SY01 Médian–A11
Durée : 2h
UNE FEUILLE RECTO-VERSO AUTORISÉE
CALCULATRICE INTERDITE

Le barème tiendra compte de la rédaction. Justifiez clairement vos réponses.

Exercice 1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité sur lequel sont définies toutes les variables aléatoires mentionées dans cet exercice.
Soit f la fonction définie par

$$f(x) = C \exp(-x) 1\!\!1_{]0,+\infty[}(x),$$

où C est une constante et où $1\!\!1_{]0,+\infty[}(x) = \begin{cases} 1 & \text{si} \quad x > 0 \\ 0 & \text{si} \quad x \leq 0 \end{cases}$ désigne la fonction indicatrice sur $]0, +\infty[$.

1. Déterminer la valeur de C pour que f soit une densité de probabilité.
2. Soit X une variable aléatoire continue admettant f comme densité de probabilité. Déterminer l'espérance et la variance de X.
3. Déterminer la fonction de répartition de X.
4. Soit Y la variable aléatoire définie par $Y = \exp(X)$. Après avoir exprimé la fonction de répartition de Y en fonction de celle de X, déterminer la fonction de répartition de Y.
5. La variable aléatoire Y admet-elle une densité de probabilité ? Si tel est le cas, déterminer cette fonction.

Correction–6 points

1. f est une densité de probabilité si C est un réel positif et si $\int_{\mathbb{R}} f(t)dt = 1$, soit

$$\begin{aligned} \int_{\mathbb{R}} f(t)dt = 1 &\iff C \int_{\mathbb{R}} \exp(-t) 1\!\!1_{]0,+\infty[}(t)dt \\ &\iff C \left[-\exp(t)\right]_0^{+\infty} = 1 \\ &\iff C \exp(0) = 1 \\ &\iff C = 1. \end{aligned}$$

2. Si elle existe, l'espérance de X est égale à

$$\begin{aligned} \int_{\mathbb{R}} t \exp(-t) 1\!\!1_{]0,+\infty[}(t)dt &= \left[-t \exp(-t)\right]_0^{+\infty} + \int_0^{\infty} \exp(-t) 1\!\!1_{]0,+\infty[}(t)dt \\ &= 1, \end{aligned}$$

qui s'obtient en effectuant une intégration par partie (IPP). Si elle existe, la variance de X est égale à $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$, il reste donc à déterminer $\mathbb{E}(X^2)$ si elle existe. A partir d'une IPP , il vient,

$$\begin{aligned} \int_{\mathbb{R}} t^2 f(t)dt &= \left[-t^2 \exp(-t)\right]_0^{+\infty} + 2 \int_0^{\infty} t \exp(-t)dt \\ &= 2\mathbb{E}(X) = 2. \end{aligned}$$

Nous en déduisons que $\mathbb{E}(X) = 1$ et $\text{Var}(X) = 1$.

3. Soit F la fonction de répartition de X. Pour tout x réel, $F(x) = \mathbb{P}(X \leq x)$ et

$$F(x) = \int_{-\infty}^{x} \exp(-t) 1\!\!1_{]0,+\infty[}(t)dt = \begin{cases} 0 & si\ x \leq 0, \\ \int_0^x \exp(-t)dt = 1 - \exp(-x) & si\ x > 0. \end{cases}$$

4. Notons G la fonction de répartition de Y; G est définie sur $\mathbb{R}$ par :
$\forall y \in \mathbb{R}, G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\exp(X) \leq y) = \begin{cases} 0 & \text{si } y \leq 0 \\ F(\log(y)) & \text{si } y > 0. \end{cases} \begin{cases} 0 & \text{si } y \leq 1 \\ 1 - \frac{1}{y} & \text{si } y > 1. \end{cases}$

5. La fonction G est continue et de fonction dérivée égale à $G'(y) = \begin{cases} 0 & \text{si } y \leq 1 \\ \frac{1}{y^2} & \text{si } y > 1, \end{cases}$ qui est continue sauf en $y = 1$; par conséquent, Y admet une densité de probabilité $g(y) = G'(y)$ pour tout y dans $\mathbb{R}$.

<u>Exercice 2.</u> Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité sur lequel sont définies toutes les variables aléatoires mentionées dans cet exercice.

- **A.** Soient Y et Z deux variables aléatoires de même loi, indépendantes et de loi commune, la loi Géométrique $\mathcal{G}(p)$, avec $p \in]0, 1[$.
  - En utilisant la convolution, déterminer la loi de la variable aléatoire $S = Z + Y$.
- **B.** Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires de même loi, indépendantes et de loi commune, la loi de Bernoulli de paramètre p, $p \in]0, 1[$. La suite $(X_n)_{n\in\mathbb{N}^\star}$ est appelée une suite d'épreuves de Bernoulli de paramètre p.
  Pour $r \in \mathbb{N}^\star$, notons T_r la variable aléatoire égale au nombre minimal d'épreuves de Bernoulli nécessaires à l'obtention d'exactement r succès et notons W_r la variable aléatoire égale au nombre d'échecs qui précédent le r-ième succès.

  1. Considérons dans cette question $r = 2$. Après avoir donné l'ensemble E_{T_2} des valeurs possibles de T_2, déterminer pour tout $k \in E_{T_2}$, $\mathbb{P}(T_2 = k)$. Comparer la loi de T_2 avec la loi de S de la question **A**.
  2. Considérons le cas général de r quelconque, $r \in \mathbb{N}^\star$. Montrer que pour un k bien choisi,

$$\begin{aligned} \{T_r = k\} &= \{\sum_{l=1}^{k-1} X_l = r - 1\} \cap \{X_k = 1\} \\ &= \{X_1 + \ldots + X_{k-1} = r - 1\} \cap \{X_k = 1\} \qquad (1) \end{aligned}$$

  3. A partir de la relation (1), montrer pour un k bien choisi que

$$\mathbb{P}(T_r = k) = C_{k-1}^{r-1} p^r (1-p)^{k-r},$$

     où C_{k-1}^{r-1} désigne les combinaisons de $(r-1)$ éléments dans un ensemble en contenant $(k-1)$.
  4. Etablir une relation simple entre T_r et W_r; en déduire l'ensemble E_{W_r} des valeurs possibles de W_r, puis pour tout $k \in E_{W_r}$, déterminer $\mathbb{P}(W_r = k)$.

Correction–9 points + 0.5 BONUS

- **A.** Les variables aléatoires Y et Z sont indépendantes; par conséquent la loi de S est le produit de convolution des lois de Y et Z; comme Y et Z suivent la même loi, la loi Géométrique $\mathcal{G}(p)$, la variable aléatoire S est à valeurs dans $\mathbb{N}^\star \setminus \{1\}$. En notant p_Y et p_Z les lois de probabilités de Y et Z, rappelons que $p_Y(l) = p_Z(l) = p(1-p)^{l-1}$, $\forall l \in \mathbb{N}^\star$ (donc $p_Y(0) = p_Z(0) = 0$, c'est ce qui justifie les bornes de la somme ci-dessous); nous obtenons alors pour tout $k \in \mathbb{N}^\star \setminus \{1\}$,

$$\begin{aligned} \mathbb{P}(S=k) &= \sum_{l=1}^{k-1} p_Y(k-l)p_Z(l) \\ &= \sum_{l=1}^{k-1} p(1-p)^{k-l-1}p(1-p)^{l-1} \\ &= p^2(1-p)^{k-2}\sum_{l=1}^{k-1} 1 \\ &= p^2(1-p)^{k-2}(k-1). \end{aligned}$$

- **B.**

1. Le nombre minimum d'expériences de Bernoulli permettant d'obtenir deux succés est 2; cette valeur est la plus plus petite valeur que peut prendre la variable aléatoire T_2; l'ensemble des valeurs possibles de T_2 est égal à $E_{T_2} = \mathbb{N}^\star \setminus \{1\}$. Pour tout $k \in E_{T_2}$, l'événement $\{T_2 = k\}$ signifie que la k-ième expérience est un succès et qu'au cours des $(k-1)$ premières expériences, se produisent 1 succès et $(k-2)$ échecs. Nous obtenons donc $\forall k \in E_{T_2}$,

$$\mathbb{P}(T_2 = k) = C_{k-1}^1(1-p)^{k-2}p^2 = (k-1)(1-p)^{k-2}p^2.$$

La loi de T_2 est donc identique à celle de la variable aléatoire S de la partie **A**.

2. Par un raisonnement similaire à la question **B.**1., l'événement $\{T_r = k\}$ signifie qu'au cours des $(k-1)$ premières expériences de Bernoulli, on a $(r-1)$ succès et donc $(k-r)$ échecs, c'est à dire $\sum_{l=1}^{k-1} X_l = r-1$ et que la k-ième expérience de Bernoulli est un succès, c'est à dire $X_k = 1$. Pour que cet événement soit non vide, k doit être supérieur ou égal à r. On retrouve ainsi la relation (1) vérifiée pour $k \geq r$, k entier.

3. La suite des $(X_l)_{l\geq 1}$ étant une suite de variables aléatoires indépendantes, les événements $\{\sum_{l=1}^{k-1} X_l = r-1\}$ et $\{X_k = 1\}$ sont indépendants (en tant que fonctions de variables indépendantes). Nous avons vu en cours et en TD que la somme de $(k-1)$ v.a. indépendantes de Bernoulli $B(p)$ suit la loi Binomiale $B(k-1,p)$; d'après la relation (1) et par indépendance des deux événements de part et d'autre du signe $\cap$, nous en déduisons que pour k entier et $k \geq r$,

$$\begin{aligned} \mathbb{P}(T_r = k) &= \mathbb{P}(\sum_{l=1}^{k-1} X_l = r-1)\mathbb{P}(X_k = 1) \\ &= C_{k-1}^{r-1}p^r(1-p)^{k-r}. \end{aligned}$$

4. De part les définitions de T_r et W_r, il facile d'établir que $T_r = r + W_r$; par ailleurs comme T_r peut prendre la valeur r et que W_r ne peut pas prendre de valeurs négatives, alors la

valeur minimale possible de W_r est 0; de part l'ensemble des valeurs possibles de T_r, nous en déduisons que $E_{W_r} = \mathbb{N}$. Pour tout $k \in E_{W_r}$,

$$\begin{aligned}\mathbb{P}(W_r = k) &= \mathbb{P}(T_r - r = k) \\ &= \mathbb{P}(T_r = k + r) \\ &= C_{k+r-1}^{r-1} p^r (1-p)^k.\end{aligned}$$

.

Exercice 3. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité sur lequel sont définies toutes les variables aléatoires mentionées dans cet exercice.
Soient N une variable aléatoire à valeurs dans $\mathbb{N}^\star$ et $(Y_l)_{l\in\mathbb{N}^\star}$ une suite de variables aléatoires discrètes de même loi, à valeurs dans $\mathbb{N}$, (mutuellement) indépendantes et de loi commune $(p_k)_{k\in\mathbb{N}}$ telle que pour tout k dans $\mathbb{N}$, $p_k = \mathbb{P}(Y_l = k), \forall l \in \mathbb{N}^\star$. Nous supposons de plus que N et la suite $(Y_l)_{l\in\mathbb{N}^\star}$ sont indépendantes.
Considérons la fonction définie par :

$$\forall \omega \in \Omega,\ Z_N(\omega) = \sum_{l=1}^{N(\omega)} Y_l(\omega).$$

- **A. Cas général.**

  1. Montrer que la fonction $Z_N : \omega \in \Omega \longrightarrow Z_N(\omega)$ est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$.
  2. Notons G_{Z_N} la fonction génératrice de Z_N, h_N la fonction génératrice de N et g la fonction génératrice de Y_l pour tout l dans $\mathbb{N}^\star$.
     - (i) Montrer que pour tout $k \in \mathbb{N}$, $\mathbb{P}(Z_N = k) = \sum_{n\geq 1} \mathbb{P}(Z_n = k \cap N = n)$
     - (ii) En déduire la relation suivante: $G_{Z_N}(s) = h_N(g(s)) = h_N \circ g(s)$.
  3. En supposant dérivables en $s = 1$, les fonctions génératrices G_{Z_N}, h_N et g, déduire de la question 2.(ii) la relation suivante: $\mathbb{E}(Z_N) = \mathbb{E}(N)\mathbb{E}(Y_l)$ où l est un entier naturel non nul quelconque.

- **B. Application.** Le nombre d'accidents en une semaine sur la nationale N31, est une variable aléatoire d'espérance $\mu > 0$ et de variance $\sigma^2 > 0$. Le nombre d'individus blessés à la suite d'accident est une variable aléatoire d'espérance $\nu > 0$ et de variance $\tau^2 > 0$. Le nombre d'individus dans des accidents différents sont indépendants entre eux et indépendants du nombre d'accidents.
  Supposons que les fonctions génératrices des variables aléatoires, nombre d'accidents en une semaine sur la nationale N31 et nombre de blessés par accident soient deux fois dérivables en 1.

  1. En utilisant la partie **A.**, déterminer l'espérance et la variance du nombre d'individus blessés au cours des accidents survenus sur la nationale N31 pendant une semaine donnée.
  2. Utiliser l'inégalité de Bienaymé-Tchebychev pour déterminer un réel $a > \mu\nu$ en fonction de μ, ν, σ^2 et τ^2 tel que la probabilité que le nombre d'individus blessés au cours des accidents survenus sur la nationale N31 pendant une semaine donnée soit supérieur à a, est inférieure à 0.1.

Correction —6 points + 1 point BONUS

- **A. Cas général.**

1. Il est clair que l'ensemble des valeurs possibles de Z_N est $\mathbb{N}$; par conséquent, montrer que Z_N est une variable aléatoire définie sur $(\Omega, \mathcal{A})$ revient à montrer que pour tout $k \in \mathbb{N}$, l'ensemble $\{\omega \in \Omega : Z_N(\omega) = k\}$ appartient à $\mathcal{A}$. Remarquons que $\Omega = \cup_{n\geq 1}\{\omega \in \Omega : N(\omega) = n\}$ et $\{\omega \in \Omega : N(\omega) = n\} \cap \{\omega \in \Omega : N(\omega) = m\} = \emptyset$ pour tous les couples $(n, m) :\ n \neq m$ d'entiers strictement positifs, c'est à dire $\{\omega \in \Omega : N(\omega) = n\}_{n\geq 1}$ est une partition de Ω.

$$\begin{aligned}
\{\omega \in \Omega : Z_N(\omega) = k\} &= \{\omega \in \Omega : Z_N(\omega) = k\} \\
&= \{\omega \in \Omega : \sum_{l=1}^{N(\omega)} Y_l(\omega) = k\} \\
&= \{\omega \in \Omega : \sum_{l=1}^{N(\omega)} Y_l(\omega) = k\} \cap (\cup_{n\geq 1}\{\omega \in \Omega : N(\omega) = n\}) \\
&= \cup_{n\geq 1}(\{\omega \in \Omega : \sum_{l=1}^{N(\omega)} Y_l(\omega) = k\} \cap \{\omega \in \Omega : N(\omega) = n\}) \\
&= \cup_{n\geq 1}(\{\omega \in \Omega : \sum_{l=1}^{n} Y_l(\omega) = k\} \cap \{\omega \in \Omega : N(\omega) = n\}), \quad (2)
\end{aligned}$$

où (2) est vraie car l'intersection des deux événements est non vide uniquement pour les $\omega \in \Omega$ tels que $N(\omega) = n$. Finalement, le terme de droite dans l'équation (2) appartient à $\mathcal{A}$ en tant que réunion dénombrable d'éléments de $\mathcal{A}$ puisqu'ils correspondent à une intersection d'élements de $\mathcal{A}$ (par hypothèse, les variables aléatoires N et $(Y_l)_{l\geq 1}$ sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$).

2. – (i) A partir de la relation (2) et par σ-additivité de $\mathbb{P}$, nous avons

$$\mathbb{P}(Z_N = k) = \sum_{n\geq 1} \mathbb{P}(Z_n = k \cap N = n) \quad (3)$$

– (ii) Par définition de la fonction génératrice, $G_{Z_N}(s) = \mathbb{E}(s^{Z_N})$. Les variables aléatoires N et $(Y_l)_{l\geq 1}$ sont indépendantes, par conséquent $\mathbb{P}(Z_n = k \cap N = n) = \mathbb{P}(Z_n = n)\mathbb{P}(N = n) = \mathbb{P}(\sum_{l=1}^{n} Y_l(\omega) = k)\mathbb{P}(N = n)$. De plus, comme Z_n est une somme de n variables aléatoires indépendantes et de même loi ($Z_n = \sum_{l=1}^{n} Y_l(\omega)$), alors $\mathbb{E}(s^{Z_n}) = \prod_{l=1}^{n} \mathbb{E}(s^{Y_l}) = (g(s))^n$. Signalons enfin que lorsque les quantités sont positives, il est permis d'échanger les signes $\sum$, c'est à dire $\sum_{n\geq 1}\sum_{k\geq 0} u_{n,k} = \sum_{k\geq 0}\sum_{n\geq 1} u_{n,k}$, avec $u_{n,k} \geq 0$ pour tout k et n.
Alors à partir de la relation (3) , nous avons pour tout k dans $\mathbb{N}$,

$$\begin{aligned}
G_{Z_N}(s) &= \sum_{k\geq 0} s^k \mathbb{P}(Z_N = k) \\
&= \sum_{k\geq 0} s^k \sum_{n\geq 1} \mathbb{P}(Z_n = k \cap N = n) \\
&= \sum_{n\geq 1}\sum_{k\geq 0} s^k \mathbb{P}(Z_n = k)\mathbb{P}(N = n) \\
&= \sum_{n\geq 1} \mathbb{P}(N = n) \sum_{k\geq 0} s^k \mathbb{P}(Z_n = k)
\end{aligned}$$

$$\begin{aligned} &= \sum_{n\geq 1}\mathbb{P}(N=n)(g(s))^n \\ &= \sum_{n\geq 1}(g(s))^n\mathbb{P}(N=n) \\ &= \mathbb{E}(g(s)^N) = h_N(g(s)). \end{aligned}$$

3. D'après le cours, $\mathbb{E}(Z_N) = G'_{Z_N}(1)$. Nous dérivons en s l'expression de la fonction G_{Z_N} de la questions 2 ii),

$$G'_{Z_N}(s) = h'_N(g(s))g'(s).$$

Remarquons que $g(1) = \mathbb{E}(1^{Y_1}) = \sum_k 1^k p_k = \sum_k p_k = 1$, par conséquent, $h'_N(g(s)) = h'_N(1) = \mathbb{E}(N)$. Nous obtenons donc $\mathbb{E}(Z_N) = \mathbb{E}(N)\mathbb{E}(Y_l)$.

- **B. Application.**

Soient N le nombre d'accidents sur la nationale N31 au cours d'un semaine, Y le nombre de blessés au cours d'un accident et $Z_N = \sum_{l=1}^N Y_l$ le nombre d'individus blessés au cours des accidents survenus sur la nationale N31 pendant une semaine donnée.

1. En utilisant directement la partie **A.**, $\mathbb{E}(Z_N) = \mathbb{E}(N)\mathbb{E}(Y_l) = \mu\nu$. Nous avons $G''_{Z_N}(s) = h''_N(g(s))(g'(s))^2 + h'_N(g(s))g''(s)$. Remarquons que $h''_N(1) = \sigma^2 + \mu^2 - \mu$ et $g''(1) = \tau^2 + \nu^2 - \nu$ puisque d'après le cours, $h''_N(1) = \mathrm{Var}(N) + (\mathbb{E}(N))^2 - \mathbb{E}(N)$ et $g''(1) = \mathrm{Var}(Y) + (\mathbb{E}(Y_l))^2 - \mathbb{E}(Y_l)$ alors pour $s = 1$,

$$\begin{aligned} G''_{Z_N}(1) &= h''_N(g(1))(g'(1))^2 + h'_N(g(1))g''(1) \\ &= h''_N(1)(g'(1))^2 + h'_N(1)g''(1) \\ &= (\sigma^2+\mu^2-\mu)\nu^2 + \mu(\tau^2+\nu^2-\nu) \\ &= (\sigma^2+\mu^2)\nu^2 + \mu(\tau^2-\nu). \end{aligned} \tag{4}$$

la variance de Z_N se déduit de la relation (4) :
$\mathrm{Var}(Z_N) = (\sigma^2+\mu^2)\nu^2 + \mu(\tau^2-\nu) + \mu\nu(1-\mu\nu) = \sigma^2\nu^2 + \mu\tau^2$.

2. D'après la question **B**1., l'espérance et la variance de Z_N existent. Rappelons que pour tout variable aléatoire C et tout réel c, l'implication suivant est vraie : "$C > c \Rightarrow |C| > c$", puisque $|C| > C$; par conséquent, on a l'inclusion des événements $\{C > c\} \subset \{|C| > c\}$, ce qui entraine une relation d'ordre sur les probabilités correspondantes, soit $\mathbb{P}(C > c) \leq \mathbb{P}(|C| > c)$.

En appliquant l'inégalité de Bienaymé-Tchebychev, nous cherchons à déterminer un réel $a > \mu\nu$ tel que

$$\begin{aligned} \mathbb{P}(Z_N > a) &= \mathbb{P}(Z_N - \mu\nu > a - \mu\nu) \\ &\leq \mathbb{P}(|Z_N - \mu\nu| > a - \mu\nu) \\ &\leq \frac{\mathrm{Var}(Z_N)}{(a-\mu\nu)^2} \\ &\leq 0.1. \end{aligned} \tag{5}$$

De (5), nous en déduisons un réel a qui vérifie l'inégalité suivante :

$$\begin{aligned} &\frac{\sigma^2\nu^2 + \mu\tau^2}{(a-\mu\nu)^2} \leq 0.1 \\ \Longleftrightarrow\ & 10 * (\sigma^2\nu^2 + \mu\tau^2) \leq (a-\mu\nu)^2 \\ \Longleftrightarrow\ & \mu\nu + \sqrt{10 * (\sigma^2\nu^2 + \mu\tau^2)} \leq a. \end{aligned}$$