Sommes de variables indépendantes S6 Polytech

1. **Couple de variables de Bernoulli.** Soit X et Y deux variables aléatoires de Bernoulli, dont la loi jointe est donnée par le tableau suivant :

$X \backslash Y$	0	1
0	$\frac{1}{2} - a$	$a + \frac{1}{3}$
1	b	$\frac{1}{6} - 2a$

   (a) Quelles valeurs sont autorisées pour a et b ?
   (b) Calculer en fonction de a et b les lois marginales de X et de Y.
   (c) Quelles valeurs de a et b correspondent à un couple (X, Y) de variables indépendantes ?

2. **Un couple de variables à densité.** Soit (X, Y) un couple de variables aléatoires à densité, avec

$$f_{(X,Y)}(x, y) = \begin{cases} C\, y\, \mathrm{e}^{-x} & \text{si } 0 \leq y \leq x, \\ 0 & \text{sinon.} \end{cases}$$

   (a) Trouver la valeur de la constante C.
   (b) Calculer les densités marginales de X et de Y. Quelles lois retrouve-t-on ?
   (c) Calculer la covariance de X et Y. Les variables X et Y sont-elles indépendantes ?

3. **Lois conditionnelles.** Soit X et Y deux variables aléatoires à valeurs dans $\mathbb{Z}$. La *loi conditionnelle* de X sachant Y est la fonction aléatoire $\mathbb{P}_{X|Y} : \mathbb{Z} \to [0, 1]$, dont les valeurs sont données par

$$\mathbb{P}_{X|Y}[k] = \mathbb{P}[X = k | Y = l] = \frac{\mathbb{P}[X = k \text{ et } Y = l]}{\mathbb{P}[Y = l]} \quad \text{si } Y = l.$$

   (a) Expliciter $\mathbb{P}_{X|Y}$ dans le cas où X et Y sont indépendantes.
   (b) Soit W et X deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres w et x. On pose $Y = W + X$, et on rappelle que Y suit une loi de Poisson de paramètre $w + x$. Calculer la loi conditionnelle $\mathbb{P}_{X|Y}$.

4. **Somme de deux loi géométriques.** Calculer la loi de la somme de deux variables aléatoires indépendantes X et Y de lois géométriques $\mathcal{G}(p)$ et $\mathcal{G}(p')$ (on distinguera les cas $p = p'$ et $p \neq p'$).

5. **Somme de deux lois gaussiennes.** Calculer la loi de la somme de deux variables aléatoires indépendantes X et Y, avec $X \sim \mathcal{N}(m_1, \sigma_1^2)$ et $Y \sim \mathcal{N}(m_2, \sigma_2^2)$. On pourra se ramener au cas où $m_1 = m_2 = 0$.

6. **Loi faible des grands nombres.** Dans cet exercice, $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes, de même loi et centrées : $\mathbb{E}[X_n] = 0$ pour tout n. On suppose aussi $\mathbb{E}[(X_n)^2] = \mathbb{E}[(X_1)^2] = \sigma^2 \in (0, +\infty)$: la variance est finie et non nulle.

(a) On pose $M_n = \frac{X_1+X_2+\cdots+X_n}{n}$. Calculer $\mathbb{E}[(M_n)^2]$.

(b) Montrer que

$$\mathbb{P}[|M_n| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2},$$

et en déduire que cette probabilité tend vers 0 pour tout $\varepsilon > 0$.

(c) On suppose plus généralement que les variables X_n sont indépendantes centrées, et que $\mathbb{E}[(X_n)^2] = (\sigma_n)^2$, les valeurs $(\sigma_n)^2$ des variances pouvant maintenant dépendre de n. Donner une condition simple sur ces variances pour que $\lim_{n\to\infty} \mathbb{P}[|M_n| \geq \varepsilon] = 0$ quelque soit la valeur de $\varepsilon > 0$.

7. **Cumulants d'une variable aléatoire.** Soit X une variable aléatoire dont les valeurs possibles restent bornées dans un intervalle $[-A, A]$. On admet que dans ce cas, on peut développer en série

$$\mathbb{E}[\mathrm{e}^{zX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(zX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!}\, z^k,$$

cette expansion étant valable pour tout $z \in \mathbb{C}$. On pose $h_X(z) = \log \mathbb{E}[\mathrm{e}^{zX}]$.

(a) Montrer que

$$h'_X(0) = \mathbb{E}[X] \qquad ; \qquad h''_X(0) = \mathrm{var}(X) \qquad ; \qquad h''''_X(0) = \kappa^{(4)}(X).$$

(b) En déduire une preuve simple des identités $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$ et $\kappa^{(4)}(X+Y) = \kappa^{(4)}(X) + \kappa^{(4)}(Y)$ lorsque X et Y sont deux variables indépendantes.

8. **Moyenne de variables aléatoires indépendantes mais pas de même loi.** Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, telles que les lois des variables X_n soient discrètes et données par les formules suivantes :

$$\mathbb{P}[X_n = 0] = 1 - \frac{1}{n\log(n+1)};$$
$$\mathbb{P}[X_n = n] = \mathbb{P}[X_n = -n] = \frac{1}{2n\log(n+1)}$$

pour tout $n \in \mathbb{N}^*$.

(a) On pose $M_n = \frac{X_1+\cdots+X_n}{n}$. Calculer $\mathbb{E}[X_n]$, $\mathrm{var}(X_n)$, $\mathbb{E}[M_n]$ et $\mathrm{var}(M_n)$.

(b) Montrer que $\mathbb{P}[|M_n| \geq \varepsilon] \to 0$ pour tout $\varepsilon > 0$.

(c) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements indépendants :

$$\mathbb{P}[A_{n_1} \cap A_{n_2} \cap \cdots \cap A_{n_k}] = \prod_{i=1}^{k} \mathbb{P}[A_{n_i}]$$

pour tout choix d'entiers $n_1 < n_2 < \cdots < n_k$. On s'intéresse à l'événement $B_{n_0} = \bigcup_{n\geq n_0} A_n$. Montrer que

$$\mathbb{P}[B_{n_0}] = 1 - \prod_{n=n_0}^{\infty}(1 - \mathbb{P}[A_n]) \geq 1 - \exp\left(-\sum_{n=n_0}^{\infty} \mathbb{P}[A_n]\right).$$

En déduire la probabilité de B_{n_0} si $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = +\infty$.

(d) Montrer que pour tout n_0 et tout $\varepsilon < 1$, $\mathbb{P}[\forall n \geq n_0,\ |M_n| \leq \varepsilon] = 0$. Commenter.

9. **Méthode de Monte Carlo.** Soit f une fonction continue sur $[0,1]$. On se donne une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ indépendantes et de loi uniforme sur $[0,1]$. Expliquer comment calculer de façon approchée l'intégrale $\int_0^1 f(x)\,dx$, en utilisant uniquement les nombres $f(X_1), f(X_2), \ldots, f(X_n)$.

10. **Limites de variables de Poisson.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires de lois de Poisson $\mathcal{P}(\lambda_n)$, avec $\lambda_n \to +\infty$ et les λ_n entiers.

    (a) Écrire X_n comme somme de λ_n variables aléatoires indépendantes et de même loi.

    (b) Quelle est la limite en loi des variables Y_n, où
$$Y_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}\ ?$$

    (c) Traiter aussi le cas où les paramètres λ_n ne sont plus supposés entiers, mais tendent toujours vers $+\infty$.

11. **Convergence vers une loi de Poisson.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires de lois $X_n \sim \mathcal{B}(n, p_n)$. On rappelle le critère sur les transformées de Fourier pour la convergence en loi : X_n converge en loi vers une variable X si et seulement si $\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X_n}] \to \mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X}]$ pour tout $\xi \in \mathbb{R}$.

    (a) Calculer $\mathbb{E}[\mathrm{e}^{\mathrm{i}\xi X_n}]$.

    (b) On suppose que $p_n = \frac{\lambda}{n}$. Montrer que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers une variable $X \sim \mathcal{P}(\lambda)$.

    (c) Étendre le résultat précédent au cas où $X_n = \sum_{i=1}^n A_{n,i}$, avec les $A_{n,i} \sim \mathcal{B}(p_{n,i})$ variables de Bernoulli indépendantes. On demande dans ce cas des conditions suffisantes simples sur les paramètres $(p_{n,i})_{n\geq i\geq 1}$ pour avoir la convergence vers une loi de Poisson $\mathcal{P}(\lambda)$.

12. **Distance entre deux probabilités discrètes.** Si X et Y sont deux variables discrètes à valeurs dans $\mathbb{Z}$, on définit leur *distance en variation totale* par la formule suivante :
$$d(X,Y) = \frac{1}{2}\sum_{k\in\mathbb{Z}} |\mathbb{P}_X[k] - \mathbb{P}_Y[k]|\,.$$

    (a) Montrer qu'on a toujours $d(X,Y) \leq 1$.

    (b) Si $X \sim \mathcal{B}(p)$ et $Y \sim \mathcal{P}(p)$, montrer que $d(X,Y) = p(1-\mathrm{e}^{-p})$, puis que $d(X,Y) \leq p^2$.

    (c) Soient $(X_n)_{n\in\mathbb{N}}$ et X des variables à valeurs dans $\mathbb{Z}$. On suppose que $(X_n)_{n\in\mathbb{N}}$ converge en loi vers X. Montrer que $d(X_n, X)$ tend vers 0. La réciproque est elle vraie ?

    (d) Soient X_1 et X_2 deux variables indépendantes, et Y_1 et Y_2 deux variables indépendantes. Montrer que
$$d(X_1 + X_2, Y_1 + Y_2) \leq d(X_1, X_2) + d(Y_1, Y_2).$$

(e) Application : soit $(X_n)_{n\in\mathbb{N}}$ une famille de variables indépendantes de lois de Bernoulli de paramètres $(p_n)_{n\in\mathbb{N}}$, et $(Y_n)_{n\in\mathbb{N}}$ une famille de variables indépendantes de lois de Poisson de mêmes paramètres $(p_n)_{n\in\mathbb{N}}$. Montrer que

$$d(X_1 + X_2 + \cdots + X_n, Y_1 + Y_2 + \cdots + Y_n) \leq \sum_{i=1}^{n} (p_i)^2.$$

Retrouver le résultat de l'exercice précédent (cette fois-ci sans utiliser le critère de Fourier).

13. **Approximation de fonctions par des polynômes.** Soit f une fonction continue sur $[0,1]$. On définit une suite de polynômes

$$P_n^f(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k \,(1-x)^{n-k}.$$

(a) Soit X_n une variable aléatoire de loi $\mathcal{B}(n,p)$. Calculer $\mathbb{E}[f(\frac{X_n}{n})]$.

(b) Quelle est la limite de $\mathbb{E}[f(\frac{X_n}{n})]$? On pourra décomposer l'espérance en deux parties :

$$\mathbb{E}\left[1_{|\frac{X_n}{n}-p|\leq\varepsilon}\, f\left(\frac{X_n}{n}\right)\right] + \mathbb{E}\left[1_{|\frac{X_n}{n}-p|>\varepsilon}\, f\left(\frac{X_n}{n}\right)\right]$$

et utiliser la loi des grands nombres.

(c) Montrer que pour tout $x \in [0,1]$, $\lim_{n\to\infty} P_n^f(x) = f(x)$. Commenter.