1. Déterminer la fonction de répartition et la fonction quantile pour les lois suivantes :
    1. La loi Béta $B(\alpha, 1)$, avec $\alpha > 0$, dont la densité est
    $$f(x) = \alpha x^{\alpha-1}\mathbf{1}_{[0,1]}(x).$$
    2. La loi de Weibull de paramètres $\alpha > 0$ et $\lambda > 0$, dont la densité est
    $$f(x) = \alpha\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}}\mathbf{1}_{\mathbb{R}^+}(x).$$

---

2. Montrer que si X est une variable aléatoire réelle, les médianes de X minimisent $a \mapsto \mathbb{E}|X - a|$. Commencer par le montrer lorsque X est une variable aléatoire discrète, uniforme sur $\{x_1, \dots, x_r\}$ avec $x_1 < \dots < x_r$.

---

3. Soit X une variable aléatoire uniforme sur l'intervalle $[1, 3]$.
Calculer la loi de la variable aléatoire $Y = \frac{1}{2}(|X - 1| + |X|)$. Calculer la moyenne et la variance de Y.

---

4. Soit X une variable aléatoire de loi uniforme sur $[1, 5]$. Déterminer la fonction de répartition et la densité de probabilité de la variable aléatoire $Y \coloneqq X/(5 - X)$.

Solution: Tout d'abord, $F_Y(\frac{1}{4}) = \mathbb{P}(Y \leqslant \frac{1}{4}) = \mathbb{P}(X \leqslant 1) = 0$, ce qui montre que $F_Y(y) = 0$ pour tout $y \leqslant \frac{1}{4}$. Supposons donc $y > \frac{1}{4}$. Dans ce cas,
$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(X \leqslant y(5 - X)) = \mathbb{P}(X \leqslant 5y/(1 + y)) = \frac{4y - 1}{4y + 4},$$
puisque $5y/(1 + y) \leqslant 5$ pour tout $y > \frac{1}{4}$. La densité de probabilité s'obtient par différentiation :
$$f_Y(y) = F'_Y(y) = \frac{5}{4(y + 1)^2}\mathbf{1}_{\{y \geqslant \frac{1}{4}\}}.$$

---

5. Soit X une variable aléatoire réelle de fonction de répartition F_X. On définit, pour tout $u \in]0, 1[$ l'inverse généralisée continue à gauche de F_X :
$$F_X^{-1}(u) = \inf\{t : F_X(t) \geqslant u\}$$
    a) Montrer que, pour tout $t \in \mathbb{R}$, $u \in]0, 1[$, $(F_X^{-1}(u) \leqslant t) \Leftrightarrow (u \leqslant F_X(t))$
    b) En déduire que si U suit la loi uniforme sur $[0, 1]$ alors $\tilde{X} = F_X^{-1}(U)$ suit la même loi que X. Remarquer que cela permet de simuler des lois non uniformes.
    c) Expliciter F_X^{-1} dans les cas d'une loi exponentielle et d'une loi de Cauchy.
    d) A l'aide de ce qui précède, montrer que toute fonction croissante F continue à droite telle que $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$ est la fonction de répartition d'une loi de probabilité sur $\mathbb{R}$.

Solution:

---

6. Soit $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes de même loi de fonction de répartition F. Soit $X_{(1)}, \dots, X_{(n)}$ les nombres $X_1, \dots, X_n$ ordonnés de façon croissante. La variable aléatoire $X_{(k)}$ est appelée statistique d'ordre de rang k et fait partie des outils fondamentaux en statistiques.
   a) Trouver une expression (sous forme d'une somme appropriée) pour la fonction de répartition de $X_{(k)}$ pour $k \in \{1, \dots, n\}$.

   Solution: On a
   $$F_{X_{(k)}}(t) = \mathbb{P}(X_{(k)} \leqslant t) = \sum_{\ell=k}^{n} \binom{n}{\ell} F(t)^{\ell}(1 - F(t))^{n-\ell}.$$
   En effet, lorsque $X_{(k)} \leqslant t$, $\ell \in \{k, \dots, n\}$ variables prennent une valeur inférieure ou égale à t, ce qui se produit avec probabilité $F(t)$ pour chacune, alors que les $n - \ell$ variables restantes prennent une valeur strictement supérieure à t, ce qui se produit avec probabilité $1 - F(t)$ pour chacune ; évidemment, il y a $\binom{n}{\ell}$ façons de choisir ces ℓ variables.

   b) Déterminer explicitement les fonctions de répartition de $X_{(1)}$ et $X_{(n)}$.

   Solution: On a évidemment $F_{X_{(n)}}(t) = F(t)^n$ et, à l'aide de la formule du binôme de Newton,
   $$F_{X_{(1)}}(t) = \sum_{\ell=1}^{n} \binom{n}{\ell} F(t)^{\ell}(1 - F(t))^{n-\ell} = 1 - (1 - F(t))^n.$$

---

7. a) Soit $X \sim \mathsf{poisson}(\lambda)$ et $Y \sim \mathsf{poisson}(\mu)$ deux variables aléatoires indépendantes. Montrer que $X + Y \sim \mathsf{poisson}(\lambda + \mu)$.

   Solution: Pour tout $k \in \mathbb{N}$,
   $$\begin{aligned}\mathbb{P}(X + Y = k) &= \sum_{\ell=0}^{k} \mathbb{P}(X = \ell, Y = k - \ell) = \sum_{\ell=0}^{k} \mathbb{P}(X = \ell)\mathbb{P}(Y = k - \ell) \\ &= \sum_{\ell=0}^{k} \frac{\lambda^{\ell}}{\ell!}e^{-\lambda} \frac{\mu^{k-\ell}}{(k-\ell)!}e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{\ell=0}^{k} \binom{k}{\ell} \lambda^{\ell}\mu^{k-\ell} = \frac{(\lambda+\mu)^k}{k!}e^{-(\lambda+\mu)}.\end{aligned}$$

   b) Soit $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant chacune une loi de Poisson de paramètre λ. Montrer que $X_1 + \dots + X_n \sim \mathsf{poisson}(n\lambda)$.

   Solution: On procède par récurrence. Le cas $n = 1$ est trivial. Supposons que $X_1 + \dots + X_{n-1} \sim \mathsf{poisson}((n-1)\lambda))$. Comme $X_1 + \dots + X_{n-1}$ est indépendant de X_n, il suit du point précédent que $(X_1 + \dots + X_{n-1}) + X_n$ suit une loi de Poisson de paramètre $(n-1)\lambda + \lambda = n\lambda$.

   c) Notons X_k, $1 \leqslant k \leqslant n$ le nombre de fautes de frappe sur la k-ième page d'un texte de n pages. On supposera que les variables aléatoires X_k sont indépendantes et suivent chacune une loi de Poisson de paramètre λ.
      i. Calculer la probabilité que la première faute se trouve sur la k-ième page.

Solution: On doit avoir $X_1 = \cdots = X_{k-1} = 0$ et $X_k \geqslant 1$. Par indépendance,

$$\mathbb{P}(X_1 = \cdots = X_{k-1} = 0, X_k \geqslant 1) = \mathbb{P}(X_k \geqslant 1) \prod_{i=1}^{k-1} \mathbb{P}(X_i = 0) = (1 - e^{-\lambda}) e^{-\lambda(k-1)}.$$

ii. Pour $1 \leqslant r \leqslant n$, notons $S_r \coloneqq X_1 + \cdots + X_r$ le nombre total de fautes dans les r premières pages. Soit $\ell \geqslant 0$; calculer $\mathbb{P}(S_r \geqslant \ell)$.

Solution: Puisque S_r suit une loi de Poisson de paramètre $r\lambda$ par le point b de l'exercice, on a

$$\mathbb{P}(S_r \geqslant \ell) = 1 - \mathbb{P}(S_r \leqslant \ell - 1) = 1 - e^{-r\lambda} \sum_{i=0}^{\ell-1} \frac{(r\lambda)^i}{i!}.$$

iii. Soit $\ell \geqslant 1$. Calculer la probabilité que la ℓ-ième faute se trouve sur la k-ième page.

Solution: Comme $\{S_{k-1} \geqslant \ell\} \subset \{S_k \geqslant \ell\}$, on peut écrire $\{S_k \geqslant \ell\} \cap \{S_{k-1} < \ell\} = \{S_k \geqslant \ell\} \setminus \{S_{k-1} \geqslant \ell\}$. Par conséquent,

$$\mathbb{P}(S_k \geqslant \ell, S_{k-1} < \ell) = \mathbb{P}(S_k \geqslant \ell) - \mathbb{P}(S_{k-1} \geqslant \ell).$$

On peut donc conclure avec le point précédent.

---

8. Soit $p_1, \dots, p_n \in (0,1)$ et soit $X_1, \dots, X_n$ des variables aléatoires indépendantes avec $X_i \sim \mathsf{géom}(p_i)$.

a) Déterminer $\mathbb{P}(X_i \geqslant k)$ pour tout $k \in \mathbb{N}^*$.

Solution: On peut soit procéder par un calcul direct,

$$\mathbb{P}(X_i \geqslant k) = p_i \sum_{\ell \geqslant k} (1 - p_i)^{\ell - 1} = p_i (1 - p_i)^{k-1} \sum_{\ell \geqslant 0} (1 - p_i)^{\ell} = (1 - p_i)^{k-1},$$

soit observer que cette probabilité est la probabilité que $k - 1$ épreuves de Bernoulli de paramètre p_i résultent toutes en un échec.

b) Soit $M \coloneqq \min\{X_1, \dots, X_n\}$. Déterminer $\mathbb{P}(M \geqslant k)$ pour tout $k \in \mathbb{N}^*$.

Solution:

$$\mathbb{P}(M \geqslant k) = \mathbb{P}(X_i \geqslant k,\ 1 \leqslant i \leqslant n) = \prod_{i=1}^{n} \mathbb{P}(X_i \geqslant k) = \Big(\prod_{i=1}^{n} (1 - p_i) \Big)^{k-1}.$$

c) Déduire du point précédent que $M \sim \mathsf{géom}(\rho)$ et exprimer ρ en fonction de $p_1, \dots, p_n$.

Solution: Posons $1 - \rho = \prod_{i_1}^{n} (1 - p_i)$. On a vu que $\mathbb{P}(M \geqslant k) = (1 - \rho)^{k-1}$. On a donc

$$\mathbb{P}(M = k) = \mathbb{P}(M \geqslant k) - \mathbb{P}(M \geqslant k + 1) = (1 - \rho)^{k-1} - (1 - \rho)^{k} = (1 - \rho)^{k-1} \rho.$$

d) En utilisant le fait que X_i a la même loi que le temps de premier succès d'une suite d'épreuves de Bernoulli de paramètre p_i, réinterpréter M comme le temps de premier succès d'une suite d'épreuves de Bernoulli appropriées et en déduire une preuve alternative du résultat du point précédent.

Tournez la page.

Solution: Considérons n suites d'épreuves de Bernoulli en parallèle, la i-ième ayant une probabilité de succès p_i. On peut alors interpréter M comme le nombre d'épreuves nécessaires pour observer un succès dans au moins une de ces n suites d'épreuves de Bernoulli. La probabilité d'un échec est donc donnée par la probabilité d'avoir un échec dans chacune des n épreuves, soit $\prod_{i=1}^n (1-p_i)$. On en conclut donc bien que M suit une loi géométrique dont la probabilité de succès est $1-\prod_{i=1}^n (1-p_i)$.

---

9. Soit X et Y deux variables aléatoires indépendantes suivant chacune la loi géométrique de paramètre $p \in (0,1)$. Soit $U \coloneqq \max\{X, Y\}$ et $V \coloneqq \min\{X, Y\}$.
   a) Pour $u, v \in \mathbb{N}^*$, calculer $\mathbb{P}(U \leqslant u, V \geqslant v)$. En déduire les lois de U et de V.

Solution:

$$\mathbb{P}(U \leqslant u, V \geqslant v) = \mathbb{P}(v \leqslant X \leqslant u)^2 = \big(\mathbb{P}(X > v-1) - \mathbb{P}(X > u)\big)^2 = \big((1-p)^{v-1} - (1-p)^u\big)^2.$$

En particulier,

$$\begin{aligned}\mathbb{P}(U \leqslant u) &= \mathbb{P}(U \leqslant u, V \geqslant 1) = \big(1-(1-p)^u\big)^2,\\ \mathbb{P}(V \geqslant v) &= \lim_{u \to +\infty} \mathbb{P}(U \leqslant u, V \geqslant 1) = (1-p)^{2v-2}.\end{aligned}$$

On trouve donc, pour tout $u \in \mathbb{N}^*$,

$$\mathbb{P}(U = u) = \mathbb{P}(U \leqslant u) - \mathbb{P}(U \leqslant u-1) = (1-(1-p)^u)^2 - (1-(1-p)^{u-1})^2,$$

et, pour tout $v \in \mathbb{N}^*$,

$$\mathbb{P}(V = v) = \mathbb{P}(V \geqslant v) - \mathbb{P}(V \geqslant v+1) = (1-p)^{2v-2} - (1-p)^{2v} = (1-(1-p)^2)(1-p)^{2(v-1)}.$$

   b) Déterminer la loi de $D \coloneqq |X - Y|$ et montrer que D et V sont indépendantes.

Solution: Observons que $D = U - V$. En particulier,

$$\forall n, m \in \mathbb{N}^*, \qquad \mathbb{P}(V = m, D = n) = 2\mathbb{P}(X = m, Y = m+n) = 2p^2(1-p)^{2m+n-2}.$$

En particulier, on a

$$\mathbb{P}(D = n) = \sum_{m \geqslant 1} \mathbb{P}(V = m, D = n) = 2p^2(1-p)^n \sum_{m \geqslant 1} (1-p)^{2m-2} = \frac{2p^2(1-p)^n}{1-(1-p)^2}.$$

On a donc

$$\forall n, m \in \mathbb{N}^*, \qquad \mathbb{P}(V = m, D = n) = \mathbb{P}(V = m)\mathbb{P}(D = n).$$

L'indépendance suit donc, puisque

$$\begin{aligned}\forall v \in \mathbb{N}^*, \forall d \in \mathbb{N}, \qquad F_{V,D}(v,d) &= \sum_{m=1}^{v} \sum_{n=0}^{d} \mathbb{P}(V = m, D = n)\\ &= \sum_{m=1}^{v} \sum_{n=0}^{d} \mathbb{P}(V = m)\mathbb{P}(D = n) = F_V(v)F_D(d),\end{aligned}$$

---

10. Albert collectionne des vignettes, vendues par pochette. Il existe n vignettes différentes, et chaque pochette contient 5 vignettes distinctes, tirées uniformément parmi toutes les combinaisons possibles. Albert achète une pochette à la fois jusqu'à ce qu'il ait complété sa collection. On note N le nombre de pochettes qu'il lui a fallu acheter. On numérote les N pochettes selon leur ordre d'achat.

    a) On note T_k le numéro de la première pochette contenant une copie de la k-ième vignette. Montrer que $N = \max\{T_1, \dots, T_n\}$.

    Solution: Par définition, si $m \geqslant T_k$, Albert possède la k-ième vignette après avoir acheté m pochettes. Par conséquent, il possède toutes les vignettes après avoir acheté m pochettes si $m \geqslant \max\{T_1, \dots, T_n\}$. D'autre part, si $m < \max\{T_1, \dots, T_n\}$, il existe k tel que $m < T_k$ et Albert ne possède donc pas encore la vignette k après avoir acheté m pochettes.

    b) On note $A_k \coloneqq \{T_k > \ell\}$. Montrer que

$$\mathbb{P}(N > \ell) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n} \mathbb{P}\big(\min\{T_{i_1}, \dots, T_{i_k}\} > \ell\big).$$

    (Indication : appliquer le principe d'inclusion-exclusion à $\mathbb{P}(A_1 \cup \dots \cup A_n)$.)

    Solution: Il suffit d'observer que, pour tout $1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n$,

$$A_{i_1} \cap \dots \cap A_{i_k} = \big\{T_{i_1} > \ell, \dots, T_{i_k} > \ell\big\} = \big\{\min\{T_{i_1}, \dots, T_{i_k}\} > \ell\big\}$$

    et que

$$\bigcup_{i=1}^{n} A_i = \big\{\exists k \in \{1, \dots, n\} \,\big|\, T_k > \ell\big\} = \big\{N > \ell\big\}.$$

    c) Calculer $\mathbb{P}\big(\min(T_{i_1}, \dots, T_{i_k}) > \ell\big)$ pour $1 \leqslant k \leqslant n$ et $1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n$.

    Solution: Pour que $\min(T_{i_1}, \dots, T_{i_k}) > \ell$, il est nécessaire et suffisant qu'aucune des vignettes $i_1, \dots, i_k$ ne se trouve dans les ℓ premières pochettes achetées. Pour chaque pochette achetée, la probabilité de n'y trouver aucune de ces vignettes est égale à $\binom{n-k}{5}/\binom{n}{5}$. Par conséquent, les contenus des pochettes étant indépendants,

$$\mathbb{P}\big(\min(T_{i_1}, \dots, T_{i_k}) > \ell\big) = \left(\frac{\binom{n-k}{5}}{\binom{n}{5}}\right)^{\ell}.$$

    d) En déduire que le nombre moyen de pochettes qu'Albert va devoir acheter pour compléter sa collection est égal à

$$\mathbb{E}[N] = \sum_{k=1}^{n} (-1)^{k-1} \frac{\binom{n}{k}\binom{n}{5}}{\binom{n}{5} - \binom{n-k}{5}}.$$

Tournez la page.

Solution: On a, par le Lemme 2.57,

$$\begin{aligned}
\mathbb{E}[N] &= \sum_{\ell \geqslant 0} \mathbb{P}(N > \ell) \\
&= \sum_{\ell \geqslant 0} \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \left(\frac{\binom{n-k}{5}}{\binom{n}{5}} \right)^{\ell} \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \frac{\binom{n}{5}}{\binom{n}{5} - \binom{n-k}{5}} \\
&= \sum_{k=1}^{n} (-1)^{k-1} \frac{\binom{n}{k}\binom{n}{5}}{\binom{n}{5} - \binom{n-k}{5}}.
\end{aligned}$$

e) Supposons que $n = 660$. Calculer explicitement le nombre moyen de pochettes à acheter.

Solution: La formule précédente implique qu'il faut en moyenne acheter 931 pochettes (soit 4655 vignettes) pour remplir l'album ! On comprend mieux pourquoi Panini est une compagnie prospère !

---

11. Une marque de céréales place une figurine à collectionner dans chacun de ses paquets. Une collection complète comporte n figurines distinctes. Albert souhaite obtenir la collection complète. Il va donc acheter un paquet à la fois, jusqu'à ce qu'il ait obtenu au moins un exemplaire de chacune des n figurines, puis il s'arrêtera. Soit T_n le nombre de paquets qu'il devra acheter. Le but de cet exercice est de démontrer qu'Albert devra acheter approximativement $n \log n$ paquets (lorsque n est grand) :

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}\Big(\Big| \frac{T_n}{n \log n} - 1 \Big| \geqslant \epsilon \Big) = 0. \tag{$\star$}$$

On considérera qu'à chaque achat, la probabilité d'obtenir chacune des n figurines est égale à $1/n$.

a) Soit τ_n^k le nombre minimum de paquets qu'Albert doit acheter afin d'avoir en sa possession k ($k \leqslant n$) figurines distinctes. Évidemment, $\tau_n^0 = 0$ et $\tau_n^n = T_n$. Exprimer T_n en termes des variables aléatoires $\xi_k := \tau_n^k - \tau_n^{k-1}$ (le nombre de paquets à acheter pour faire passer sa collection de $k-1$ figurines à k figurines).

Solution: Clairement, $T_n = \tau_n^n = (\tau_n^n - \tau_{n-1}^n) + (\tau_{n-1}^n - \tau_{n-2}^n) + \cdots + (\tau_n^1 - \tau_n^0) = \sum_{k=1}^n \xi_k$.

b) Expliquer pourquoi les variables aléatoires $\xi_1, \xi_2, \ldots, \xi_n$ sont indépendantes.

Solution: Notons X_k la variable aléatoire donnant le numéro de la vignette obtenue lors du k-ième tirage. Alors, les $(X_k)_{k\geqslant 1}$ sont i.i.d., de loi uniforme sur $\{1, \ldots, n\}$. Notons $\mathcal{A}_{k-1}$ le sous-ensemble (aléatoire) de $\{1, \ldots, n\}$ correspondant aux numéros des $k-1$ figurines obtenues au temps τ_n^{k-1}. On a alors, pour tout $1 \leqslant k \leqslant n$,

$$\begin{aligned}
\mathbb{P}(\xi_k = \ell_k \,|\, \xi_1 = \ell_1, \ldots, \xi_{k-1} = \ell_{k-1}, \mathcal{A}_{k-1} = A) &= \mathbb{P}(X_{r+1} \notin A, \ldots, X_{r+\ell_k-1} \notin A, X_{r+\ell_k} \in A) \\
&= (1 - p_k)^{\ell_k - 1} p_k,
\end{aligned}$$

où l'on a posé $r := \ell_1 + \cdots + \ell_{k-1}$ et $p_k := (n-k+1)/n$. Il suffit en effet d'observer que l'événement $\{X_{r+1} \notin A, \ldots, X_{r+\ell_k-1} \notin A, X_{r+\ell_k} \in A\}$ est indépendant de l'événement $\{\xi_1 = \ell_1, \ldots, \xi_{k-1} = \ell_{k-1}, \mathcal{A}_{k-1} = A\}$, ce dernier ne dépendant que des variables aléatoires $X_1, \ldots, X_r$.

On a donc, en particulier,

$$\begin{aligned}
\mathbb{P}(\xi_k = \ell_k) &= \sum_{\ell_1,\dots,\ell_{k-1},A} \mathbb{P}(\xi_k = \ell_k \mid \xi_1 = \ell_1, \dots, \xi_{k-1} = \ell_{k-1}, \mathcal{A}_{k-1} = A) \\
&\qquad\qquad\qquad \times \mathbb{P}(\xi_1 = \ell_1, \dots, \xi_{k-1} = \ell_{k-1}, \mathcal{A}_{k-1} = A) \\
&= (1-p_k)^{\ell_k - 1} p_k \sum_{\ell_1,\dots,\ell_{k-1},A} \mathbb{P}(\xi_1 = \ell_1, \dots, \xi_{k-1} = \ell_{k-1}, \mathcal{A}_{k-1} = A) \\
&= (1-p_k)^{\ell_k - 1} p_k .
\end{aligned}$$

Donc,

$$\begin{aligned}
\mathbb{P}(\xi_1 = \ell_1, \dots, \xi_n = \ell_n) &= \sum_{A_{n-1}} \mathbb{P}(\xi_n = \ell_n \mid \xi_1 = \ell_1, \dots, \xi_{n-1} = \ell_{n-1}, \mathcal{A}_{n-1} = A_{n-1}) \\
&\qquad\qquad\qquad \times \mathbb{P}(\xi_1 = \ell_1, \dots, \xi_{n-1} = \ell_{n-1}, \mathcal{A}_{n-1} = A_{n-1}) \\
&= \mathbb{P}(\xi_n = \ell_n) \sum_{A_{n-1}} \mathbb{P}(\xi_1 = \ell_1, \dots, \xi_{n-1} = \ell_{n-1}, \mathcal{A}_{n-1} = A_{n-1}) \\
&= \mathbb{P}(\xi_n = \ell_n) \mathbb{P}(\xi_1 = \ell_1, \dots, \xi_{n-1} = \ell_{n-1}) \\
&= \cdots \\
&= \mathbb{P}(\xi_n = \ell_n) \mathbb{P}(\xi_{n-1} = \ell_{n-1}) \cdots \mathbb{P}(\xi_1 = \ell_1),
\end{aligned}$$

ce qui établit l'indépendance de $\xi_1, \dots, \xi_n$.

Remarque : *ci-dessus, on a utilisé la convention habituelle d'écrire* $\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(B \mid C_i)\mathbb{P}(C_i)$ *pour une partition* $(C_i)_{i \in I}$ *de* Ω*, même dans le cas où* $\mathbb{P}(C_i) = 0$ *pour certains* $i \in I$*. Cela ne pose pas de problème, car le produit* $\mathbb{P}(B \mid C_i)\mathbb{P}(C_i) = \mathbb{P}(B \cap C_i) = 0$ *reste bien défini. Ceci nous a évité de devoir spécifier explicitement les valeurs de* $\ell_1, \ell_2, \dots$ *pour lesquelles les probabilités sont strictement positives.*

c) Quelle est la loi de ξ_k ? Déterminer son espérance et sa variance.

Solution: Comme on l'a vu ci-dessus, $\xi_k \sim \mathsf{géom}(p_k)$ avec $p_k = (n-k+1)/n$. En particulier,

$$\mathbb{E}[\xi_k] = \frac{1}{p_k} = \frac{n}{n-k+1}\,, \qquad \mathrm{Var}[\xi_k] = \frac{1-p_k}{p_k^2} = \frac{n(k-1)}{(n-k+1)^2}\,.$$

d) En déduire que

$$\mathbb{E}[T_n] = n \sum_{k=1}^{n} k^{-1}\,, \qquad \mathrm{Var}[T_n] \leqslant n^2 \sum_{k=1}^{n} k^{-2},$$

et donc, en particulier, que $\lim_{n\to\infty} \mathbb{E}[T_n]/(n \log n) = 1$.

Solution: Par linéarité de l'espérance,

$$\mathbb{E}[T_n] = \sum_{k=1}^{n} \mathbb{E}[\xi_k] = n \sum_{k=1}^{n} (n-k+1)^{-1} = n \sum_{k=1}^{n} k^{-1}\,.$$

Comme

$$\sum_{k=1}^{n} k^{-1} \leqslant 1 + \int_1^n x^{-1}\,\mathrm{d}x = 1 + \log n$$

et

$$\sum_{k=1}^{n} k^{-1} \geqslant \int_1^{n+1} x^{-1}\,\mathrm{d}x = \log(n+1) = \log(1 + \frac{1}{n}) + \log n,$$

on a bien $\lim_{n\to\infty} \mathbb{E}[T_n]/(n\log n) = 1$. Finalement, par indépendance des ξ_k,

$$\mathrm{Var}[T_n] = \sum_{k=1}^{n} \mathrm{Var}[\xi_k] = n\sum_{k=1}^{n} \frac{(k-1)}{(n-k+1)^2} = n\sum_{k=1}^{n} \frac{n-k}{k^2} \leqslant n^2 \sum_{k=1}^{n} k^{-2}\,.$$

e) Soient $(S_k)_{k\geqslant 1}$ des variables aléatoires telles que $\mathbb{E}[S_n] = \mu_n$ et $\mathrm{Var}[S_n] = \sigma_n^2$. Soit $(b_n)_{n\geqslant 1}$ une suite de nombres réels tels que $\lim_{n\to\infty} \sigma_n^2/b_n^2 = 0$. Démontrer, à l'aide de l'inégalité de Bienaymé–Tchebychev, que

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}(|S_n - \mu_n| > \epsilon b_n) = 0.$$

Solution: La conclusion suit immédiatement de Bienaymé-Tchebychev :

$$\mathbb{P}(|S_n - \mu_n| > \epsilon b_n) \leqslant \frac{\sigma_n^2}{\epsilon^2 b_n^2} \xrightarrow{n\to\infty} 0.$$

f) Vérifier qu'on peut appliquer le point précédent avec $S_n = T_n$ et $b_n = n\log n$. En déduire ($\star$).

Solution: Par le point précédent, pour tout $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(\big|T_n - \mathbb{E}[T_n]\big| \geqslant \tfrac{1}{2}\epsilon n\log n\big) = 0\,,$$

puisque $\mathrm{Var}[T_n] \leqslant \frac{\pi^2}{6}n^2$ et donc $\mathrm{Var}[T_n]/(n\log n)^2 \leqslant 1/(\log n)^2 \to 0$.
Comme $\mathbb{E}[T_n]/(n\log n) \to 1$, on a $\{|T_n - n\log n| \geqslant \epsilon n\log n\} \subset \{|T_n - \mathbb{E}[T_n]| \geqslant \frac{1}{2}\epsilon n\log n\}$ pour tout n suffisamment grand, et ($\star$) suit.

---

12. (Urne de Pólya) On considère une urne contenant initialement b boules bleues et r boules rouges. On tire une boule au hasard uniformément et on la replace dans l'urne, ainsi que d boules supplémentaires de la même couleur que cette dernière. On répète ensuite cette procédure autant de fois que nécessaire. On note $X_k = 1$ si la k-ième boule tirée est bleue et $X_k = 0$ sinon. Soit $S_n \coloneqq X_1 + \cdots + X_n$ le nombre total de boule bleues extraites lors des n premiers tirages.

a) Calculer la loi de X_1 et la loi de X_2.

Solution: Manifestement, $\mathbb{P}(X_1 = 1) = b/(b+r)$. Par la formule de la probabilité totale,

$$\begin{aligned}\mathbb{P}(X_2 = 1) &= \mathbb{P}(X_2 = 1 \,|\, X_1 = 1)\mathbb{P}(X_1 = 1) + \mathbb{P}(X_2 = 1 \,|\, X_1 = 0)\mathbb{P}(X_1 = 0)\\ &= \frac{b+d}{b+r+d}\cdot\frac{b}{b+r} + \frac{b}{b+r+d}\cdot\frac{r}{b+r} = \frac{b}{b+r}.\end{aligned}$$

Ainsi, X_1 et X_2 suivent toutes deux une loi de Bernoulli de paramètre $b/(b+r)$.

b) Calculer $\mathbb{P}(X_{n+1} = 1 \,|\, S_n = k)$.

Solution: Lorsque $S_n = k$, l'urne contient, après n tirages, $b + r + nd$ boules, dont $b + kd$ sont bleues. On a donc

$$\mathbb{P}(X_{n+1} = 1 \,|\, S_n = k) = \frac{b+kd}{b+r+nd}.$$

c) En déduire une expression pour $\mathbb{P}(X_{n+1} = 1)$ en termes de $\mathbb{E}[S_n]$.

Solution: Par la formule de la probabilité totale,

$$\begin{aligned}\mathbb{P}(X_{n+1}=1) &= \sum_{k=0}^{n} \mathbb{P}(X_{n+1}=1 \mid S_n=k)\,\mathbb{P}(S_n=k) = \sum_{k=0}^{n} \frac{b+kd}{b+r+nd}\mathbb{P}(S_n=k) \\ &= \frac{b}{b+r+nd}\sum_{k=0}^{n}\mathbb{P}(S_n=k) + \frac{d}{b+r+nd}\sum_{k=0}^{n} k\mathbb{P}(S_n=k) \\ &= \frac{b}{b+r+nd} + \frac{d}{b+r+nd}\mathbb{E}[S_n].\end{aligned}$$

d) Exprimer $\mathbb{E}[S_n]$ en termes de $\mathbb{E}[S_{n-1}]$ et $\mathbb{P}(X_n=1)$.

Solution:

$$\mathbb{E}[S_n] = \mathbb{E}[S_{n-1}+X_n] = \mathbb{E}[S_{n-1}] + \mathbb{P}(X_n=1).$$

e) En déduire que $\mathbb{P}(X_{n+1}=1) = \mathbb{P}(X_n=1)$.

Solution: Par les deux points précédents,

$$\begin{aligned}\mathbb{P}(X_{n+1}=1) &= \frac{b}{b+r+nd} + \frac{d}{b+r+nd}\mathbb{E}[S_n] \\ &= \frac{b}{b+r+nd} + \frac{d}{b+r+nd}\mathbb{E}[S_{n-1}] + \frac{d}{b+r+nd}\mathbb{P}(X_n=1) \\ &= \frac{b+r+(n-1)d}{b+r+nd}\mathbb{P}(X_n=1) + \frac{d}{b+r+nd}\mathbb{P}(X_n=1) \\ &= \mathbb{P}(X_n=1).\end{aligned}$$

f) Quelle est la loi de X_n ?

Solution: Par le point précédent, $X_n \overset{\text{loi}}{=} X_1$. Par conséquent, $X_n \sim \text{bernoulli}(b/(b+r))$.

---

13. Soit $r : \mathbb{N} \to \mathbb{R}$ telle que $\lim_{\ell\to\infty} r(\ell) = 0$. Soit $(X_n)_{n\geqslant 1}$ des variables aléatoires réelles satisfaisant, pour tout $n \geqslant 1$, $\mathbb{E}[X_n] = \mu$ et $\mathrm{Cov}[X_n, X_{n+\ell}] \leqslant r(\ell)$ pour tout $\ell \geqslant 0$. Soit $\overline{X}_n$ la moyenne empirique de $X_1, \ldots, X_n$. Montrer que

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| \geqslant \epsilon) = 0.$$

(Remarque : Cet exercice montre que la loi faible des grands nombres s'étend à certaines suites de variables aléatoires corrélées.)

Solution: Soit $\delta > 0$ et L tel que $\forall \ell \geqslant L$, $r(\ell) \leqslant \delta$.

$$\mathrm{Var}[\overline{X}_n] = \frac{1}{n^2}\sum_{1\leqslant i,j\leqslant n}\mathrm{Cov}[X_i,X_j] \leqslant \frac{2}{n^2}\sum_{i=1}^{n}\sum_{\ell=0}^{n-i} r(\ell) \leqslant \frac{2}{n^2}\sum_{i=1}^{n}\sum_{\ell=0}^{L-1} r(\ell) + \frac{2}{n^2}\sum_{i=1}^{n}\sum_{\ell=L}^{n}\delta \leqslant \frac{CL}{n} + 2\delta,$$

où $C \coloneqq 2\max\{r(\ell) \mid 0 \leqslant \ell < L\}$. Par conséquent, il suit de l'inégalité de Bienaymé-Tchebychev que

$$\mathbb{P}(|\overline{X}_n - \mu| \geqslant \epsilon) \leqslant \frac{\mathrm{Var}[\overline{X}_n]}{\epsilon^2} \leqslant \frac{CL}{\epsilon^2 n} + \frac{2\delta}{\epsilon^2}.$$

Donc,

$$\forall \epsilon > 0, \forall \delta > 0, \qquad \limsup_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| \geqslant \epsilon) \leqslant \frac{2\delta}{\epsilon^2},$$

Tournez la page.

ce qui donne

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| \geqslant \epsilon) = 0.$$

---

14. Soit $(X_n)_n$ une suite de variables aléatoires telle que chaque X_n suit une loi uniforme sur l'ensemble $\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$. Montrer que cette suite converge en loi vers une variable aléatoire X de loi uniforme sur l'intervalle $[0, 1]$.

   Solution:

---

15. Soit $\theta > 0$ et soit $(X_n)_n$ une suite de variables aléatoires telle que chaque X_n suit une loi géométrique de paramètre $p_n = \frac{\theta}{n}$. Montrer que la suite T_n/n converge en loi et déterminer sa limite.

   Solution:

---

16. Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires de Bernoulli de paramètre p_i, indépendantes.
    1. Si $p_i \to p$, montrer que la suite X_i converge en loi vers une variable aléatoire de Bernoulli de paramètre p.
    2. Si $\sum_i p_i < \infty$, montrer que X_i converge presque sûrement vers 0.
    3. Si $p_i \to 0$ et $\sum_i p_i = \infty$, montrer que X_i diverge p.s. mais converge en loi vers une variable constante égale à 0.