

En effet, si $j \geq i + 2$, les variables Y_i et Y_j sont indépendantes (il n'y a pas de tirage commun) et si $j = i+1, Y_i \cdot Y_j = 0$ et $\mathrm{cov}(Y_i, Y_j) = -E(Y_i)E(Y_j) = -(pq)^2$. Finalement :

$$V(X) = (4n-6)pq - (12n-20)(pq)^2$$

Exercice 3-7

Soit n un entier naturel strictement positif, inconnu a priori.

Une urne contient des jetons à deux faces portant chacun, sur une des faces un numéro bleu, et sur l'autre face un numéro rouge. On sait que, sur l'ensemble des jetons de l'urne, k exactement portent le numéro bleu k, ceci pour $k = 1, 2, \ldots, n$, et que, parmi les k jetons portant le numéro bleu k, un et un seul porte le numéro rouge i, ceci pour $i = 1, 2, \ldots, k$.

1. Déterminer, en fonction de n, le nombre de jetons contenus dans l'urne.

On tire au hasard un jeton de l'urne. On désigne par B la variable aléatoire associée à son numéro bleu, et par R la variable aléatoire associée a son numéro rouge. On pose, d'autre part, $G = B - R$.

2. a) Déterminer la loi du couple (B, R).

b) En déduire les lois de B et de R. Calculer leurs espérances et leurs variances.

c) Suite au tirage d'un jeton, on gagne G. Préciser l'espérance de G et calculer la variance de G.

3. Déduire de b) et c), trois estimateurs sans biais de n, l'un étant une fonction de B, le second une fonction de R et le troisième une fonction de G. Préciser leurs variances respectives. Lequel est le plus indiqué pour estimer n ?

Solution :

1. Le nombre total de jetons est égal à :

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

2. a) Le couple aléatoire (B, R) vérifie :

$$(B, R)(\Omega) = \{(i, j) \in \mathbb{N}^2 \mid 1 \leq j \leq i \leq n\}$$

et pour tout $(i, j) \in (B, R)(\Omega)$:

$$P((B, R) = (i, j)) = \frac{1}{\frac{n(n+1)}{2}} = \frac{2}{n(n+1)}$$

b) Il vient immédiatement que $B(\Omega) = \{1, \ldots, n\}$ et pour tout $1 \leq i \leq n$:

$$P(B = i) = \sum_{j=1}^{i} P((B, R) = (i, j)) = \frac{2i}{n(n+1)}$$

Ainsi :

$$E(B) = \frac{2}{n(n+1)} \sum_{i=1}^{n} i^2 = \frac{2n+1}{3},$$

$$E(B^2) = \frac{2}{n(n+1)} \sum_{i=1}^{n} i^3 = \frac{n(n+1)}{2},$$

$$V(B) = \frac{n^2 + n - 2}{18}$$

La loi de R est tout aussi immédiate. On a $R(\Omega) = \{1, \dots, n\}$ et pour tout $1 \leq j \leq n$:

$$P(R = j) = \sum_{i=j}^{n} P((B, R) = (i, j)) = \frac{2(n - j + 1)}{n(n + 1)}$$

Un calcul élémentaire donne $E(R) = \dfrac{n+2}{3}, V(R) = V(B)$.

c) La variable aléatoire G est égale à $B - R$. Aussi $E(G) = E(B) - E(R) = \dfrac{n-1}{3}$, et :

$$V(G) = V(B) + V(R) - 2\mathrm{cov}(B, R) = \frac{n^2 + n - 2}{18}$$

En effet :

$$\begin{aligned} E(BR) &= \sum_{i=1}^{n} \sum_{j=1}^{i} ij P((B, R) = (i, j)) \\ &= \frac{2}{n(n+1)} \sum_{i=1}^{i} \sum_{j=1}^{n} ij \\ &= \frac{n(n+1)}{4} + \frac{2n+1}{6} \end{aligned}$$

et :

$$\mathrm{cov}(B, R) = \frac{n(n+1)}{4} + \frac{2n+1}{6} - \frac{(2n+1)(n+2)}{9}$$

3. Posons $B' = \dfrac{3B - 1}{2}, R' = 3R - 2, G' = 3G + 1$. Ces trois variables aléatoires ont pour espérance n et sont des estimateurs sans biais de n. Et :

$$V(B') = \frac{n^2 + n - 2}{8}, \quad V(R') = V(G') = \frac{n^2 + n - 2}{2}$$

indique que B' est un meilleur estimateur que R' et G'.

Exercice 3-8

Une épreuve écrite de concours se présente sous forme d'un Q.C.M.
50 questions, supposées mutuellement indépendantes , y sont posées. Pour chacune des 50 questions, il y a quatre sous-questions supposées mutuellement indépendantes.
Pour chaque sous-question, le candidat a le choix entre deux réponses : « vrai » ou « faux ». Les réponses aux questions sont présentées en ligne, une ligne par

question, chaque ligne étant constituée de quatre cases à cocher, une par sous-question.
On suppose que le candidat donne une réponse à chaque sous-question et qu'il répond à chaque fois au hasard.
On note F la variable aléatoire décomptant le nombre de sous-questions dont la réponse est erronée.

1) Dans une première éventualité, on suppose que le Q.C.M. est corrigé « question par question », c'est-à-dire qu'une ligne est considérée comme juste et rapporte 4 points au candidat si et seulement si toutes les réponses de la ligne sont correctes. Dans le cas contraire, le candidat est pénalisé d'un point.
On note X la variable aléatoire égale au nombre de points obtenus par le candidat.

Déterminer la loi de X, son espérance, et sa variance.

2) Désormais, on suppose que les sous-questions sont corrigées indépendamment les unes des autres, chaque case ayant une réponse correcte rapportant 1 point au candidat, et chaque case ayant une réponse incorrecte pénalisant la note de $1/4$ de point.
On note Y la variable aléatoire égale au nombre de points obtenus dans cette deuxième correction.

Déterminer la loi de Y, son espérance, et sa variance.

Solution :

1. Soit L le nombre de lignes non globalement justes. L prend ses valeurs dans $\{0, 1, \ldots, 50\}$ et le nombre de points obtenus par le candidat est

$$4(50 - L) - L = 200 - 5L.$$

Pour tout $j \in \{0, \ldots, 50\}, P(X = 200 - 5j) = P(L = j)$. Or L suit une loi binomiale de paramètres $(50, p)$, avec $p = 1 - \dfrac{1}{2^4} = \dfrac{15}{16}$. En effet, une ligne est considérée comme correcte si les 4 réponses sont correctes, et ceci avec la probabilité $\dfrac{1}{16}$. Les réponses étant considérées indépendamment ligne par ligne, on obtient la justification demandée.

Comme $X = 200 - 5L$, il vient

$$E(X) = 200 - 250p \text{ et } V(X) = 25V(L) = 1250p(1 - p).$$

2. Soit F le nombre de réponses fausses. F prend ses valeurs dans $\{0, 1, \ldots, 200\}$ et le nombre de points obtenus par le candidat est $(200 - F) - F/4 = 200 - \frac{5}{4}F$.

Pour tout $j \in \{0, \ldots, 200\}, P(Y = 200 - 5j/4) = P(F = j)$. Or F suit une loi binomiale de paramètres $(200, 1/2)$.
Donc $Y(\Omega) = \{200 - 5j/4, 0 \leq j \leq 200\}$, et pour tout $j \in \{0, \ldots, 200\}$:

$$P(Y = 200 - 5j/4) = P(F = j) = C_{200}^{j} \left(\frac{1}{2}\right)^n$$

Comme $Y = 200 - 5F/4$, il vient $E(Y) = 75$ et $V(Y) = \dfrac{25}{64}$.

Exercice 3-9

Soient X_1, X_2, X_3, X_4, quatre variables aléatoires indépendantes, définies sur le même espace probabilisé $(\Omega, \mathcal{B}, P)$, et toutes de loi uniforme sur l'intervalle $[0, b]$, b étant un nombre réel strictement positif inconnu.

On définit les variables réelles à densité suivantes :

$$Y = \min(X_1, X_2), Z = \max(X_1, X_2), A = \min(X_3, X_4), B = \max(X_3, X_4)$$

1. Déterminer la loi de Y (fonction de répartition et densité).

2. Déterminer la loi de Z (fonction de répartition et densité).

On définit, d'autre part, les variables aléatoires :

$$S = \max(A, Y), \text{ et } T = \min(B, Z)$$

et on note F et G, respectivement, les fonctions de répartition de S et T.

3. Déterminer F et G, et montrer que :

$$(*) \quad F(x) - G(x) \geq 0, \forall\, x \in \mathbb{R}$$

4. En intégrant F et G par parties sur l'intervalle $[0, b]$, et en utilisant la relation $(*)$, montrer que $E(S) < E(T)$, où E désigne l'opérateur espérance.

5. Déterminer des densités de S et T, et calculer les espérances et les variances de S et T.

6. Soient a et b deux réels, avec $a < b$, et U et V, deux variables aléatoires réelles à densité, définies toutes deux sur l'intervalle $[a, b]$, dont les fonctions de répartitions respectives, F_U et F_V sont supposées de classe $\mathcal{C}^1$ et vérifiant la relation : $\forall\, x \in \mathbb{R}, F_U(x) - F_V(x) \geq 0$.
A-t-on $E(U^k) \leq E(V^k)$ pour tout k entier naturel ?

Solution :

1. Les variables aléatoires X_1 et X_2 suivant une loi uniforme sur $[0, b]$ et étant indépendantes, il vient $Y(\Omega) = [0, b]$, et pour tout $y \in \mathbb{R}$:

$$P(Y > y) = P((X_1 > y) \cap (X_2 > y)) = (1 - F_X(y))^2$$

et :

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 - \left(1 - \dfrac{y}{b}\right)^2 & \text{si } 0 \leq y \leq b \\ 1 & \text{si } y > b \end{cases}$$

d'où une densité de Y :

$$f_Y(y) = \begin{cases} 0 & \text{si } y \notin [0, b] \\ \dfrac{2}{b}\left(1 - \dfrac{y}{b}\right) & \text{autrement} \end{cases}$$

2. Les variables aléatoires X_1 et X_2 suivant une loi uniforme sur $[0, b]$ et étant indépendantes, il vient $Z(\Omega) = [0, b]$, et pour tout $z \in \mathbb{R}$:

$$P(Z < z) = P((X_1 < z) \cap (X_2 < z)) = (F_X(z))^2$$

et :

$$F_Z(z) = \begin{cases} 0 & \text{si } z < 0 \\ \left(\dfrac{z}{b}\right)^2 & \text{si } 0 \leq z \leq b \\ 1 & \text{si } z > b \end{cases}$$

d'où une densité de Z :

$$f_Z(z) = \begin{cases} 0 & \text{si } z \notin [0, b] \\ \dfrac{2z}{b^2} & \text{autrement} \end{cases}$$

3. Par indépendance de A et Y, il vient, pour tout s réel :

$F(s) = p((A < s) \cap (Y < s)) = [F_Y(s)]^2$. Donc :

$$F(s) = \begin{cases} 0 & \text{si } s < 0 \\ \dfrac{s^2}{b^4}(2b - s)^2 & \text{si } 0 \leq s \leq b \\ 1 & \text{si } s > b \end{cases}$$

De même, par indépendance de B et Z, il vient, pour tout t réel :

$$1 - G(t) = p((B > t) \cap (Z > t)) = [P(Z > t)]^2.$$

Donc :

$$G(t) = \begin{cases} 0 & \text{si } t < 0 \\ \dfrac{t^2}{b^4}(2b^2 - t^2) & \text{si } 0 \leq t \leq b \\ 1 & \text{si } t > b \end{cases}$$

Enfin, pour tout x réel : $G(x) - F(x) = -\dfrac{2x^2}{b^4}(b - x)^2 \leq 0$.

4. Comme F et G sont des fonctions de classe C^1, les espérances de S et T existent et :

$$\int_0^b F(x)dx = [xF(x)]_0^b - \int_0^b xf_S(x)dx = b - E(S)$$

$$\int_0^b G(x)dx = [xG(x)]_0^b - \int_0^b xf_T(x)dx = b - E(T)$$

La relation, $F(x) - G(x) \geq 0$, pour tout x réel entraîne donc que $E(S) \leq E(T)$. De plus, on a l'égalité $E(S) = E(T)$ si et seulement si $\int_0^b (G(x) - F(x))dx = 0$ ce qui entraîne $F = G$ (puisque $F - G$ est positive et continue), ce qui n'est pas le cas ici.

5. Un calcul immédiat donne :

$$f_S(s) = \begin{cases} 0 & \text{si } s \notin [0,b] \\ \dfrac{4}{b^4}s(s-b)(s-2b) & \text{sinon} \end{cases}, f_T(t) = \begin{cases} 0 & \text{si } t \notin [0,b] \\ \dfrac{4t}{b^4}(b^2-t^2) & \text{sinon} \end{cases}$$

et

$$E(S) = \frac{7b}{15},\ V(S) = \frac{11b^2}{225},\ E(T) = \frac{8b}{15},\ V(T) = \frac{11b^2}{225}$$

6. Le cas $k = 0$ est évident et le cas $k = 1$ a été traité dans une question précédente. Pour $k \geq 2$, considérons :

$$\begin{aligned} \int_a^b kx^{k-1}(F_U - F_V)(x)dx &= \left[x^k(F_U - F_V)(x)\right]_a^b - \int_a^b x^k(f_U - f_V)(x)dx \\ &= E(V^k) - E(U^k) \end{aligned}$$

Ainsi :

$$E(U^k) \leq E(V^k) \Leftrightarrow \int_a^b kx^{k-1}(F_U - F_V)(x)dx \geq 0$$

- pour k impair cette inégalité est toujours vérifiée.
- pour k pair, elle est vérifiée si $a \geq 0$, mais n'est pas toujours vérifiée si $a < 0$. En effet, dans ce cas, il est possible de trouver deux fonctions de répartition F_U et F_V telles que l'intégrale reste négative.

Exercice 3-10

On considère une fonction f de classe $\mathcal{C}^2$ de $[0,1]$ dans $\mathbb{R}$ telle que $f(0) = 0$, $f'(0) = \lambda \in]0,1[$ et pour tout $x \in]0,1]$, $f'(x) > 0$ et $f''(x) < 0$.

1. a) Montrer que pour tout $x \in]0,1]$, on a $f(x) \leq \lambda x$.

b) On prend $x_0 \in]0,1]$ et pour $n \in \mathbb{N}$ on définit par récurrence $x_{n+1} = f(x_n)$

Montrer que la suite (x_n) est décroissante et converge vers 0.

2. On considère des variables aléatoires indépendantes de Bernoulli $(X_n)_{n\in\mathbb{N}^*}$ de paramètre x_n, c'est-à-dire que la probabilité de succès la n-ième fois (ce qui correspond à $X_n = 1$) est x_n.

a) Quelle est la probabilité p_n d'obtenir n échecs aux n premières épreuves ?

b) Montrer que la suite (p_n) est décroissante et converge. On note ℓ sa limite. Comment interpréter la propriété $\ell = 0$?

c) Montrer que $\ell = 0$ si et seulement si la série de terme général x_n diverge.

3. On pose $f(x) = \dfrac{x}{1+x}$.

a) La fonction f vérifie-t-elle les hypothèses du 1 ?

b) Calculer $\dfrac{1}{x_{n+1}} - \dfrac{1}{x_n}$ · En déduire un équivalent de x_n quand n tend vers l'infini.

Que vaut, dans ce cas, ℓ définie au 2) ?

Solution :

1. a) La fonction f étant de classe C^2 et vérifiant $f(0) = 0, f'(0) = \lambda, f''(t) < 0$, la formule de Taylor avec reste intégral permet d'écrire, pour tout $x \in [0,1]$:

$$f(x) = \lambda x + \int_0^x (x-t)f''(t)dt < \lambda x$$

b) Par la question précédente, comme $\lambda \in]0,1]$, la suite (x_n) est décroissante, minorée par 0. Elle converge donc vers une limite $L \geq 0$ qui vérifie $L \leq \lambda L$ donc $L = 0$ lorsque $\lambda < 1$.

2. a) Par indépendance, on a :

$$p_n = (1-x_1)(1-x_2)\dots(1-x_n) = \prod_{k=1}^{n}(1-x_k)$$

b) Comme $\dfrac{p_{n+1}}{p_n} = 1 - x_n < 1$, la suite (p_n) est décroissante, et minorée par 0. Elle admet donc une limite ℓ.
$\ell = 0$ signifie obtenir une infinité d'échecs, car les événements E_n=«obtenir n échecs lors des n premières épreuves», forment une suite décroissante d'événements.

c) On a $\ln p_n = \displaystyle\sum_{k=1}^{n} \ln(1-x_k)$. La suite (x_n) étant convergente vers 0, on sait que pour n grand $\ln(1-x_n) \sim -x_n$. Ainsi la série $\sum x_n$ diverge si et seulement si la série $\sum \ln(1-x_n)$ diverge c'est-à-dire si et seulement si la limite $\lim\limits_{n\to+\infty} \ln(p_n)$ n'existe pas ou encore si et seulement si $\ell = 0$.

3. On vérifie aisément que la fonction $f : x \mapsto \dfrac{x}{1+x}$ vérifie les conditions de la première question,sauf pour $\lambda = 1$. Mais, pour tout $n \in \mathbb{N}$:

$$\frac{1}{x_{n+1}} - \frac{1}{x_n} = \frac{x_n}{1+x_n} - \frac{1}{x_n} = 1$$

La suite $\left(\dfrac{1}{x_{n+1}} - \dfrac{1}{x_n}\right)_n$ est convergente vers 1 (ici même égale...). On peut écrire :

$$\sum_{k=0}^{n-1}\left(\frac{1}{x_{k+1}} - \frac{1}{x_k}\right) = \left(\frac{1}{x_n} - \frac{1}{x_0}\right) = n$$

donc $\dfrac{1}{x_n} \sim n$ et $x_n \sim \dfrac{1}{n}$. Ainsi la série $\sum x_n$ diverge et $\ell = 0$.

Exercice 3-11

1. Soient X et Y deux variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{B}, P)$, indépendantes, de même loi, à valeurs dans $\mathbb{N}$ et telles que :

$$\forall n \in \mathbb{N}, P(X = n) 0.$$

Les variables $X + Y$ et $X - Y$ sont-elles indépendantes ?

2. Réciproquement, on suppose que X et Y, à valeurs dans $\mathbb{N}$, sont telles que $X + Y$ et $X - Y$ sont indépendantes.
Les variables aléatoires X et Y sont-elles indépendantes ?

3. Soient X et Y à valeurs dans $\mathbb{N}$, possédant une variance, telles que $X + Y$ et $X - Y$ soient indépendantes. Comparer la variance de X et celle de Y.

Solution :

1. La réponse est négative. En effet, on a $P(X + Y = 1) \neq 0, P(X - Y = 0) \neq 0$, mais $P((X + Y = 1) \cap (X - Y = 0)) = 0$.

2. Ici encore la réponse est négative. En effet, si X est une variable aléatoire à valeurs dans $\mathbb{N}$ et si $Y = X$, alors $X + Y = 2X, X - Y = 0$, et $2X$ est indépendante de 0, alors que X ne l'est pas de X.

3. Comme $X + Y$ et $X - Y$ ont une variance, leur produit a une espérance et par indépendance $E((X + Y)(X - Y)) = E(X + Y)E(X - Y)$, soit en développant :

$$E(X^2) - E(Y^2) = E^2(X) - E^2(Y) \Rightarrow V(X) = V(Y)$$

Exercice 3-12

On admet le résultat suivant :
Soit $(u_{n,k})_{n\in\mathbb{N},k\in\mathbb{N}}$ une suite à deux indices de réels positifs ou nuls telle que :

a) pour tout $k \in \mathbb{N}$, la série $\sum\limits_{n=0}^{\infty} u_{n,k}$ converge. On note S_k sa somme ;

b) la série $\sum\limits_{k=0}^{\infty} S_k$ converge.

Alors on peut permuter les symboles $\sum$, c'est-à-dire que l'on a :

$$\sum_{k=0}^{\infty} S_k = \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} u_{n,k} \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} u_{n,k} \right)$$

1. Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et (X, N) un couple de variables aléatoires discrètes définies sur Ω à valeurs dans $\mathbb{N}$. On suppose que X admet une espérance. Montrer que si on note $E(X \mid N = n)$ l'espérance de la loi conditionnelle de X sachant $(N = n)$, on a :

$$E(X) = \sum_{n=0}^{\infty} E(X \mid N = n).P(N = n)$$

2. Une expérience consiste à effectuer des lancers successifs et indépendants d'une pièce donnant « pile » avec la probabilité $p \in]0, 1[$ et « face » avec la probabilité $q = 1 - p$.

On note N la variable aléatoire égale au nombre de tirages nécessaires pour obtenir pour la première fois « pile ».

Dans un deuxième temps, si « pile » est apparu pour la première fois au n-ième tirage, on effectue n lancers de la même pièce et on note X le nombre de « pile » obtenus lors de cette seconde série de lancers.
Déterminer l'espérance de X (on admettra que cette espérance existe).

3. Dans un casino, un jeu se déroule de la façon suivante : un croupier mélange trois cartes, qui sont l'As de Coeur, le 2 de Coeur et le Valet de Pique et les présente sur la table face cachée. Un joueur choisit une carte au hasard. Si celle-ci est un Coeur, il gagne la somme correspondante (1 Euro si c'est l'As ou 2 Euros si c'est le 2) et rejoue, le croupier mélangeant à nouveau les trois cartes ; si la carte tirée est le Valet de Pique, le jeu s'arrête.
On note N le nombre de cartes de Coeur tirées jusqu'à l'apparition du Valet de Pique et S la somme des Euros obtenus.
a) Déterminer la loi de N. Quelle est la probabilité que le Valet de Pique n'apparaisse jamais ?

b) Déterminer la loi conditionnelle de S sachant que $N = n$.

c) Quel prix minimum le casino doit-il faire payer une telle partie pour ne pas être perdant en moyenne ?

Solution :

1. La série $\displaystyle\sum_{k=0}^{\infty} kP(X = k|N = n)$ est une série convergente, car pour tout $k \geq 1$, on a :

$$0 \leq kP(X = k|N = n) \leq kP(X = k)$$

et on sait que X admet une espérance.

On a donc, par définition $E(X|N=n)=\sum_{k=0}^{\infty} kP(X=k|N=n)$, et, par la propriété admise :

$$\begin{aligned}
E(X) &= \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k\left(\sum_{n=0}^{\infty} P(X=k|N=n)P(N=n)\right)\\
&= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{\infty} kP(X=k|N=n)\right)P(N=n)\\
&= \sum_{n=0}^{\infty} E(X|N=n)P(N=n)
\end{aligned}$$

2. N représente le temps d'attente du premier « pile ». On sait que N suit une loi géométrique de paramètre p, soit pour tout $n\in\mathbb{N}^*$: $P(N=n)=pq^{n-1}$.
Or la loi conditionnelle de X conditionnée par $(N=n)$ représente le nombre de
« piles » obtenus lors de n lancers ; c'est donc une loi binomiale de paramètres (n,p). Donc pour tout $0\leq k\leq n$:

$$P(X=k|N=n)=C_n^k p^k q^{n-k},\qquad E(X=k|N=n)=np$$

En utilisant la question précédente, il vient :

$$\begin{aligned}
E(X) &= \sum_{n=1}^{\infty} E(X|N=n)P(N=n)\\
&= \sum_{n=1}^{\infty} np\cdot pq^{n-1} = p^2\left(\sum_{n=1}^{\infty} nq^{n-1}\right)\\
&= p^2\frac{1}{(1-q)^2}=1
\end{aligned}$$

3. a) N représente le nombre de parties jouées avant l'apparition du Valet de Pique. N suit une loi géométrique sur $\mathbb{N}$ de paramètre $\frac{1}{3}$, soit :

$$P(N=n)=\frac{1}{3}\left(\frac{2}{3}\right)^n, \forall\ n\in\mathbb{N}$$

On a alors $\sum_{n=0}^{\infty} P(N=n)=1$ et la probabilité que le Valet de Pique n'apparaisse jamais est nulle.

b) L'événement $(N=n)$ correspond au tirage de n cartes Cœur, qui rapporte chacune 1 ou 2 Euros. Donc $(S|N=n)(\Omega)=\{n,\dots,2n\}$.

c) $(S = k|N = n)$ correspond au tirage de x cartes As de Cœur et y cartes deux de Cœur, avec $\begin{cases} x + y = n \\ x + 2y = k \end{cases}$. Donc $\begin{cases} x = 2n - k \\ y = k - n \end{cases}$.
Ainsi, pour tout $k \in \{n, \dots, 2n\}$:

$$P(X = k|N = n) = \frac{C_n^{k-n}}{2^n}$$

et

$$\begin{aligned} E(X|N = n) &= \sum_{k=n}^{2n} k\frac{C_n^{k-n}}{2^n} = \frac{1}{2^n}\sum_{j=0}^{n}(n + j)C_n^j \\ &= \frac{1}{2^n}n\sum_{j=0}^{n} C_n^j + \frac{1}{2^n}\sum_{j=0}^{n} jC_n^j \\ &= n + \frac{n}{2} = \frac{3n}{2} \end{aligned}$$

et

$$E(S) = \sum_{n=0}^{\infty} E(S|N = n)P(N = n) = \sum_{n=0}^{\infty} \frac{3n}{2}\cdot\frac{1}{3}\cdot\left(\frac{2}{3}\right)^n = \frac{1}{3}\sum_{n=1}^{\infty} n\left(\frac{2}{3}\right)^{n-1} = 3$$

Le prix minimum que le casino doit faire payer pour ne pas être perdant en moyenne doit être supérieur ou égal à 3 Euros.

Exercice 3-13

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et T une variable aléatoire définie sur Ω suivant une loi normale centrée réduite $\mathcal{N}(0, 1)$.

1. Rappeler l'expression d'une densité φ de T. On notera Φ la fonction de répartition de T.

2. a) Montrer que pour tout $x > 0$, on a : $0 < 1 - \Phi(x) \leq \dfrac{1}{2x^2}$.
(on pourra utiliser l'inégalité de Bienaymé-Chebicheff).
b) En déduire l'existence de l'intégrale $\displaystyle\int_0^{+\infty} \big(1 - \Phi(x)\big)\, dx$, puis la calculer.

3. On associe à T les variables aléatoires suivantes :

$$X = |T|, \quad Y = \lfloor T \rfloor, \quad Z = \lfloor X \rfloor$$

où $|\ |$ désigne la valeur absolue et $\lfloor\ \rfloor$ désigne la partie entière.
a) Déterminer une densité de X.

b) Exprimer, pour tout $k \in \mathbb{N}$, la probabilité $p(Z = k)$ à l'aide de la fonction Φ.

c) Déterminer, pour tout $k \in \mathbb{Z}$, la probabilité $P(Y = k)$.

d) Montrer la convergence de la série $\sum_{k \geq 1} k\big(P(Y = k) - P(Y = -k)\big)$ (on pourra raisonner à partir des sommes partielles). En déduire l'espérance de Y.

e) Montrer que Z admet une espérance.

Solution :

1. C'est effectivement une question de cours ! Une densité de T est définie pour tout x réel par $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$ et la fonction de répartition de T est alors $\Phi(x) = \displaystyle\int_{-\infty}^{x} \varphi(t)dt$.

2. Par définition d'une fonction de répartition, on sait que pour tout x réel, $0 \leq \Phi(x) \leq 1$, et, ici, comme $e^{-t^2/2} > 0$ pour tout t réel, on a : $\Phi(x) < 1$.

a) L'inégalité de Bienaymé-Chebicheff affirme que, pour tout x réel :

$$P(|T| > x) \leq \frac{1}{x^2} \Leftrightarrow P((T > x) \cup (T < -x)) \leq \frac{1}{x^2}$$

donc que $2P(T > x) \leq \dfrac{1}{x^2}$ ce qui correspond à $1 - \Phi(x) \leq \dfrac{1}{2x^2}$.

b) La majoration que l'on vient d'obtenir montre que l'intégrale $\displaystyle\int_0^{+\infty} (1 - \Phi(x))dx$ converge. De plus, pour tout $A > 0$:

$$\int_0^A (1 - \Phi(t))dt = [t(1 - \Phi(t))]_0^A - \int_0^A t\varphi(t)dt$$

Or $0 < A(1 - \Phi(A)) \leq A\dfrac{1}{2A^2}$ entraîne que $\lim\limits_{A \to +\infty} A(1 - \Phi(A)) = 0$, donc que :

$$\int_0^{+\infty} (1 - \Phi(t))dt = \int_0^{+\infty} t\varphi(t)dt = \frac{1}{\sqrt{2\pi}}$$

3. a) Calculons la fonction de répartition de $X = |T|$.

- pour tout $x \leq 0, P(X \leq x) = 0$.
- pour $x > 0$, $P(X \leq x) = P(-x \leq T \leq x) = \displaystyle\int_{-x}^{x} \varphi(t)dt = 2\Phi(x) - 1$.

Une densité de X est alors :

$$\varphi_1(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 2\varphi(x) & \text{si } x > 0 \end{cases}$$

et son espérance est $E(X) = 2\int_0^{+\infty} \varphi(t)dt = \sqrt{\dfrac{2}{\pi}}$.

b) Pour tout $k \geq 0$, on a $(Z = k) = (\lfloor X \rfloor = k) = (k \leq X < k+1)$. Ainsi :

$$P(Z = k) = 2\int_k^{k+1} \varphi(t)dt = 2[\Phi(k+1) - \Phi(k)]$$

c) de la même façon, pour $k \in \mathbb{Z}$: $(Y = k) = (\lfloor T \rfloor = k) = (k \leq T < k+1)$, et

$$P(Y = k) = \int_k^{k+1} \varphi(t)dt = \Phi(k+1) - \Phi(k)$$

d) Utilisons, comme suggéré dans l'énoncé, les sommes partielles. Pour tout $N \in \mathbb{N}$:

$$\begin{aligned}\sum_{k=1}^{N} k\big(p(Y = k) - p(Y = -k)\big) &= \sum_{k=1}^{N} k\big(\Phi(k+1) - \Phi(k)\big) \\ &\quad - \sum_{k=1}^{N} k\big(\Phi(-k+1) - \Phi(-k)\big)\end{aligned}$$

Or :

$$\begin{aligned}\sum_{k=1}^{N} k\big(\Phi(k+1) - \Phi(k)\big) &= \sum_{k=1}^{N}[(k+1)\Phi(k+1) - k\Phi(k)] - \sum_{k=1}^{N} \Phi(k+1) \\ &= (N+1)\Phi(N+1) - \Phi(1) - \sum_{k=2}^{N+1} \Phi(k)\end{aligned}$$

et comme pour tout $x \geq 0, \Phi(-x) = 1 - \Phi(x)$:

$$\begin{aligned}\sum_{k=1}^{N} k\big(\Phi(-k+1) - \Phi(-k)\big) &= \sum_{k=1}^{N} k\big(\Phi(k) - \Phi(k-1)\big) \\ &= \sum_{k=1}^{N}[k\Phi(k) - (k-1)\Phi(k-1)] - \sum_{k=1}^{N} \Phi(k-1) \\ &= N\Phi(N) - \sum_{k=0}^{N-1} \Phi(k)\end{aligned}$$

Donc :

$$\sum_{k=1}^{N} k\big(p(Y=k)-p(Y=-k)\big) = N(\Phi(N+1)-\Phi(N)) + \Phi(N+1)$$
$$-\Phi(1)+\Phi(0)+\Phi(1)-\Phi(N)-\Phi(N+1)$$

Or :

$$N(\Phi(N+1)-\Phi(N)) = N\int_N^{N+1} \frac{e^{-t^2/2}}{\sqrt{2\pi}}dt \leq \frac{N}{\sqrt{2\pi}}e^{-N^2/2}$$

tend vers 0 lorsque N tend vers l'infini. Ainsi :

$$\lim_{N\to+\infty} \sum_{k=1}^{N} k\big(p(Y=k)-p(Y=-k)\big) = 1-\Phi(0) = \frac{1}{2}$$

e) La série $\sum k(\Phi(k+1)-\Phi(k))$ est convergente, car

$$0 < k(\Phi(k+1)-\Phi(k)) = k\int_k^{k+1} \frac{e^{-t^2/2}}{\sqrt{2\pi}}dt \leq \frac{ke^{-k^2/2}}{\sqrt{2\pi}}$$

qui est le terme général d'une série convergente. Ceci signifie que l'espérance de Z existe.

Exercice 3-14

On considère n variables aléatoires $X_1, X_2, \ldots, X_n$ indépendantes suivant la même loi de Poisson de paramètre inconnu λ.

1. On pose $S_n = \sum_{k=1}^{n} X_k$ et $\overline{X_n} = \frac{1}{n}S_n$.

Rappeler la loi de S_n.

Montrer que $\overline{X_n}$ est un estimateur de λ sans biais et convergent.

2. On pose $\theta = P(X=0) = \mathrm{e}^{-\lambda}$ (où $P(A)$ désigne la probabilité de l'événement A) et on cherche un estimateur sans biais du paramètre θ.
(Dans le cas où X représente le nombre de pannes que subit un appareil, $P(X=0)$ est la probabilité qu'il n'y ait aucune panne et c'est en ce sens qu'il est intéressant de l'estimer).

a) Calculer l'espérance de $T_n = \mathrm{e}^{-\overline{X_n}} = \exp(-\overline{X_n})$ et montrer que T_n n'est pas un estimateur sans biais de θ.

b) Soit s un entier quelconque. Montrer que la loi conditionnelle de X_1 sachant $[S_n = s]$ est une loi binomiale dont on déterminera les paramètres.
Préciser en particulier la valeur de $P(X_1 = 0 \mid S_n = s)$

c) On considère la variable aléatoire :

$$\widehat{\theta} = \left(1 - \frac{1}{n}\right)^{S_n}$$

Montrer que $\widehat{\theta}$ est un estimateur sans biais de θ.

d) Calculer la variance de $\widehat{\theta}$ et montrer que $\lim\limits_{n\to+\infty} V(\widehat{\theta}\,) = 0$.

Solution :

1. Les n variables aléatoires $X_1, X_2, \ldots, X_n$ étant indépendantes et suivant la même loi de Poisson de paramètre λ, on sait que S_n suit une loi de Poisson de paramètre $n\lambda$. Son espérance et sa variance sont donc égales à $n\lambda$. De plus $E(\overline{S_n}) = \lambda$ et $V(\overline{S_n}) = \dfrac{\lambda}{n}$.

Comme $\lim\limits_{n\to+\infty} V(\overline{S_n}) = 0$, $\overline{S_n}$ est un estimateur sans biais et convergent de λ.

2. a) On sait, d'après le théorème de transfert, que :

$$\begin{aligned}
E(T_n) &= \sum_{k=0}^{+\infty} e^{-\frac{k}{n}} P(S_n = k) = \sum_{k=0}^{+\infty} e^{-\frac{k}{n}} e^{-n\lambda} \frac{(n\lambda)^k}{k!} \\
&= e^{-n\lambda} \sum_{k=0}^{+\infty} \frac{\left(e^{-1/n}(n\lambda)\right)^k}{k!} \\
&= e^{(n\lambda)(e^{-1/n}-1)}
\end{aligned}$$

On a $E(T_n) \neq \theta$ ce qui signifie que T_n n'est pas un estimateur sans biais de θ. On a tout de même $\lim\limits_{n\to+\infty} E(T_n) = \theta$ ce qui signifie que T_n est asymptotiquement sans biais.

b) On sait que :

$$P(X_1 = k | S_n = s) = \frac{P((X_1 = k) \cap (S_n = s))}{P(S_n = s)}$$

Or, $P((X_1 = k) \cap (S_n = s)) = P(X_1 = k \cap (\sum_{j\neq 1} X_j = s - k))$. Les variables aléatoires X_1 et $\displaystyle\sum_{j\neq 1} X_j$ sont indépendantes et $\displaystyle\sum_{j\neq 1} X_j$ suit une loi de Poisson de paramètre $(n-1)\lambda$. On obtient alors :

$$P(X_1 = k | S_n = s) = e^{-\lambda} \frac{\lambda^k}{k!} e^{-(n-1)\lambda} \times \frac{((n-1)\lambda)^{s-k}}{(s-k)!} \times \frac{s!}{e^{-n\lambda}(n\lambda)^s}$$

et

$$P(X_1 = k|S_n = s) = C_s^k \frac{(n-1)^{s-k}}{n^s} = C_s^k \left(\frac{1}{n}\right)^k \left(\frac{n-1}{n}\right)^{s-k}$$

Ainsi la loi conditionnelle de X_1 sachant $(S_n = s)$ est la loi binomiale de paramètres $s, \frac{1}{n}$. Enfin :

$$P(X_1 = 0|S_n = s) = \left(1 - \frac{1}{n}\right)^s$$

c) D'après le théorème de transfert :

$$E(\widehat{\theta}) = \sum_{s=0}^{+\infty} P(X_1 = 0|S_n = s)P(S_n = s) = P(X_1 = 0) = e^{-\lambda}$$

et $\widehat{\theta}$ est un estimateur sans biais de θ.

d) Comme dans les questions précédentes :

$$\begin{aligned} E(\widehat{\theta}^2) &= \sum_{s=0}^{+\infty} \left(1 - \frac{1}{n}\right)^{2s} P(S_n = s) \\ &= \sum_{s=0}^{+\infty} \left(1 - \frac{1}{n}\right)^{2s} e^{-n\lambda} \frac{(n\lambda)^s}{s!} \\ &= e^{-n\lambda} \sum_{s=0}^{+\infty} \frac{\left(\left(1 - \frac{1}{n}\right)^2 (n\lambda)\right)^s}{s!} \\ &= e^{-2\lambda} e^{\frac{\lambda}{n}} \\ V(\widehat{\theta}) &= e^{-2\lambda}\left(e^{\frac{\lambda}{n}} - 1\right) \end{aligned}$$

quantité qui tend vers 0 lorsque n tend vers l'infini.

Exercice 3-15

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Soit r un réel tel que $0 < r < 1$. Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre r et, pour $n \geq 2$, S_n une variable aléatoire suivant une loi binomiale de paramètres (n, r).

1. a) Pour tout réel $t > 0$, on pose $g(t) = E(\exp(-tX))$. (E désigne l'espérance, exp la fonction exponentielle). Calculer $g(t)$ en fonction de t et r.

b) Justifier l'égalité :

$$E\left(\exp\left(-t\frac{S_n}{n}\right)\right) = \left(g\left(\frac{t}{n}\right)\right)^n$$

2. Soit a un réel tel que $0 < a < r < 1$ et t un réel positif fixé.
a) Montrer que :

$$p\left(\frac{S_n}{n} \leq a\right) = p\left[\exp\left(-t\left(\frac{S_n}{n} - a\right)\right) \geq 1\right]$$

b) Soit Y une variable aléatoire prenant un nombre fini de valeurs positives. Montrer que $p(Y \geq 1) \leq E(Y)$. En déduire que :

$$p\left(\frac{S_n}{n} \leq a\right) \leq \exp(at)\left(g\left(\frac{t}{n}\right)\right)^n$$

3. a) Etudier les variations de la fonction $t \mapsto \ln\left[\exp(at)g(t)\right]$

b) En déduire qu'il existe une constante $M > 0$ telle que :

$$p\left(\frac{S_n}{n} \leq a\right) \leq \exp(-nM)$$

4. Soit $\varepsilon > 0$ donné. Donner une interprétation probabiliste du résultat précédent lorsque $a = r - \varepsilon$.

Solution :

1. a) On sait que $e^{-tX}(\Omega) = \{1, e^{-t}\}$. Ainsi :

$$g(t) = (1-r) + re^{-t}$$

b) Soient $(X_1, X_2, \ldots, X_n)$ n variables indépendantes suivant une loi de Bernoulli de paramètre r, alors $\dfrac{S_n}{n} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ et :

$$E\left(\exp\left(-\frac{tS_n}{n}\right)\right) = E\left(\exp\left(-\frac{t}{n}\sum_{i=1}^{n} X_i\right)\right) = E\left(\prod_{i=1}^{n}\exp(-\frac{tX_i}{n})\right)$$

$$= \prod_{i=1}^{n} E[\exp(-\frac{tX_i}{n})] = \left(g\left(\frac{t}{n}\right)\right)^n$$

2. a) Comme $t > 0$ et comme la fonction exponentielle est bijective, on peut écrire : $\left(\dfrac{S_n}{n} \leq a\right) = \left(\dfrac{S_n}{n} - a \leq 0\right) = \left(-t\left(\dfrac{S_n}{n} - a\right) \geq 0\right)$

$$= \left[\exp\left(-t\left(\frac{S_n}{n} - a\right)\right) \geq 1\right]$$

b) On suppose que $Y(\Omega) = \{y_1, y_2, \dots, y_n\}$, avec $y_i \geq 0$, pour tout $i \in \{1, \dots, n\}$. On a alors :

$$\begin{aligned} E(Y) = \sum_{i=1}^{n} y_i P(Y = y_i) &= \sum_{i|y_i \leq 1} y_i P(Y = y_i) + \sum_{i|y_i \geq 1} y_i P(Y = y_i) \\ &\geq \sum_{i|y_i \geq 1} y_i P(Y = y_i) \geq \sum_{i|y_i \geq 1} P(Y = y_i) = P(Y \geq 1) \end{aligned}$$

(On remarquera que cette inégalité reste vérifiée même si $P(Y \geq 1) = 0$, puisque l'espérance de Y est positive). Par les deux résultats précédents, pour tout $t > 0$:

$$\begin{aligned} P\left(\frac{S_n}{n} \leq a\right) &= P\left[\exp\left(-t\left(\frac{S_n}{n} - a\right)\right) \geq 1\right] \\ &\leq E\left[\exp\left(-t\left(\frac{S_n}{n} - a\right)\right)\right] \\ &= e^{at} E\left[\exp\left(-t\left(\frac{S_n}{n}\right)\right)\right] \\ &= e^{at}\left(g\left(\frac{t}{n}\right)\right)^n \end{aligned}$$

3. a) On peut écrire $\varphi : t \mapsto at + \ln(re^{-t} + 1 - r)$. Cette fonction est continue et dérivable sur $\mathbb{R}^+$ et, pour tout $t \geq 0$:

$$\varphi'(t) = \frac{r}{e^t(r-1) - r} + a$$

On a $\varphi'(t) = 0$ si et seulement si $t = t_0 = \ln\left(\frac{(a-1)r}{a(r-1)}\right) > 0$ (car $0 < a < r < 1$). Ainsi la fonction φ admet un minimum en t_0 qui est strictement négatif, puisque $\varphi(0) = 0$.

b) Notons $-M$ ce minimum. On remarque que :

$$e^{at}\left(g\left(\frac{t}{n}\right)\right)^n = \left(e^{at/n} g\left(\frac{t}{n}\right)\right)^n = \psi^n(\frac{t}{n})$$

avec $\psi(t) = e^{\varphi(t)}$. Donc :

$$P\left(\frac{S_n}{n} \leq a\right) \leq \left(e^{\varphi(t/n)}\right)^n = e^{n\varphi(t/n)} \leq e^{-nM}$$

4. Si $a \leq r - \varepsilon$, il vient :

$$P\left(\frac{S_n}{n} \leq r - \varepsilon\right) \leq e^{-nM}$$

qui tend vers 0 lorsque n tend vers l'infini. Ainsi :

$$\lim_{n\to+\infty} P\left(\frac{S_n}{n} \geq r - \varepsilon\right) = 1$$

Exercice 3-16

On lance une pièce de monnaie jusqu'à ce que l'on obtienne pour la première fois une série d'au moins deux résultats identiques suivis d'un résultat contraire. On arrête alors les lancers.
On suppose qu'on a une probabilité p d'obtenir «pile» et $q = 1 - p$ d'obtenir «face» lorsqu'on lance cette pièce et que les résultats des divers lancers sont indépendants.

On note :

- X la variable aléatoire égale au nombre de lancers effectués,
- Y la variable aléatoire égale au rang du lancer où commence la première série de résultats identiques ,
- Z la variable aléatoire égale à 0 si la première série de résultats identiques est une série de faces, égale à 1 si la première série de résultats identiques est une série de piles.

Par exemple :
pour $PFPFPPPPF$, on a $X = 9$, $Y = 5$ et $Z = 1$.
pour $FFFFP$, on a $X = 5$, $Y = 1$ et $Z = 0$.

1. Déterminer la loi du triplet (X, Y, Z), puis du couple (X, Y) (on pourra séparer les cas Y pair et Y impair).

2. On se place dans le cas où $p = 1/2$. Déterminer la loi et l'espérance de X.

3. On considère la fonction Pascal suivante :

```
Function f(p : real) : char ;
Var ok : boolean ;
Begin
ok :=random<=p ;
If ok Then f :='p' Else f :='f'
End ;
```

La fonction `random`, sans paramètre, est une fonction qui renvoie une valeur de type `real` prise au hasard dans l'intervalle $[0, 1]$. Chaque appel à cette fonction donne une nouvelle valeur indépendante des précédentes.

Que rend la fonction `f` lorsqu'on l'exécute ?

4. On désire simuler l'expérience décrite dans cet exercice, lorsque $p = 0.8$.

A l'aide de la fonction `f`, écrire un programme affichant à l'écran une série de lancers, et la valeur correspondante des variables aléatoires X et Y sous la forme : $PFPFPFPPPPPF$ 12 7.
(on utilisera une variable `r : array[1..1000] of char` dans laquelle on rangera les résultats des lancers successifs obtenus.)

Solution :

1. Il est évident que $Z(\Omega) = \{0, 1\}$ et que :
$$(X, Y))(\Omega) = \{(\ell, m) \in \mathbb{N}^{*2} \mid \ell \geq m + 2\}.$$
Calculons la probabilité $P((X, Y, Z) = (\ell, m, n))$.
• l'événement $(X, Y, Z) = (\ell, 2k, 0)$ correspond à l'obtention d'une suite de PF k fois, puis d'une suite de F $(\ell - 2k - 1)$ fois qui commence au coup $2k$ et un P au rang ℓ. Ainsi $P((X, Y, Z) = (\ell, 2k, 0)) = p^{k+1} q^{\ell-k-1}$.
• l'événement $(X, Y, Z) = (\ell, 2k, 1)$ correspond à l'obtention d'une suite de FP k fois, puis d'une suite de P $(\ell - 2k - 1)$ fois et un F au rang ℓ. Ainsi $P((X, Y, Z) = (\ell, 2k, 0)) = p^{\ell-k-1} q^{k+1}$.
• l'événement $(X, Y, Z) = (\ell, 2k + 1, 1)$ correspond à l'obtention d'une suite de PF k fois, puis d'une suite de P $(\ell - 2k - 1)$ fois et un F au rang ℓ. Ainsi $P((X, Y, Z) = (\ell, 2k + 1, 0)) = p^{\ell-k-1} q^{k+1}$.
• l'événement $(X, Y, Z) = (\ell, 2k + 1, 0)$ correspond à l'obtention d'une suite de FP k fois, puis d'une suite de F $(\ell - 2k - 1)$ fois et un P au rang ℓ. Ainsi $P((X, Y, Z) = (\ell, 2k + 1, 0)) = p^{k+1} q^{\ell-k-1}$.
On sait que $(X, Y))(\Omega) = \{(\ell, m) \in \mathbb{N}^{*2} \mid \ell \geq m + 2\}$. Donc, pour tout $1 \leq m \leq \ell - 2$:
$$P((X = \ell) \cap (Y = m)) = p^{k+1} q^{\ell-k-1} + q^{k+1} p^{\ell-k-1}$$
2. Lorsque $p = 1/2$, pour tout $1 \leq m \leq \ell - 2$:
$$P((X = \ell) \cap (Y = m)) = \left(\frac{1}{2}\right)^{\ell-1}$$
et la loi de X est définie pour tout $\ell \geq 3$ par :
$$P(X = \ell) = \sum_{m=1}^{\ell-2} P((X = \ell) \cap (Y = m)) = (\ell - 2) \left(\frac{1}{2}\right)^{\ell-1}$$
On sait que pour $|x| < 1$:
$$\sum_{k=2}^{+\infty} k(k-1) x^{k-2} = \frac{2}{(1-x)^3} \Rightarrow \sum_{k=2}^{+\infty} k(k-1) x^{k-1} = \frac{2x}{(1-x)^3}$$

et

$$\sum_{k=2}^{+\infty} k(k-2)x^{k-1} = \sum_{k=2}^{+\infty} k(k-1)x^{k-1} - \sum_{k=2}^{+\infty} kx^{k-1} = \frac{2x}{(1-x)^3} - \frac{1}{(1-x)^2} + 1$$

Donc :

$$E(X) = \sum_{\ell=3}^{+\infty} \ell(\ell-2)\left(\frac{1}{2}\right)^{\ell-1} = 5$$

3. La fonction proposée simule le lancer d'une pièce avec la probabilité p d'obtenir Pile. Elle retourne la lettre p avec la probabilité p et la lettre f avec la probabilité $1-p$.

4. Voici un programme :

```
Program Pile_Face ;
Uses CRT ;
Var r : array[1..1000] of char ;
i,y : integer ;
Function f(p : real) : char ;
Var ok : boolean ;
Begin
ok :=random<=p ;
If ok Then f :='p' Else f :='f'
End ;
Begin
Randomize ;
r[1] :=f(0.8) ; i := 1 ;
Write (r[i]) ;
Repeat
i :=i+1 ; r[i] :=f(0.8) ; write r[i]
Until r[i-1]=r[i] ;
y :=i-1 ;
Repeat
i :=i+1 ; r[i] :=f(0.8) ; write r[i]
Until r[i-1]<>r[i] ;
Write ('  ',i,'   ',y)
End.
```

Exercice 3-17

Soit $n \in \mathbb{N}^*$. Un jeu consiste à tirer un numéro parmi les nombres $\{0, \ldots, n\}$. Le joueur gagne la somme x égale au numéro tiré si ce numéro est pair, ou perd cette somme x si le numéro tiré est impair.

1. On suppose que le numéro tiré suit une loi uniforme sur $\{0, ..., n\}$.
Déterminer la loi du gain (positif ou négatif) du joueur, et son espérance.
Quelle est la probabilité que le joueur gagne à ce jeu ?
Y-a-t-il des valeurs de n qui optimisent l'espérance du gain ?

2. Reprendre les questions précédentes dans le cas où le numéro tiré suit une loi binomiale de paramètres n et $1/2$.

Solution :

1. On a immédiatement $X(\Omega) = \{(-1)^k k, 0 \leq k \leq n\}$. La variable aléatoire X prend chacune de ces valeurs avec la probabilité $\dfrac{1}{n+1}$. Ainsi :

$$E(X) = \sum_{k=0}^{n} (-1)^k \frac{k}{n+1}$$

Donc

- si $n = 2p, E(X) = \dfrac{p}{2p+1}$,
- si $n = 2p+1, E(X) = -\dfrac{1}{2}$.

De même :

- si $n = 2p, P(X > 0) = \displaystyle\sum_{k=1}^{p} P(X = 2k) = \frac{p}{2p+1}$,
- si $n = 2p+1, P(X > 0) = \dfrac{p}{2(p+1)}$.

La probabilité que le joueur gagne de l'argent est inférieure à $\dfrac{1}{2}$.

2. Dans cette question $X(\Omega)$ n'a pas changé, mais pour tout $0 \leq k \leq n$,

$$P(X = (-1)^k k) = \frac{C_n^k}{2^n}$$

Ainsi

$$E(X) = \sum_{k=1}^{n} (-1)^k \frac{kC_n^k}{2^n} = \left(\frac{1}{2}\right)^n \sum_{k=1}^{n} (-1)^k n C_{n-1}^{k-1} = \begin{cases} -1/2 & \text{si } n = 1 \\ 0 & \text{si } n > 1 \end{cases}$$

Le jeu est donc équilibré si $n > 1$. Enfin :

$$P(X > 0) = \sum_{k \geq 1} P(X = 2k) = \sum_{k \geq 1} \frac{C_n^{2k}}{2^n} = \frac{2^{n-1} - 1}{2^n} = \frac{1}{2} - \frac{1}{2^n}$$

Là encore, la probabilité que le joueur gagne de l'argent est inférieure à $\frac{1}{2}$.

Exercice 3-18

Soient $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$, $n + m$ variables aléatoires indépendantes suivant une loi de Bernoulli de paramètre inconnu p.
On se propose d'estimer p. On suppose dans la suite que $n > m$.
Soit $M_1 = \frac{1}{n} \sum_{k=1}^{n} X_k$, $M_2 = \frac{1}{m} \sum_{k=1}^{m} Y_k$ et $N = \frac{M_1 + M_2}{2}$.
1. Vérifier que M_1, M_2 et N sont des estimateurs sans biais et convergents de p.
Quel est le meilleur des trois ? (c'est-à-dire, celui dont la variance est la plus petite)
On discutera suivant les valeurs de n et m.

2. Parmi les estimateurs de p de la forme $aM_1 + bM_2$ quel est le meilleur estimateur sans biais ?

Solution :

1. De façon évidente, $E(M_1) = E(M_2) = E(N) = p$. Ces trois estimateurs sont sans biais. De plus :

$V(M_1) = \frac{pq}{n}$, $V(M_2) = \frac{pq}{m}$, $V(N) = \frac{1}{4}(V(M_1) + V(M_2)) = \frac{pq}{4}\left(\frac{1}{n} + \frac{1}{m}\right)$

- si $n > m, V(M_1) < V(M_2)$ et M_1 est un meilleur estimateur que M_2.
- $\frac{V(N)}{V(M_2)} = \frac{m+n}{4n} < 1$ car $n > m$. Ainsi N est un meilleur estimateur que M_2.
- $\frac{V(N)}{V(M_1)} = \frac{m+n}{4m}$. Donc si $m < n < 3m$, N est meilleur que M_1 lui même meilleur que M_2. Si $3m < n$, M_1 est meilleur que N lui même meilleur que M_2.

2. On sait que $E(aM_1 + bM_2) = (a+b)p$. Donc $aM_1 + bM_2$ est un estimateur sans bais de p si et seulement si $a + b = 1$. De plus :

$$V(aM_1 + bM_2) = a^2 \frac{pq}{n} + b^2 \frac{pq}{m}$$

Il faut donc minimiser cette fonction de (a, b) sous la contrainte $a + b = 1$. On étudie donc :

$$F : a \mapsto F(a) = a^2 \frac{pq}{n} + (1-a)^2 \frac{pq}{m}$$

qui présente un minimum pour $a = \dfrac{n}{n+m}$. Le meilleur estimateur sans biais du type $aM_1 + bM_2$ est donc $\dfrac{n}{n+m} M_1 + \dfrac{m}{n+m} M_2$.

Exercice 3-19

1. On considère la fonction f définie sur $\mathbb{R}$ par $t \mapsto f(t) = \frac{1}{2}\exp(-|t|)$.

Montrer que f est la densité de probabilité d'une variable aléatoire X dont on déterminera la fonction de répartition, l'espérance et l'écart-type.
On dit alors que X suit la loi de Laplace.

2. Soient X_1 et X_2 deux variables aléatoires indépendantes, de même loi que X.
Déterminer une densité $g : x \mapsto g(x)$ de $Y = X_1 + X_2$. On distinguera les cas $x < 0$ et $x > 0$.
Quelles sont les valeurs de l'espérance et de l'écart-type de Y ?

3. a) Donner sans calculs une densité de la variable aléatoire $X_1 - X_2$.

b) Dans un laboratoire, un technicien effectue indépendamment l'une de l'autre, deux pesées successives d'un même objet. On suppose que l'erreur qu'il commet à chaque pesée suit une loi de Laplace.
Soit a réel positif. Quelle est la probabilité que les deux mesures effectuées ne diffèrent pas de plus de a ?

c) Pour quelles valeurs de a l'inégalité de Bienaymé-Chebychev donne-t-elle une information sur cette probabilité ?

Solution :

1. La fonction f est continue sur $\mathbb{R}$, positive ; il suffit donc de vérifier qu'elle est d'intégrale 1. La convergence est évidente et, par parité :

$$\int_{-\infty}^{+\infty} f(t)\, dt = \frac{2}{2}\int_0^{+\infty} \mathrm{e}^{-t}\, dt = 1$$

Soit F la fonction de répartition de X, on a :

Si $x \leq 0$, $F(x) = \dfrac{1}{2}\displaystyle\int_{-\infty}^{x} \mathrm{e}^{t}\, dt = \frac{1}{2}\,\mathrm{e}^{x}$

Si $x \geq 0$, par symétrie, $F(x) = 1 - F(-x) = 1 - \frac{1}{2}\,\mathrm{e}^{-x}$.

$\star$ La convergence (absolue) de l'intégrale définissant l'espérance est banale et, par symétrie, l'espérance est nulle.

$\star$ Enfin la variance est égale au moment d'ordre 2 et :

$$V(X) = \int_{-\infty}^{+\infty} \frac{1}{2}\,t^2\mathrm{e}^{-|t|}\,dt = \int_0^{+\infty} t^2\mathrm{e}^{-t}\,dt = \Gamma(3) = 2$$

2. Une densité g de Y est donnée, par convolution, par :

$$g(x) = \int_{-\infty}^{+\infty} \frac{1}{4}\mathrm{e}^{-|t|}\mathrm{e}^{-|x-t|}\,dt$$

Pour $x < 0$, $4g(x) = \displaystyle\int_{-\infty}^{x} \mathrm{e}^{2t-x}\,dt + \int_x^0 \mathrm{e}^{x}\,dt + \int_0^{+\infty} \mathrm{e}^{x-2t}\,dt$

D'où : $g(x) = \frac{1}{4}(1-x)\mathrm{e}^{x}$

Tandis que pour $x > 0$: $4g(x) = \displaystyle\int_{-\infty}^{0} \mathrm{e}^{2t-x}\,dt + \int_0^x \mathrm{e}^{-x} + \int_x^{+\infty} \mathrm{e}^{x-2t}\,dt$

D'où : $g(x) = \frac{1}{4}(x+1)\mathrm{e}^{-x}$.

On peut donc écrire : $\forall\, x \in \mathbb{R}, g(x) = \frac{1}{4}(1+|x|)\mathrm{e}^{-|x|}$.

Par linéarité de l'espérance : $E(Y) = 0$ et, par indépendance :

$V(Y) = V(X_1) + V(X_2) = 4$.

3. a) La variable $-X_2$ a même loi que X_2, donc, toujours par indépendance, $X_1 - X_2$ a même loi que $Y = X_1 + X_2$.

b) Soit X_1 l'erreur aléatoire commise lors de la première mesure et X_2 celle commise lors de la seconde. On cherche la probabilité de l'événement $|X_1 - X_2| \leq a$, c'est-à-dire $-a \leq X_1 - X_2 \leq a$. On peut donc écrire :

$$p = P(-a \leq X_1 - X_2 \leq a) = \int_{-a}^{a} g(x)\,dx = \frac{1}{2}\int_0^a (x+1)\mathrm{e}^{-x}\,dx$$

Une intégration par parties donne la solution et :

$$p = \frac{1}{2}\big(2 - (a+2)\mathrm{e}^{-a}\big)$$

c) L'inégalité de Bienaymé-Chebychev donne une indication dès que la largeur de l'intervalle considéré est égale au moins à deux fois l'écart-type de Y, c'est-à-dire pour $a \geq 2$.

Exercice 3-20

Soient X et Y des variables aléatoires indépendantes suivant la même loi exponentielle de paramètre 1.

Soit $Z = \dfrac{X}{X+Y}$.

1. Pour tout $t \in [0,1]$, déterminer une densité sur $\mathbb{R}^+$ de la variable $t(X+Y) - X$ (qui s'écrit également $(t-1)X + tY$).

2. En déduire la loi de Z.

Solution :

1. Y suit la loi exponentielle de paramètre 1, donc tY suit la loi exponentielle de densité $f(u) = \frac{1}{t}\exp(-\frac{u}{t})$, pour $u \geq 0$ et $f(u) = 0$, pour $u < 0$ (écrire $P(tY \leq u) = P(Y \leq \frac{u}{t})$, ce qui donne la fonction de répartition et dériver).

De la même façon $(1-t)X$ suit la loi exponentielle de paramètre $\dfrac{1}{1-t}$ et $(t-1)X$, qui est son opposée, admet pour densité la fonction g définie par : $g(u) = \dfrac{1}{1-t}\exp(\dfrac{u}{1-t})$, si $u \leq 0$ et $g(u) = 0$, si $u > 0$.

Une densité h de $T = (1-t)X + tY$ est donc donnée par convolution :

$$\forall\, x \in \mathbb{R}, h(x) = \int_{-\infty}^{+\infty} f(x-u)g(u)\,du$$

La fonction à intégrer est non nulle si et seulement si $x - u \geq 0$ et $u \leq 0$. Donc, pour $x \geq 0$:

$$\begin{aligned} h(x) &= \int_{-\infty}^{0} \frac{1}{t}\exp(-\frac{x-u}{t})\frac{1}{1-t}\exp(\frac{u}{1-t})\,du \\ &= \frac{1}{t(1-t)}\exp(-\frac{x}{t})\int_{-\infty}^{0}\exp(\frac{u}{t(1-t)})\,du = \exp(-\frac{x}{t}) \end{aligned}$$

2. La variable Z prend clairement ses valeurs entre 0 et 1 et, pour $t \in [0,1]$:

$$[Z \leq t] = [X \leq t(X+Y)] = [(t-1)X + tY \geq 0] = [T \geq 0]$$

C'est pour cela qu'il suffisait de connaître une densité sur $\mathbb{R}^+$ de T et :

$$P(Z \leq t) = P(T \geq 0) = \int_{0}^{+\infty} \exp(-\frac{x}{t})\,dx = t$$

La variable Z suit la loi uniforme sur $[0,1]$.

Exercice 3-21

Pour $a > 0$ et $b > 0$ on pose $B(a,b) = \displaystyle\int_0^1 u^{a-1}(1-u)^{b-1}du$.

1. a) Justifier l'existence de $B(a,b)$.

b) Calculer $B(a,b)$ lorsque a et b sont des entiers naturels non nuls.

2. On rappelle que X_a suit la loi $\Gamma(1,a)$ si une densité f_a de X_a est définie par :

$$f_a(t) = \begin{cases} \dfrac{1}{\Gamma(a)} t^{a-1} \mathrm{e}^{-t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

Pour $x > 0$, exprimer $\displaystyle\int_{-\infty}^{+\infty} f_a(t) f_b(x-t)\, dt$ en fonction de $B(a,b)$.

En déduire que pour $a > 0$ et $b > 0$, $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Exprimer $B(a+1,b)$ en fonction de $B(a,b)$.

3. A l'aide du changement de variable $u = 2t - 1$, calculer :

$$B(1/2, 1/2) = \int_0^1 \frac{dt}{\sqrt{t}\sqrt{1-t}}.$$

En déduire la valeur de $\Gamma(1/2)$.

4. Soit X une variable aléatoire de densité :

$$f(t) = \begin{cases} \dfrac{1}{B(1/2,1/2)} \cdot \dfrac{1}{\sqrt{t}\sqrt{1-t}} & \text{si } t \in \,]0,1[\\ 0 & \text{sinon} \end{cases}$$

Déterminer avec un minimum de calculs son espérance et sa variance.

Solution :

1. a) Si $a \geq 1$ et $b \geq 1$, l'intégrale ne pose pas de problème d'existence et si $0 < a < 1$ ou $0 < b < 1$, la règle de Riemann assure la convergence pour la borne concernée. Ainsi, par disjonction des cas, l'intégrale proposée converge.

b) Pour $b \geq 1$ et $a > 1$, une intégration par parties donne :

$$B(a,b) = \frac{a-1}{b} B(a-1, b+1)$$

Soit : $\qquad B(a,b) = \dfrac{a-1}{b}\, \dfrac{a-2}{b+1} \cdots \dfrac{1}{a+b-2} B(1, a+b-1)$

et comme $B(1, a+b-1) = \dfrac{1}{a+b-1}$, il vient :

$$\forall\, a, b \in \mathbb{N}^*,\, B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$$

2. On sait que si $X \hookrightarrow \Gamma(1,a)$ et $Y \hookrightarrow \Gamma(1,b)$ et si X et Y sont des variables aléatoires indépendantes, alors $X + Y \hookrightarrow \Gamma(1, a+b)$. Comme on sait qu'une

densité de $X+Y$ s'obtient alors par convolution, il vient, avec les notations de l'énoncé :

$$\forall\, x \in \mathbb{R}_+\,,\, f_{a+b}(x) = \int_{-\infty}^{+\infty} f_a(t) f_b(x-t)\, dt$$

Il n'y a du grain à moudre que si $t \in [0,x]$ et :

$$\forall\, x \in \mathbb{R}_+\,,\, f_{a+b}(x) = \int_0^x \frac{1}{\Gamma(a)} t^{a-1} \mathrm{e}^{-t} \frac{1}{\Gamma(b)} (x-t)^{b-1} \mathrm{e}^{t-x}\, dt$$

Soit : $\dfrac{1}{\Gamma(a+b)} x^{a+b-1} \mathrm{e}^{-x} = \dfrac{\mathrm{e}^{-x}}{\Gamma(a)\Gamma(b)} \displaystyle\int_0^x t^{a-1}(x-t)^{b-1}\, dt$

On effectue le changement de variable $t = xu$ et on obtient :

$$\frac{1}{\Gamma(a+b)} x^{a+b-1} \mathrm{e}^{-x} = \frac{\mathrm{e}^{-x}}{\Gamma(a)\Gamma(b)} x^{a+b-1} B(a,b), \text{ soit :}$$

$$\forall\, a,b \in \mathbb{R}_+^*\,,\, B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Ainsi : $B(a+1,b) = \dfrac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \dfrac{a}{a+b} B(a,b)$.

3. Le changement de variable donne : $B(1/2,1/2) = \displaystyle\int_{-1}^{1} \frac{du}{\sqrt{1-u^2}}$.

En posant $u = \sin t$, on obtient finalement $B(1/2,1/2) = \displaystyle\int_{-\pi/2}^{\pi/2} dt = \pi$.

Comme $B(1/2,1/2) = \dfrac{\Gamma^2(1/2)}{\Gamma(1)}$, on en déduit que $\Gamma(1/2) = \sqrt{\pi}$.

4. En écrivant l'intégrale idoine, on a : $E(X) = \dfrac{B(3/2,1/2)}{B(1/2,1/2)} = \dfrac{1/2}{1} = \dfrac{1}{2}$.

De même $E(X^2) = \dfrac{3}{8}$ et, par la formule de Koenig-Huygens : $V(X) = \dfrac{1}{8}$.

Exercice 3-22

Soit $a \in \,]0,1[$ et X une variable aléatoire, définie sur un espace probabilisé $(\Omega, \mathcal{B}, P)$, à valeurs dans $[0,1]$.

On suppose que la loi conditionnelle de X conditionnée par $[X \leqslant a]$ est la loi uniforme sur $[0,a]$ et que sa loi conditionnelle conditionnée par $[X > a]$ est la loi uniforme sur $[a,1]$.

On suppose de plus que $P[X \leqslant a] = \dfrac{1}{2}$.

1. Déterminer une densité de X, son espérance et sa variance.

On cherche maintenant à estimer le paramètre inconnu a.

2. Soit $(X_1, \dots, X_n)$ n variables aléatoires indépendantes de même loi que X. On pose $M_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$. Déterminer l'espérance $E(M_n)$ et la variance $V(M_n)$.
En déduire un estimateur sans biais T_n de a.
La suite d'estimateurs $(T_n)_{n \geq 1}$ est-elle convergente ?

3. Donner une majoration de $V(T_n)$ quand $a \in]0, 1[$.
Pour n et α donnés, préciser, en utilisant l'inégalité de Bienaymé-Chebychev, les valeurs de ε pour lesquelles $[T_n - \varepsilon, T_n + \varepsilon]$ est un intervalle de confiance au risque α pour a.

(c'est-à-dire : $P[T_n - \varepsilon \leqslant a \leqslant T_n + \varepsilon] \geqslant 1 - \alpha$)

4. Quelle est la limite en loi de la suite $\left(\dfrac{T_n - a}{\sqrt{V(T_n)}}\right)_{n \geq 1}$?

On suppose que n est suffisamment grand pour identifier la loi de $\dfrac{T_n - a}{\sqrt{V(T_n)}}$ à cette loi limite.
En déduire un intervalle de confiance pour a à un risque inférieur à 5%.

5. Au niveau de risque 5%, comparer les longueurs des deux intervalles de confiance trouvés dans les questions 3 et 4.

Solution :

1. Une densité de X vaut $\dfrac{1}{2a}$ sur $[0, a]$ et $\dfrac{1}{2(1-a)}$ sur $[a, 1]$ et 0 ailleurs.

En utilisant la formule de Chasles pour calculer les intégrales, on trouve alors facilement :

$$E(X) = \frac{2a+1}{4} \text{ et } V(X) = \frac{4 + (2a-1)^2}{48}$$

2. $E(M_n) = E(X) = \dfrac{2a+1}{4}$ et $V(M_n) = \dfrac{1}{n} V(X) = \dfrac{4 + (2a-1)^2}{48.n}$.

Ainsi $T_n = 2M_n - \dfrac{1}{2}$ est d'espérance a, donc est un estimateur sans biais de a. La variance de T_n vaut $4V(M_n)$, donc est de limite nulle et (T_n) est une suite d'estimateurs convergente.

3. On a : $V(T_n) = \dfrac{4 + (2a-1)^2}{12.n} \leq \dfrac{5}{12.n}$ et, grce à l'inégalité de Bienaymé-Chébychev :

$$\text{Pour tout } a \text{ de } [0,1],\ P(|T_n - a| \geq \varepsilon) \leq \frac{5}{12.n\varepsilon^2}$$

Si on veut que $[T_n - \varepsilon, T_n + \varepsilon]$ soit un intervalle de confiance au risque α, il suffit de prendre ε tel que $\dfrac{5}{12.n\varepsilon^2} \leq \alpha$, donc tel que $\varepsilon \geq \sqrt{\dfrac{5}{12.n\alpha}}$.

4. Le théorème de la limite centrée indique que $\left(\dfrac{T_n - a}{\sigma(T_n))}\right)_n$ converge en loi vers une variable suivant la loi normale centrée réduite.

Soit alors $t_{\alpha/2}$ le quantile d'ordre $\alpha/2$ de la loi normale centrée réduite. On a :

$$P[T_n - t_{\alpha/2}\sigma(T_n) \leq a \leq T_n + t_{\alpha/2}\sigma(T_n)] = 1 - \alpha$$

L'intervalle précédent n'est pas observable, puisque $\sigma(T_n)$ dépend du paramètre inconnu a, mais $V(T_n)$ étant majoré par $\dfrac{5}{12.n}$, cet intervalle est toujours inclus dans l'intervalle $\left[T_n - t_{\alpha/2}\sqrt{\dfrac{5}{12.n}}, T_n + t_{\alpha/2}\sqrt{\dfrac{5}{12.n}}\right]$, qui est, lui, observable et est *a fortiori* un intervalle de confiance pour a de risque inférieur à α.

Pour $\alpha = 5\%$, on a $t_{\alpha/2} = 1{,}96$.

5. Le rapport de longueur des deux intervalles de confiance est $\sqrt{\alpha}.t_{\alpha/2}$. Pour $\alpha = 5\%$, il vaut approximativement 0,44.

Exercice 3-23

1. Montrer que pour tout n de $\mathbb{N}$, la fonction f_n définie par :

$$f_n(x) = \begin{cases} (n+1)(1-x)^n & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

est une densité de probabilité.

Soit $Y_n = nX_n$ où X_n est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{B}, P)$ et qui admet f_n pour densité.

Pour $x \in [0,n]$, calculer $P[Y_n \leqslant x]$. En déduire que la suite de variables $(Y_n)_{n\in\mathbb{N}}$ converge en loi vers une variable suivant une loi exponentielle.

2. Montrer que pour tout n de $\mathbb{N}$, la fonction f_n définie par :

$$f_n(x) = \begin{cases} (n+1)(n+2)x(1-x)^n & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

est une densité de probabilité.

Soit $Y_n = nX_n$ où X_n admet pour densité f_n.

Pour $x \in [0,n]$, montrer que $P[Y_n \leqslant x] = 1 - \left(\dfrac{n+1}{n}x + 1\right)\left(1 - \dfrac{x}{n}\right)^{n+1}$.

En déduire que la suite de variables $(Y_n)_{n\geq 0}$ converge en loi vers une variable dont on précisera la loi.

Solution :

1. La fonction f_n est positive, continue sauf au point 0 et :

$$\int_{-\infty}^{+\infty} f_n(t)\,dt = (n+1)\int_0^1 (1-t)^n\,dt = 1$$

Toutes les conditions sont réunies pour affirmer que f_n est bien une densité de probabilité.

La variable aléatoire Y_n prend ses valeurs entre 0 et n et, pour $x \in [0,n]$:

$$P(Y_n \leq x) = P(X_n \leq \frac{x}{n}) = (n+1)\int_0^{x/n} (1-t)^n\,dt = 1-(1-\frac{x}{n})^{n+1}$$

$\star$ En notant F_n la fonction de répartition de Y_n, on a donc : $\lim\limits_{n\to\infty} F_n(x) = 0$, si $x < 0$; $\lim\limits_{n\to\infty} F_n(x) = 1 - \mathrm{e}^{-x}$, si $x \geq 0$.

(En effet $\lim\limits_{n\to\infty}(1-\frac{x}{n})^n = \mathrm{e}^{-x}$, comme on le voit facilement en considérant les logarithmes et si $x \geq 0$, alors pour n assez grand, on a $0 \leq x \leq n$)

Par conséquent, la suite (Y_n) converge donc en loi vers une variable suivant la loi exponentielle de paramètre 1.

2. $\star$ A nouveau, f_n est positive, continue sauf au point 0 et :

$$\int_{-\infty}^{+\infty} f_n(t)\,dt = (n+1)(n+2)\int_0^1 t(1-t)^n\,dt = (n+1)(n+2)\int_0^1 u^n(1-u)\,du = 1$$

Donc f_n est une densité de probabilité.

$\star$ De la même façon, Y_n prend ses valeurs entre 0 et n et, pour $x \in [0,n]$:

$$\begin{aligned} P(Y_n \leq x) = P(X_n \leq \frac{x}{n}) &= (n+1)(n+2)\int_0^{x/n} t(1-t)^n\,dt \\ &= (n+1)(n+2)\int_{1-\frac{x}{n}}^1 u^n(1-u)\,du \end{aligned}$$

soit :

$$P(Y_n \leq x) = (n+2)\big[1-(1-\frac{x}{n})^{n+1}\big] - (n+1)\big[1-(1-\frac{x}{n})^{n+2}\big]$$

Ce qui s'écrit aussi :

$$P(Y_n \leq x) = 1 - \big(\frac{n+1}{n}\,x+1\big)(1-\frac{x}{n})^{n+1}$$

Pour $x \geq 0$, on a donc $\lim\limits_{n\to\infty} P(Y_n \leq x) = 1-(x+1)\mathrm{e}^{-x}$, tandis que pour $x < 0$, on a : $\lim\limits_{n\to\infty} P(Y_n \leq x) = 0$.

En dérivant, ou directement, on reconnaît dans la loi limite la loi Gamma de paramètres $b = 1$ et $\tau = 2$.

Exercice 3-24

On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, définies sur le même espace probabilisé $(\Omega, \mathcal{B}, P)$, et suivant toutes la loi uniforme sur l'intervalle $[0, 1]$.

1. Déterminer une densité de la variable aléatoire $Y_k = \max(X_1, ..., X_k)$, $k \geqslant 1$.

2. Déterminer une densité de la variable aléatoire $Z_k = -Y_k$.

3. En déduire la probabilité de l'événement $A_n = \{X_n \geqslant \max(X_1, ..., X_{n-1})\}$ pour $n \geqslant 2$.

Solution :

1. La variable Y_k prend ses valeurs entre 0 et 1, et pour $t \in [0, 1]$, on a, par indépendance et équirépartition :

$$P(Y_k \leq t) = P(X_1 \leq t).P(X_2 \leq t) \cdots P(X_k \leq t) = t^k$$

On en déduit qu'une densité f_{Y_k} de Y_k est donnée par :

$$f_{Y_k}(t) = k.t^{k-1}, \text{ si } 0 \leq t \leq 1\,;\ f_{Y_k}(t) = 0, \text{ sinon}$$

2) La variable Z_k prend ses valeurs entre -1 et 0, et pour $t \in [-1, 0]$:

$$P(Z_k \leq t) = P(Y_k \geq -t) = 1 - P(Y_k \leq -t) = 1 - (-t)^k$$

Une densité f_{Z_k} de Z_k est donc donnée par :

$$f_{Z_k}(t) = k.(-t)^{k-1}, \text{ si } -1 \leq t \leq 0\,;\ f_{Z_k}(t) = 0, \text{ sinon}$$

3) On a : $P(A_n) = P(X_n + Z_{n-1} \geq 0)$, et la variable aléatoire X_n étant indépendante de toute fonction de $X_1, \ldots, X_{n-1}$, la variable X_n est indépendante de Z_{n-1}. D'où, par convolution :

$$P(A_n) = \int_0^{+\infty} \Big[\int_{-\infty}^{+\infty} f_{Z_{n-1}}(t) f_{X_n}(x - t)\, dt\Big] dx$$

Mais $f_{Z_{n-1}}$ est nulle en dehors de $[-1, 0]$, donc :

$$\int_{-\infty}^{+\infty} f_{Z_{n-1}}(t) f_{X_n}(x - t)\, dt = \int_{-1}^{0} (n - 1)(-t)^{n-2} f_{X_n}(x - t)\, dt$$

Puis f_{X_n} vaut 1 si $x - t$ appartient à $[0, 1]$ et 0 sinon. L'intégrale précédente est donc nulle si $x \notin [0, 1]$, et si $x \in [0, 1]$ elle se réduit à $\displaystyle\int_{x-1}^{0} (n-1)(-t)^{n-2}\, dt$ et vaut donc $(1 - x)^{n-1}$.

Ainsi : $P(A_n) = \int_0^1 (1-x)^{n-1}\,dx = \frac{1}{n}$.

Exercice 3-25

1. Montrer que l'application $x \mapsto x^2$ est convexe.

2. Trois autobus contiennent respectivement n_1, n_2, n_3 passagers en plus de leur chauffeur. On considère deux expériences :

- dans la première on choisit au hasard un des trois chauffeurs, et on note X la variable aléatoire égale au nombre de passagers de l'autobus qu'il conduit.
- dans la seconde, on choisit au hasard un des $n_1 + n_2 + n_3$ passagers et on note Y le nombre de passagers de l'autobus dans lequel ce passager se trouve.

Comparer les espérances des variables aléatoires X et Y.

Solution :

1. La fonction $g : x \mapsto x^2$ est de classe $\mathcal{C}^2$ et $g''(x) = 2 > 0$, donc g est convexe sur $\mathbb{R}$.

2. $\star$ La variable aléatoire X prend les valeurs n_1, n_2, n_3 avec la même probabilité $\frac{1}{3}$, donc $E(X) = \frac{1}{3}(n_1 + n_2 + n_3)$ (on suppose donc implicitement que n_1, n_2, n_3 sont deux à deux distincts, mais le résultat concernant l'espérance reste valide dans tous les cas).

$\star$ La variable aléatoire Y prend les mêmes valeurs et $P(Y = n_i) = \dfrac{n_i}{n_1 + n_2 + n_3}$, d'où $E(Y) = \dfrac{n_1^2 + n_2^2 + n_3^2}{n_1 + n_2 + n_3}$ (même remarque que pour la variable X).

$\star$ La convexité de g permet d'affirmer que l'on a :

$$g\left(\frac{n_1 + n_2 + n_3}{3}\right) \leq \frac{g(n_1) + g(n_2) + g(n_3)}{3}$$

C'est-à-dire : $\dfrac{(n_1 + n_2 + n_3)^2}{9} \leq \dfrac{n_1^2 + n_2^2 + n_3^2}{3}$, *i.e.* $E(X) \leq E(Y)$.

Exercice 3-26

Soient a et b deux réels positifs distincts.
On considère X et Y deux variables aléatoires indépendantes suivant des lois de Pareto de densités respectives f_a et f_b définies par :

$$f_a(t) = \begin{cases} at^{-a-1} & \text{si } t \geq 1 \\ 0 & \text{si } t < 1 \end{cases}, \qquad f_b(t) = \begin{cases} bt^{-b-1} & \text{si } t \geq 1 \\ 0 & \text{si } t < 1 \end{cases}$$

1. Reconnaître les lois de $\ln(X)$ et de $\ln(Y)$.

2. En déduire la loi de $Z = XY$.

Solution :

1. Les variables aléatoires X et Y prennent leurs valeurs dans $[1, +\infty[$, donc $\ln X$ et $\ln Y$ prennent leurs valeurs dans $\mathbb{R}_+$. Pour $x \geq 0$, on a :

$$P(\ln X \leq x) = P(X \leq \mathrm{e}^x) = \int_1^{\mathrm{e}^x} a.t^{-a-1}\, dt = 1 - \mathrm{e}^{-ax}$$

Donc $\ln X$ suit la loi exponentielle de paramètre a. De même $\ln Y$ suit la loi exponentielle de paramètre b. Une densité g_a de $\ln X$ est la fonction définie par :

$$\text{Si } x \geq 0,\ g_a(x) = a\mathrm{e}^{-ax} \text{ et sinon } g_a(x) = 0$$

On notera de même g_b une densité de $\ln Y$.

2. Les variables aléatoires X et Y étant indépendantes, il en est de même des variables aléatoires $\ln X$ et $\ln Y$. Une densité g de $\ln Z = \ln X + \ln Y$ s'obtient donc par convolution :

$$g(x) = \int_{-\infty}^{+\infty} g_a(t) g_b(x-t)\, dt$$

$\ln Z$ prenant ses valeurs dans $\mathbb{R}_+$, on se limite au cas $x \geq 0$ et l'intégrale devient :

$$g(x) = \int_0^x a\mathrm{e}^{-at} b\mathrm{e}^{-b(x-t)}\, dt = ab\frac{\mathrm{e}^{-ax} - \mathrm{e}^{-bx}}{b-a}$$

Enfin, pour $x \geq 1$:

$$\begin{aligned} P(Z \leq x) = P(\ln Z \leq \ln x) &= \int_0^{\ln x} ab\frac{\mathrm{e}^{-at} - \mathrm{e}^{-bt}}{b-a}\, dt \\ &= \frac{ab}{b-a}\left[\frac{1}{b}(x^{-b} - 1) - \frac{1}{a}(x^{-a} - 1)\right] \end{aligned}$$

Donc une densité de Z sur $[1, +\infty[$ est $x \mapsto \dfrac{ab}{b-a}(x^{-a-1} - x^{-b-1})$.

Exercice 3-27

Une urne contient n boules numérotées de 1 à n. On tire plusieurs fois au hasard et avec remise une boule de l'urne. On arrête les tirages dès que le dernier numéro obtenu est supérieur ou égal au numéro obtenu lors du précédent tirage. On appelle X la variable aléatoire égale au nombre de tirages effectués.

1. Déterminer $X(\Omega)$.

2. a) Déterminer la probabilité des événements $X \geq 2$, $X \geq 3$ et $X = 2$.

b) Pour tout $k \in X(\Omega)$, déterminer la probabilité de l'événement $(X \geq k)$.

c) En déduire la loi et l'espérance de X, puis la limite de cette dernière lorsque n tend vers l'infini.

3. On veut simuler l'expérience précédente lorsqu'il y a 10 boules dans l'urne.

Compléter les lignes manquantes du programme Pascal suivant afin qu'il affiche la série de numéros obtenus ainsi que la valeur de X correspondante :

```
Var u : array[1..11] of integer ;
......
Begin
u[1] :=1+random(10) ;
......
......
End.
```

L'affichage à l'écran se fera sous la forme 5 3 2 6 $X = 4$.

On précise que l'instruction `x :=random(n)` met dans la variable x déclarée de type `integer` une valeur « au hasard » entre 0 et $n-1$ (chaque appel à `random(n)`, pour n fixé renvoie une nouvelle valeur, indépendante des valeurs déjà obtenues).

Solution :

1. On effectue au minimum deux tirages et au maximum $n+1$ (lorsque les n premiers tirages ont amené, dans cet ordre, les résultats $n, n-1, n-2, \ldots, 1$). Toutes les situations intermédiaires étant possibles, on a donc $X(\Omega) = [\![2, n+1]\!]$.

2. a) $P(X \geq 2) = 1$, $P(X \geq 3) = \dfrac{C_n^2}{n^2}$, car l'événement $(X \geq 3)$ est réalisé lorsque les numéros n_1, n_2 obtenus aux deux premiers tirages vérifient $n_1 > n_2$. Or il existe n^2 façons de tirer deux fois un numéro, toutes équiprobables, et C_n^2 façons de choisir un couple (n_1, n_2) de $[\![1, n]\!]^2$ tel que $n_1 > n_2$.

On en déduit : $P(X = 2) = P(X \geq 2) - P(X \geq 3) = 1 - \dfrac{C_n^2}{n^2}$.

b) Le même raisonnement montre que $P(X \geq k) = \dfrac{C_n^{k-1}}{n^{k-1}}$.

c) Par conséquent $P(X=k)=\dfrac{C_n^{k-1}}{n^{k-1}}-\dfrac{C_n^k}{n^k}$, le résultat étant valable même pour $k=n+1$, avec les conventions habituelles concernant les coefficients binomiaux, et aussi pour $k=1$ (ce qui donne bien une probabilité nulle).

On a :

$E(X)=\sum\limits_{k=2}^{n+1} k.P(X=k)=\sum\limits_{k=2}^{n+1} k.P(X\geq k)-\sum\limits_{k=2}^{n+1} k.P(X\geq k+1)$

Un changement d'indice dans la première somme donne alors :

$E(X)=\sum\limits_{k=1}^{n}(k+1)P(X\geq k+1)-\sum\limits_{k=2}^{n+1} k.P(X\geq k+1)$

D'où, après regroupements et simplifications :

$$E(X)=\sum_{k=0}^{n}\frac{C_n^k}{n^k}=(1+\frac{1}{n})^n \underset{n\to\infty}{\longrightarrow} \mathrm{e}$$

3.

```
PROGRAM boules ; uses crt ;
Var u :Array[1..11] of integer ; i : integer ;
Begin
(randomize ;
u[1] :=1+random(10) ; i :=1 ; write(u[1], ' ') ;
repeat i :=i+1 ;
   u[i] :=1+random(10) ;
   write(u[i],' ') ;
   until u[i]>=u[i-1] ;
writeln ('X=',i) ; readkey ;
End.
```

Exercice 3-28

Dans cet exercice, on considère trois variables aléatoires Y, X_1, X_2 admettant des moments d'ordre deux, et on s'intéresse à l'existence et à l'unicité d'un *minimum* pour la quantité

$$E\big((Y-aX_1-bX_2)^2\big)$$

lorsque (a,b) parcourt $\mathbb{R}^2$, E désignant l'espérance.

1. Soit $A=\begin{pmatrix} x & y \\ z & t\end{pmatrix}\in\mathcal{M}_2(\mathbb{R})$. Montrer que :

$$(A \text{ est inversible}) \iff (xt-yz\neq 0)$$

2. Dans cette question et la suivante, on suppose que les variables aléatoires X_1 et X_2 vérifient $E(X_1^2).E(X_2^2) - (E(X_1X_2))^2 \neq 0$.

a) Montrer que $E(X_1^2)$ et $E(X_2^2)$ sont des réels strictement positifs.

b) Plus généralement, montrer que si $(a,b) \in \mathbb{R}^2 \setminus \{(0,0)\}$, $E\big((aX_1+bX_2)^2\big)$ est un réel strictement positif.

3. On définit l'application : $f : \mathbb{R}^2 \to \mathbb{R}, (a,b) \mapsto f(a,b) = E\big((Y - aX_1 - bX_2)^2\big)$

a) Montrer que f est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$, et qu'il existe un unique couple $(a_0,b_0) \in \mathbb{R}^2$ tel que : $\dfrac{\partial f}{\partial a}(a_0,b_0) = \dfrac{\partial f}{\partial b}(a_0,b_0) = 0$.

b) Montrer que, dans ces conditions, on a $\forall\,(a,b) \in \mathbb{R}^2$,

$$E\big((Y - a_0X_1 - b_0X_2)(aX_1 + bX_2)\big) = 0$$

puis que

$$E\big((Y-aX_1-bX_2)^2\big) = E\big((Y-a_0X_1-b_0X_2)^2\big)+E\big(((a_0-a)X_1+(b_0-b)X_2)^2\big)$$

c) Etudier les extremums de f.

4. **Cas particulier**.
On suppose ici que X_1 et X_2 sont deux variables aléatoires à densité, indépendantes, suivant une loi uniforme sur $[0,1]$ et on pose $Y = X_1^2$. Déterminer (a_0,b_0) pour lequel $E\big((Y - a_0X_1 - b_0X_2)^2\big)$ est *minimum*.

Pour cela, on admettra que la propriété relative à l'espérance d'un produit de variables aléatoires discrètes indépendantes reste vraie pour des variables aléatoires à densité indépendantes.

Solution :

1. Ce résultat est connu.

2. a) Si $E(X_1^2) = 0$, alors $P(X_1 = 0) = 1$, donc $P(X_1X_2 = 0) = 1$ et $E(X_1X_2) = 0$, ce qui contredit l'hypothèse. On conclut de la même façon pour la variable aléatoire X_2.

b) Supposons b non nul, alors :

$$\forall\, a \in \mathbb{R}, E\big((aX_1 + bX_2)^2\big) = a^2E(X_1^2) + 2abE(X_1X_2) + b^2E(X_2^2) \geq 0$$

Par conséquent : $\Delta' = b^2[E^2(X_1X_2) - E(X_1^2)E(X_2^2)] \leq 0$.
Par hypothèse ce discriminant n'est pas nul, donc il est strictement négatif et le trinôme est toujours strictement positif.
Enfin, si $b = 0$, il reste $E(a^2X_1^2) > 0$ et la conclusion reste valable.

3. a) On a :

$f(a,b)=E(Y^2)+a^2E(X_1^2)+b^2E(X_2^2)+2abE(X_1X_2)-2aE(X_1Y)-2bE(X_2Y)$

La fonction f est polynomiale, donc de classe $\mathcal{C}^\infty$ et :

$$\begin{cases} \dfrac{\partial f}{\partial a}(a,b) = 2aE(X_1^2) + 2bE(X_1X_2) - 2E(X_1Y) \\ \dfrac{\partial f}{\partial b}(a,b) = 2bE(X_2^2) + 2aE(X_1X_2) - 2E(X_2Y) \end{cases}$$

Il existe un unique point critique (a_0, b_0), car la matrice du système correspondant est inversible, d'après l'hypothèse de l'énoncé.

b) $\star$ $\dfrac{\partial f}{\partial a}(a_0, b_0) = 0$, donc $E\big((Y - a_0X_1 - b_0X_2)X_1\big) = 0$ et, de la même façon $E\big((Y - a_0X_1 - b_0X_2)X_2\big) = 0$. Par linéarité de l'espérance, on a donc :

$$E\big((Y - a_0X_1 - b_0X_2)(aX_1 + bX_2)\big) = 0$$

$\star$ On écrit $Y - aX_1 - bX_2 = Y - a_0X_1 - b_0X_2 + (a_0 - a)X_1 + (b_0 - b)X_2$, on développe alors le carré et le résultat précédent montre que l'espérance du « double produit » est nul, ce qui donne la formule voulue.

c) Ainsi $f(a,b) = f(a_0,b_0) + E\big(((a_0 - a)X_1 + (b_0 - b)X_2)^2\big) \geq f(a_0, b_0)$ l'égalité ne pouvant avoir lieu que si l'espérance du terme complémentaire est nulle, c'est-à-dire pour $(a,b) = (a_0, b_0)$.

La fonction f admet un minimum absolu atteint uniquement en (a_0, b_0). Comme f n'a pas d'autre point critique, elle ne possède pas d'autre extremum local et *a fortiori* pas de maximum absolu.

4. On a :

$E(X_1^2) = E(X_2^2) = \dfrac{1}{3}$ et, par indépendance $E(X_1X_2) = E(X_1)E(X_2) = \dfrac{1}{4}$.

On est bien dans les conditions de l'exercice et de plus $E(X_1Y) = E(X_1^3) = \dfrac{1}{4}$, $E(X_2Y) = E(X_2)E(X_1^2) = \dfrac{1}{6}$.

Par conséquent $E\big((Y - aX_1 - bX_2)^2\big)$ est minimum au point (a_0, b_0) solution du système $\begin{cases} \dfrac{a_0}{3} + \dfrac{b_0}{4} = \dfrac{1}{4} \\ \dfrac{a_0}{4} + \dfrac{b_0}{3} = \dfrac{1}{6} \end{cases}$

Soit pour $(a_0, b_0) = \big(\frac{6}{7}, -\frac{1}{7}\big)$.

Exercice 3-29

Soit X une variable aléatoire réelle suivant la loi exponentielle de paramètre 1.

On pose $Y = \ln(\mathrm{e}^X - 1)$.

1. Déterminer la fonction de répartition de Y et une densité de Y.

2. Calculer $E(Y)$.

Solution :

1. La variable aléatoire X prend ses valeurs dans $\mathbb{R}_+$, donc $\mathrm{e}^X - 1$ aussi et puisque l'événement $(X = 0)$ est quasi-imposible, Y est définie et prend ses valeurs dans $\mathbb{R}$.

Soit alors $y \in \mathbb{R}$, on a :

$$\begin{aligned} P(Y \leq y) &= P(\mathrm{e}^X \leq \mathrm{e}^y + 1) = P(X \leq \ln(\mathrm{e}^y + 1)) = 1 - \exp(-\ln(\mathrm{e}^y + 1)) \\ &= \frac{\mathrm{e}^y}{\mathrm{e}^y + 1} \end{aligned}$$

Une densité de Y est donc la fonction f_Y définie sur $\mathbb{R}$ par :

$$f_Y(y) = \frac{\mathrm{e}^y}{(\mathrm{e}^y + 1)^2}$$

2. Sous réserve de convergence (absolue) , l'espérance de Y est donnée par :

$$E(Y) = \int_{-\infty}^{+\infty} \frac{x.\mathrm{e}^x}{(\mathrm{e}^x + 1)^2}\, dx$$

La fonction φ à intégrer est continue sur $\mathbb{R}$ et $\lim\limits_{x \to +\infty} x^2\varphi(x) = \lim\limits_{x \to -\infty} x^2\varphi(x) = 0$, la convergence en résulte en appliquant deux fois la règle de Riemann.

Le changement de variable $y = -x$, légitime, donne alors :

$$\int_{-\infty}^{+\infty} \frac{x.\mathrm{e}^x}{(\mathrm{e}^x + 1)^2}\, dx = -\int_{-\infty}^{+\infty} \frac{y.\mathrm{e}^y}{(\mathrm{e}^y + 1)^2}\, dy$$

L'intégrale est donc nulle et $E(Y) = 0$.

Exercice 3-30

On considère une urne contenant m boules numérotées de 1 à m. On procède à partir de cette urne à des tirages successifs d'une boule selon le protocole suivant :

Au premier tirage, on tire une boule « au hasard » de l'urne.

Si à un tirage quelconque, on a obtenu la boule numéro k, on la replace dans l'urne, et toutes les boules portant un numéro strictement inférieur à k sont remplacées par un nombre égal de boules portant le numéro k ; on peut alors procéder au tirage suivant.

On note X_n la variable aléatoire égale au numéro obtenu lors du n-ème tirage. $P(A)$ désigne la probabilité de l'événement A.

1. Déterminer la loi de X_1.

2. i) Montrer que : $\forall\, k \in [\![1, m]\!], P(X_2 = k) = \dfrac{2k-1}{m^2}$.

ii) Montrer que : $\forall\, n \in \mathbb{N}^*, P(X_{n+1} = m) = \dfrac{m-1}{m} P(X_n = m) + \dfrac{1}{m}$.

iii) Calculer $P(X_n = m)$ en fonction de m et de n.

3. Comparer les événements $(X_n = m)$ et $\bigcup\limits_{k=1}^{n} (X_k = m)$. En déduire un calcul direct de $P(X_n = m)$.

Solution :

1. Pour le premier tirage tous les numéros sont représentés une fois et X_1 suit la loi uniforme sur $[\![1, m]\!]$.

2. i) Appliquons la formule des probabilités totales, avec le système complet $\big(X_1 = i\big)_{1 \leq i \leq m}$:

$$\forall\, k \in [\![1, m]\!], P(X_2 = k) = \sum_{i=1}^{m} P(X_2 = k / X_1 = i) P(X_1 = i)$$

Si $i \leq k-1, P(X_2 = k \mid X_1 = i) = \dfrac{1}{m}$ (une boule porte le numéro k)

$$P(X_2 = k \mid X_1 = k) = \frac{k}{m} \text{ (}k\text{ boules portent le numéro }k\text{)}$$

si $i > k, P(X_2 = k \mid X_1 = i) = 0$ (il n'y a plus de boule portant le numéro k)

Il reste donc : $P(X_2 = k) = (k-1)\dfrac{1}{m^2} + k\,\dfrac{1}{m^2} = \dfrac{2k-1}{m^2}$

ii) De la même façon : $P(X_{n+1} = m) = \sum\limits_{i=1}^{m} P(X_{n+1} = m / X_n = i) P(X_n = i)$.

Or, si X_n prend une valeur inférieure à m, il y a 1 boule portant le numéro m pour le tirage suivant, tandis que si $X_n = m$ est réalisé, l'urne ne contient plus que des boules numérotées m. Ainsi :

$$P(X_{n+1}=m)=\frac{1}{m}\sum_{i=1}^{m-1}P(X_n=i)+P(X_n=m)$$
$$=\frac{1}{m}(1-P(X_n=m))+P(X_n=m)$$

Ce qui est le résultat voulu.

iii) La suite $\big(P(X_n = m)\big)_{n\geq 1}$ est une suite arithmético-géométrique de raison $\frac{m-1}{m}$, de point fixe 1 et de premier terme $\frac{1}{m}$. Il vient alors classiquement :

$$\forall\, n\geq 1, P(X_n=m)=1-(1-\frac{1}{m})^n$$

3. Si on obtient la boule m au cours d'un des n premiers tirages, on est sur d'obtenir la boule numérotée m au n-ème tirage. Réciproquement si on obtient la boule m au n-ème tirage, on a obtenu la boule m au cours des n premiers tirages.

Donc : $(X_n = m) = \bigcup\limits_{k=1}^{n}(X_k = m)$ et :

$$P(X_n=m)=1-P\big(\bigcap_{k=1}^{n}(X_k m)\big)$$

Mais, tant que l'on n'obtient pas la boule m, on a une chance sur m de l'obtenir au tirage suivant, c'est-à-dire la probabilité $1-\frac{1}{m}$ de ne pas l'obtenir au tirage suivant. Ainsi :

$$P(X_n=m)=1-\big(1-\frac{1}{m}\big)^n$$

Exercice 3-31

On dispose de dix pièces de monnaie numérotées de 1 à 10, telles que la k-ème pièce amène «Pile» avec la probabilité $\frac{k}{10}$.

On prend une pièce au hasard, on la lance et on obtient «Face». Quelle est la probabilité d'avoir lancé la cinquième pièce ?

Solution :

Soit A_k l'événement «on choisit la pièce numéro k» et F l'événement «on obtient Face». La formule du Révérend Thomas Bayes donne :

$$P(A_5/F)=\frac{P(F/A_5)P(A_5)}{P(F)}$$

et la formule des probabilités totales donne :

$$P(F)=\sum_{i=1}^{10}P(F/A_i)P(A_i)$$

sachant que $\sum_{i=1}^{10} i = 55$, il vient $P(A_5/F) = \frac{5}{55} = \frac{1}{11}$.

Exercice 3-32

On note $VP(\lambda, \theta)$ la loi de Pareto de paramètres $\lambda > 0, \theta > 0$ et 0 , c'est-à-dire la loi de densité f définie par : $f(x) = \dfrac{\lambda}{\theta}\Big(\dfrac{\theta}{x}\Big)^{\lambda+1}$ si $x > \theta$ et $f(x) = 0$ sinon.

Un phénomène économique suit une loi $VP(\lambda, \theta)$, θ étant un paramètre connu et λ un paramètre que l'on veut estimer. On dispose pour cela d'un échantillon $(X_1, \dots, X_n)$ de variables aléatoires indépendantes suivant cette loi.

On pose $T = \sum_{i=1}^{n} \ln\Big(\dfrac{X_i}{\theta}\Big)$ et $\widehat{\lambda} = \dfrac{n}{T}$.

1. Soit X une variable aléatoire suivant la loi $VP(\lambda, \theta)$ et $Y = \ln\Big(\dfrac{X}{\theta}\Big)$.

a) Déterminer la fonction de répartition F_X de X .

b) Déterminer, lorsqu'elles existent, l'espérance et la variance de X .

c) Montrer que Y suit une loi Γ dont on précisera les paramètres.

2. Quelle est la loi de T ? En donner une densité.

3. Calculer l'espérance et la variance de $\widehat{\lambda}$.

4. Déduire de $\widehat{\lambda}$ un estimateur $\widehat{\lambda_1}$ sans biais de λ. L'estimateur $\widehat{\lambda_1}$ est-il convergent ?

5. On admet que la suite de variables aléatoires $Z_n = \dfrac{\sqrt{n}(\widehat{\lambda} - \lambda)}{\lambda}$ converge en loi vers la loi normale centrée réduite $\mathcal{N}(0,1)$, et on rappelle que la fonction de répartition Φ de la loi normale centrée réduite vérifie $\Phi(1,96) \simeq 0{,}975$.

En utilisant la loi $\mathcal{N}(0,1)$ comme approximation de la loi de Z_n, donner, en fonction de n (supposé assez grand) et de la valeur observée λ_0 de $\widehat{\lambda}$, un intervalle de confiance à 95% de λ.

Application numérique : $n = 100$ et $\lambda_0 = 5$.

Solution :

1. Des calculs sans surprises donnent :

a) Pour $x \leq \theta$, $F_X(x) = 0$ et pour $x > \theta$, $F_X(x) = 1 - \left(\frac{\theta}{x}\right)^\lambda$

b) La variable aléatoire X admet une espérance si et seulement si $\lambda > 1$ et dans ce cas : $E(X) = \frac{\lambda\theta}{\lambda - 1}$.
De même, la variable aléatoire X admet une variance si et seulement si $\lambda > 2$ et alors : $V(X) = \frac{\lambda\theta^2}{(\lambda-2)(\lambda-1)^2}$.

c) La variable aléatoire Y prend ses valeurs dans $\mathbb{R}_+^*$ et :

$$\forall\, x > 0, F_Y(x) = P(X \leq \theta.\mathrm{e}^x) = F_X(\theta.\mathrm{e}^x) = 1 - \mathrm{e}^{-\lambda x}$$

Donc Y suit la loi exponentielle de paramètre λ, c'est-à-dire la loi Γ de paramètres $\frac{1}{\lambda}$ et 1.

2. La variable aléatoire T est la somme de n variables aléatoires indépendantes de même loi $\Gamma(\frac{1}{\lambda}, 1)$, donc T suit la loi $\Gamma(\frac{1}{\lambda}, n)$.
Une densité f_T de T est donnée par :

$$\forall\, t \leq 0, f_T(x) = 0,\ \forall\, t > 0, f_T(x) = \frac{\mathrm{e}^{-\lambda x} x^{n-1} \lambda^n}{(n-1)!}$$

3. $\star$ On a $\widehat{\lambda} = \frac{n}{T}$, donc sous réserve d'existence :

$$E(\widehat{\lambda}) = \int_{-\infty}^{+\infty} \frac{n}{x} f_T(x)\, dx = \frac{n\lambda}{n-1} \int_0^{+\infty} \frac{\mathrm{e}^{-\lambda x} x^{n-2} \lambda^{n-1}}{(n-2)!}\, dx$$

On reconnaît l'intégrale (intégrale d'une densité d'une variable aléatoire suivant la loi $\Gamma(\frac{1}{\lambda}, n-1)$), ce qui prouve que l'espérance existe et vaut :

$$E(\widehat{\lambda}) = \frac{n\lambda}{n-1}$$

$\star$ En procédant de la même façon on montre que $\widehat{\lambda}$ admet un moment d'ordre 2 valant $\frac{n^2\lambda^2}{(n-1)(n-2)}$, donc une variance valant : $V(\widehat{\lambda}) = \frac{n^2\lambda^2}{(n-1)^2(n-2)}$

4. $\widehat{\lambda}_1 = \frac{n-1}{n}\widehat{\lambda}$ vérifie $E(\widehat{\lambda}_1) = \lambda$ et $V(\widehat{\lambda}_1) = \frac{\lambda^2}{n-2}$. Cette variance est de limite nulle lorsque n tend vers l'infini, donc $\widehat{\lambda}_1$ est un estimateur sans biais et convergent de λ.

5. En approchant la loi de Z par la loi normale centrée réduite, on obtient :

$$P(-1,96 \leq \frac{\sqrt{n}(\widehat{\lambda} - \lambda)}{\lambda} \leq 1,96) = 0,95$$

C'est-à-dire :

$$P(\frac{\sqrt{n}}{\sqrt{n} + 1,96}\widehat{\lambda} \leq \lambda \leq \frac{\sqrt{n}}{\sqrt{n} - 1,96}\widehat{\lambda}) = 0,95$$

On obtient donc comme intervalle de confiance pour λ, au niveau de confiance $0,95$:

$$[\frac{\sqrt{n}}{\sqrt{n}+1,96}\lambda_0, \frac{\sqrt{n}}{\sqrt{n}-1,96}\lambda_0]$$

L'application numérique donne : $I = [4{,}18; 6{,}22]$.

Exercice 3-33

Soient X et Y deux variables aléatoires indépendantes définies sur le même espace probabilisé $(\Omega, \mathcal{B}, P)$ et suivant respectivement les lois binomiales $\mathcal{B}(n, \frac{1}{4})$ et $\mathcal{B}(n, \frac{3}{4})$.

On pose, pour $\omega \in \Omega$, $A(\omega) = \begin{pmatrix} X(\omega) & 0 \\ 2 & Y(\omega) \end{pmatrix}$.

1. Calculer la probabilité que $A(\omega)$ soit diagonalisable.

2. Calculer la probabilité que $A(\omega)$ soit inversible.

Solution :

1. Les valeurs propres de $A(\omega)$ sont $X(\omega)$ et $Y(\omega)$.

Si ces deux valeurs propres sont distinctes, $A(\omega)$ est diagonalisable.

Si elles sont égales, $A(\omega)$ n'est pas diagonalisable, sinon elle serait semblable à une matrice scalaire, donc serait une matrice scalaire, ce qui est évidemment faux.

En notant p_1 la probabilité que $A(\omega)$ soit diagonalisable, on a donc :

$$p_1 = P(XY) = 1 - P(X = Y)$$

Par indépendance de X et Y, on peut écrire :

$$P(X = Y) = \sum_{k=0}^{n} P(X = k)P(Y = k) = \left(\frac{3}{16}\right)^n \sum_{k=0}^{n} C_n^k C_n^{n-k} = \left(\frac{3}{16}\right)^n C_{2n}^n$$

[La dernière relation, appelée formule de Vandermonde résulte du fait que la somme de deux variables indépendantes suivant la loi binomiale $\mathcal{B}(n,p)$ suit la loi $\mathcal{B}(2n,p)$]

$$p_1 = 1 - \left(\frac{3}{16}\right)^n C_{2n}^n$$

2. $A(\omega)$ est inversible si et seulement si $X(\omega)0$ et $Y(\omega)0$. Donc, en notant p_2 la probabilité cherchée :

$$p_2 = P[(X0) \cap (Y0)] = [1 - P(X = 0)].[1 - P(Y = 0)]$$

$$p_2 = \left[1 - \left(\tfrac{3}{4}\right)^n\right].\left[1 - \left(\tfrac{1}{4}\right)^n\right]$$

Exercice 3-34

Soit X une variable aléatoire réelle de densité f définie sur $\mathbb{R}$ par :

$$f(x) = \begin{cases} \mathrm{e}^{-(x-a)} & \text{si } x > a \\ 0 & \text{sinon} \end{cases}$$

1. Vérifier que f est une densité de probabilité.

2. Montrer que X admet une espérance et donner sa valeur.

3. F désignant la fonction de répartition de X, déterminer m tel que $F(m) = \dfrac{1}{2}$.

4. Soit $X_1, X_2, \ldots, X_n$, n variables aléatoires indépendantes et de même loi que X. On pose $Y = \inf(X_1, \ldots, X_n)$.

 a) Déterminer la loi de probabilité de Y.

 b) Déterminer l'espérance et la variance de Y.

Solution :

1. La fonction f est continue sur $\mathbb{R}^*$, positive et on vérifie facilement que :

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_a^{+\infty} \mathrm{e}^{-(x-a)}\,dx = 1.$$

La fonction f est bien une densité de probabilité.

2. La convergence de l'intégrale $\displaystyle\int_a^{+\infty} x.\mathrm{e}^{-(x-a)}\,dx$ est connue et le calcul aisé, en intégrant par parties. On obtient : $E(X) = a + 1$.
(on peut aussi dire que $X = a + Y$, la variable aléatoire Y suivant la loi exponentielle de paramètre 1.)

3. Pour $x \geq a$, on a $F_X(x) = \displaystyle\int_a^x \mathrm{e}^{-(t-a)}\,dt = 1 - \mathrm{e}^{-(x-a)}$ et sinon $F_X(x) = 0$.

L'équation $F_X(m) = \dfrac{1}{2}$ admet donc pour unique solution : $m = a + \ln 2$.

4. a) La variable aléatoire Y prend ses valeurs dans $[a, +\infty[$ et, par indépendance :

$\forall\, x \geq a, P(Y \leq x) = 1 - P(Y > x) = 1 - [P(X > x)]^n = 1 - [1 - F_X(x)]^n$

Soit :

$$P(Y \leq x) = 1 - \mathrm{e}^{-n(x-a)}$$

b) Une densité g de Y est donc définie par : $g(x) = n.\mathrm{e}^{-n(x-a)}$ si $x \geq a$ et $g(x) = 0$ sinon. La convergence des intégrales rencontrées est évidente, par négligeabilité classique de l'exponentielle devant la puissance et, à l'aide d'intégrations par parties :

$\star$ $E(Y) = \displaystyle\int_a^{+\infty} nx.\mathrm{e}^{-n(x-a)}\,dx = a + \frac{1}{n}$.

$\star$ $E(Y^2) = \displaystyle\int_a^{+\infty} nx^2.\mathrm{e}^{-n(x-a)}\,dx = a^2 + 2a + \frac{2}{n}$, puis :

$V(Y) = 2a(1 - \frac{1}{n}) + \frac{2}{n} - \frac{1}{n^2}$.

Exercice 3-35

On considère n équipes de football de 1-ère division et n équipes de 2-ème division. On tire au sort n rencontres entre ces $2n$ équipes (on représente chaque équipe par une boule placée dans une urne et on extrait au hasard et une par une les boules de cette urne, l'équipe correspondant au $(2k+1)$-ème tirage rencontrant l'équipe correspondant au $(2k+2)$-ème tirage, pour $k \in [\![0, n-1]\!]$).

1. Calculer la probabilité p_n que tous les matchs opposent une équipe de 1-ère division à une équipe de 2-ème division.

2. Calculer la probabilité q_n que tous les matchs opposent deux équipes de la même division.

3. Montrer que $\dfrac{2^{2n-1}}{n} \leq C_{2n}^n \leq 2^{2n}$.

4. Calculer $\lim\limits_{n\to\infty} p_n$ et $\lim\limits_{n\to\infty} q_n$.

Solution :

Selon le protocole choisi, il y a $(2n)!$ façons de vider de vider l'urne, mais pour tout k de $[\![1, n]\!]$, on peut permuter les résultats des $(2k-1)$-ème et $(2k)$-ème tirages, sans que cela modifie la liste des matchs à jouer. Il y a donc en fait $\dfrac{(2n)!}{2^n}$ listes différentes de matchs, toutes équiprobables. (Attention, nous parlons bien de listes de matchs, c'est-à-dire que nous décidons qu'il y a un ordre pour les différents matchs, par exemple en décidant qu'il y a n terrains numérotés et que la k-ème paire ira jouer sur le k-ème terrain.)

1. Pour réaliser l'événement demandé, la procédure est simple : on prend toutes les équipes de division 1 et on les répartit en en plaçant une sur chaque

terrain, ce qui peut se faire de $n!$ façons, puis on procède de même pour les équipes de division 2, ce qui peut se faire aussi de $n!$ façons. On a donc :

$$p_n = \frac{2^n (n!)^2}{(2n)!} = \frac{2^n}{C_{2n}^n}$$

2. Soit B_n l'événement de l'énoncé. Si n est impair, on a $q_n = P(B_n) = 0$ (il y aura au moins une rencontre entre quipes de divisions différentes).
Si n est pair, posons $n = 2m$. On choisit alors les m terrains où s'affronteront les équipes de division 1 (les équipes de division 2 iront donc sur les autres terrains !). Le raisonnement fait dans l'introduction montre qu'il y a $\dfrac{(2m)!}{2^m}$ façons d'organiser les rencontres entre les équipes de première division et le même nombre de façons d'organiser les rencontres entre équipes de division 2. Ainsi :

$$q_{2m} = C_{2m}^m \times \frac{(2m)!}{2^m} \times \frac{(2m)!}{2^m} \times \frac{2^{2m}}{(4m)!} = \frac{C_{2m}^m}{C_{4m}^{2m}}$$

3. On a :

$$C_{2n}^n = \frac{2n(2n-1)(2n-2)\cdots 1}{n!n!} = \frac{2^n(2n-1)(2n-3)\cdots 5.3.1}{n!}$$

Ainsi :

$$\frac{2^n(2n-2)(2n-4)\cdots 4.2}{n!} \leq C_{2n}^n \leq \frac{2^n(2n)(2n-2)\cdots 4.2.1}{n!}$$

Soit :

$$\frac{2^{2n-1}}{n} \leq C_{2n}^n \leq 2^{2n}$$

4. $\star$ On a donc : $0 \leq p_n \leq \dfrac{n}{2^{n-1}}$ et $\lim\limits_{n\to\infty} p_n = 0$. Puis :

$\star$ $0 \leq q_{2m} \leq \dfrac{2^{2m}}{C_{4m}^{2m}} = p_{2m}$ et $\lim\limits_{m\to\infty} q_{2m} = 0$. D'où $\lim\limits_{n\to\infty} q_n = 0$.