Leçon 21 Exercices corrigés

(*Une étoile * désignera une question de difficulté supérieure.*)

Exercice 1. Soit $(X_1, \dots, X_n)$ un échantillon de variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes de même loi exponentielle $\mathcal{E}(\frac{1}{\xi})$ de paramètre $\frac{1}{\xi} > 0$. Pour tout $n \geq 1$, poser comme dans la leçon, $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

a) Rappeler la moyenne et la variance de X_1.

b) Démontrer que pour tout $t > 0$,

$$\mathbb{P}\big(|\overline{X}_n - \xi| \geq t\big) \leq \frac{\xi^2}{t^2 n}.$$

c) Calculer un intervalle de confiance au seuil $\mathrm{s_e} = 1 - \kappa$ de ξ pour l'estimateur de la moyenne lorsque $\kappa n > 1$.

d) Mêmes questions si les variables $X_1, \dots, X_n$ sont indépendantes de même loi uniforme $\mathcal{U}(0, a)$ sur $[0, a]$, où $a > 0$ est un paramètre inconnu à estimer.

Indication. Sur la base de l'inégalité de Tchebychev b), choisir $t > 0$ tel que $\frac{\xi^2}{t^2 n} = \kappa$, autrement dit $t = \frac{\xi}{\sqrt{\kappa n}}$. Dire que $|\overline{X}_n - \xi| < t$ revient alors à

$$\frac{\overline{X}_n}{1 + \frac{1}{\sqrt{\kappa n}}} < \xi < \frac{\overline{X}_n}{1 - \frac{1}{\sqrt{\kappa n}}}$$

fournissant l'intervalle de confiance demandé au seuil $1 - \kappa$.

Exercice 2. Un vol Toulouse-Paris est assuré par un avion de $N = 150$ places. Pour un vol de ce type, des estimations ont montré que la probabilité pour qu'une personne confirme son billet est $p = 0,75$. La compagnie vend n billets, $n > 150$. Soit X la variable aléatoire « nombre de personnes parmi les n possibles ayant confirmé leur réservation pour ce vol ».

a) Quelle est la loi exacte suivie par X ?

b) Quel est le nombre maximum de places que la compagnie peut vendre pour que, à au moins 95%, elle soit sûre que tout le monde puisse monter dans l'avion (c'est-à-dire trouver n tel que $\mathbb{P}(X > 150) \leq 0,05$) ? (*Indication* : utiliser l'inégalité de Tchebychev ; proposer une estimation plus précise à partir du théorème central limite.)

c) Discuter suivant les valeurs de N, p et n.

Corrigé. a) La loi de X est une loi binomiale $\mathcal{B}(n,p)$ de moyenne np et de variance $np(1-p)$. b) D'après l'inégalité de Tchebychev,

$$\mathbb{P}(X > 150) = \mathbb{P}\big(X - np > 150 - np\big) \leq \frac{np(1-p)}{(150-np)^2}$$

(sous réserve que $150 - np > 0$). Avec la valeur $p = 0,75$, la fraction $\frac{np(1-p)}{(150-np)^2}$ est plus petite que $0,05$ dès que $n \leq 166$. Ainsi, en vendant moins de 166 billets, la compagnie ne prend qu'un risque inférieur à 5% de devoir indemniser des voyageurs en surnombre. D'après le théorème central limite, la probabilité

$$\mathbb{P}(X > 150) = \mathbb{P}\left(\frac{X-np}{\sqrt{np(1-p)}} > \frac{150-np}{\sqrt{np(1-p)}}\right)$$

est « proche » de $\mathbb{P}\big(G > \frac{150-np}{\sqrt{np(1-p)}}\big)$ où G suit une loi $\mathcal{N}(0,1)$. Une table de loi normale indique que $\mathbb{P}(G > t)$ est de l'ordre de $0,05$ si t est de l'ordre de $1,645$. Il faut alors résoudre $\frac{150-np}{\sqrt{np(1-p)}} \geq 1,645$ pour que la probabilité précédente soit inférieure à $0,05$. Il ressort que $n \leq 187$ (ce qui est plus souple que la restriction fournie par l'inégalité de Tchebychev, mais qui demande en revanche à préciser le mot « proche », source de tous les risques !). c) En faisant varier les paramètres précédents (dans l'approximation par le théorème central limite) :

$N = 150$, $p = 0,5$, $n \leq 272$.

$N = 300$, $p = 0,75$, $n \leq 381$.

$N = 300$, $p = 0,5$, $n \leq 561$.

Exercice 3 (*Processus de Poisson*). Soit X_n, $n \in \mathbb{N}$, une suite de variables aléatoires indépendantes de même loi exponentielle $\mathcal{E}(\alpha)$ de paramètre $\alpha > 0$. Poser $S_n = X_1 + \cdots + X_n$, $n \geq 1$, et par convention $S_0 = 0$. Pour tout $t > 0$, soit $N_t = \sup\{n \geq 0\,; S_n < t\}$.

a) Quelle est la loi de S_n ?

b) Établir que, pour tout $n \in \mathbb{N}$,

$$\mathbb{P}(N_t = n) \ = \ \mathbb{P}(S_n < t) - \mathbb{P}(S_{n+1} < t).$$

En déduire la loi de N_t.

Indication. a) Comme somme de variables aléatoires indépendantes de même loi exponentielle $\mathcal{E}(\alpha)$, la loi de S_n est une loi Gamma $\gamma(n, \alpha)$ (Leçon 13). b) Observer que $\{S_{n+1} < t\} \subset \{S_n < t\}$ ce qui conduit à l'égalité demandée. Ensuite, exprimer la probabilité $\mathbb{P}(S_{n+1} < t)$ à partir de la densité de la loi de S_{n+1} et effectuer une intégration par parties pour se ramener à celle de S_n. Il faut trouver que N_t suit une loi de Poisson (ainsi que l'indique le titre de l'exercice).

Exercice 4 (*Loi de Student*). Soient X et Y deux variables indépendantes, X de loi $\mathcal{N}(0,1)$ et Y de loi Gamma $\gamma(\frac{n}{2}, \frac{1}{2})$ de paramètres $\frac{n}{2}$ et $\frac{1}{2}$, aussi appelée loi du χ^2 à $n \geq 1$ degrés de liberté. La loi de

$$T_n \ = \ \frac{X}{\sqrt{\frac{Y}{n}}}$$

est appelée loi de Student à n degrés de liberté. Montrer que cette loi a une densité, et la décrire (*ainsi que son histoire !*).

Exercice 5. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de même loi normale centrée réduite $\mathcal{N}(0,1)$ et soit X le vecteur $(X_1, \ldots, X_n)$.

Soit O une matrice $n \times n$ orthogonale, et soit $Y = OX$; rappeler pourquoi $\sum_{k=1}^n X_k^2 = \sum_{k=1}^n Y_k^2$.

a) Démontrer que Y a même loi que X.

Poser

$$\overline{X} = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{et} \quad S^2 = \frac{1}{n-1} \sum_{k=1}^n \big(X_k - \overline{X}\big)^2.$$

b) En choisissant O dont tous les termes de la dernière ligne sont égaux à $\frac{1}{\sqrt{n}}$, vérifier que

$$(n-1)S^2 = \sum_{k=1}^n X_k^2 - n\overline{X}^2 = \sum_{k=1}^n Y_k^2 - Y_n^2 = \sum_{k=1}^{n-1} Y_k^2.$$

c) Déduire de ce qui précède que $\overline{X}$ et S^2 sont indépendantes et que $(n-1)S^2$ a la même loi que $\sum_{k=1}^{n-1} X_k^2$.

Corrigé. Une application orthogonale conserve les longueurs, donc $|Y| = |OX|$. a) Par indépendance des coordonnées, la loi de X est la loi gaussienne standard $\mathcal{N}(0, \mathrm{Id})$ sur $\mathbb{R}^n$. Comme transformation linéaire de X, $Y = OX$ est également gaussienne, centrée, de matrice de covariance décrite par

$$\mathbb{E}\big(\langle c, Y\rangle^2\big) = \mathbb{E}\big(\langle c, OX\rangle^2\big) = \mathbb{E}\big(\langle {}^\top Oc, X\rangle^2\big) = |{}^\top Oc|^2 = |c|^2$$

pour tout $c \in \mathbb{R}^n$. Ainsi Y est aussi de loi $\mathcal{N}(0, \mathrm{Id})$. b) Comme

$$\sum_{k=1}^n \big(X_k - \overline{X}\,\big)^2 = \sum_{k=1}^n \big(X_k^2 - 2\overline{X}X_k + \overline{X}^2\,\big) = \sum_{k=1}^n X_k^2 - n\overline{X}^2$$

(c'est une formule de variance !), la première égalité est satisfaite. Par le choix de O, $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k = \sqrt{n}\,\overline{X}$, d'où la deuxième égalité puisque $\sum_{k=1}^n X_k^2 = \sum_{k=1}^n Y_k^2$. c) Puisque $Y = (Y_1, \ldots, Y_n)$ suit la loi gaussienne standard $\mathcal{N}(0, \mathrm{Id})$ de $\mathbb{R}^n$, les composantes $Y_1, \ldots, Y_n$ sont indépendantes (de même loi $\mathcal{N}(0,1)$).

Donc $\overline{X} = \frac{1}{\sqrt{n}} Y_n$ est indépendante de $\sum_{k=1}^{n-1} Y_k^2 = (n-1)S^2$. Enfin, comme Y a même loi que X, il est clair de la même identité $(n-1)S^2 = \sum_{k=1}^{n-1} Y_k^2$ que $(n-1)S^2$ a même loi que $\sum_{k=1}^{n-1} X_k^2$.

Exercice 6 (*Théorème de Glivenko*[1]*-Cantelli*[2]). Soit X_n, $n \in \mathbb{N}$, une suite de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes de même loi P sur les boréliens de $\mathbb{R}$ de fonction de répartition F.

Pour tout $n \geq 1$, et tout $t \in \mathbb{R}$, considérer la variable aléatoire, dite *fonction de répartition empirique*, $F_n(t) : (\Omega, \mathcal{A}, \mathbb{P}) \to [0,1]$ définie pour tout $\omega \in \Omega$ par

$$F_n(t)(\omega) \;=\; \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{]-\infty, t]}\big(X_k(\omega)\big).$$

C'est la fonction de répartition de la mesure (discrète) $\frac{1}{n} \sum_{k=1}^{n} \delta_{X_k(\omega)}$ sur $\mathcal{B}(\mathbb{R})$.

a) Démontrer que pour tout $t \in \mathbb{R}$, $\lim_{n\to\infty} F_n(t) = F(t)$ presque sûrement.

b) Soit $L \geq 1$ un entier et soient $t_0 < t_1 < \cdots < t_L$ des réels ; démontrer que, presque sûrement, $\lim_{n\to\infty} \max_{0 \leq \ell \leq L} |F_n(t_\ell) - F(t_\ell)| = 0$.

Il est supposé dans la suite du problème que P est la loi uniforme sur $[0,1]$.

c) Démontrer que, pour tout $L \geq 1$ et tout $n \geq 1$,

$$\sup_{t \in [0,1]} \big|F_n(t) - t\big| \;\leq\; \max_{0 \leq \ell \leq L} \big|F_n\big(\tfrac{\ell}{L}\big) - \tfrac{\ell}{L}\big| + \tfrac{1}{L}\,.$$

d) Déduire de ce qui précède que l'ensemble

$$A \;=\; \Big\{\omega \in \Omega\,;\, \lim_{n\to\infty} \sup_{t \in [0,1]} \big|F_n(t)(\omega) - t\big| = 0\Big\}$$

contient un ensemble mesurable de probabilité égale à 1.

1. Valery Glivenko, mathématicien ukrainien et soviétique (1896–1940).
2. Francesco Paolo Cantelli, mathématicien italien (1875–1966).

Corrigé. a) Pour $t \in \mathbb{R}$ fixé, les variables aléatoires

$$\omega \in \Omega \to Y_k(\omega) = \mathbb{1}_{]-\infty,t]}\big(X_k(\omega)\big), \quad k \geq 1,$$

sont indépendantes de même loi intégrable $\mathbb{E}(Y_1) = \mathbb{P}(X_1 \leq t) = F(t)$. Ainsi, d'après la loi des grands nombres, $\frac{1}{n}\sum_{k=1}^n Y_k \to F(t)$ presque sûrement. Donc la fonction de répartition empirique $F_n(t)$ au point t (qui donne la fréquence des éléments de l'échantillon $(X_1, \ldots, X_n)$ plus petits que t) approche la fonction de répartition théorique $F(t)$ de la loi commune des X_k, a priori inconnue. Bien évidemment, une application statistique performante souhaitera que cette approximation ait lieu pour le plus de $t \in \mathbb{R}$ possibles. Or, à ce stade, l'ensemble de probabilité 1 sur lequel la convergence a lieu dépend de t, ce qui constitue la matière de l'exercice. b) D'après la question précédente, pour chaque $\ell = 0, 1, \ldots, L$, il existe $\Omega_\ell \in \mathcal{A}$ tel que $\mathbb{P}(\Omega_\ell) = 1$ et $\lim_{n\to\infty} F_n(t_\ell)(\omega) = F(t_\ell)$ pour tout $\omega \in \Omega_\ell$. L'ensemble $\bigcap_{\ell=0}^L \Omega_\ell$ est de probabilité 1 (car la probabilité de son complémentaire est inférieure ou égale à $\sum_{\ell=0}^L \mathbb{P}(\Omega_\ell^c) = 0$) ; si $\omega \in \bigcap_{\ell=0}^L \Omega_\ell$, $\lim_{n\to\infty} F_n(t_\ell)(\omega) = F(t_\ell)$ pour tout $\ell = 0, 1, \ldots, L$, ce qui entraîne la conclusion demandée. c) L'argument s'appuie simplement sur les propriétés de croissance des fonctions de répartition (empiriques). En effet, si $t \in]0, 1]$, il existe $k = 1, \ldots, L$ tel que $\frac{k-1}{L} < t \leq \frac{k}{L}$. Alors (en tout ω, omis dans les écritures),

$$F_n(t) - t \leq F_n\big(\tfrac{k}{L}\big) - \tfrac{k-1}{L} = \big(F_n\big(\tfrac{k}{L}\big) - \tfrac{k}{L}\big) + \tfrac{1}{L}$$

de sorte que

$$F_n(t) - F(t) \leq \max_{0 \leq \ell \leq L} \big|F_n\big(\tfrac{\ell}{L}\big) - \tfrac{\ell}{L}\big| + \tfrac{1}{L}\,.$$

De la même façon,

$$F_n(t) - t \geq F_n\big(\tfrac{k-1}{L}\big) - \tfrac{k}{L} = \big(F_n\big(\tfrac{k-1}{L}\big) - \tfrac{k-1}{L}\big) - \tfrac{1}{L}$$

et donc aussi

$$F_n(t) - F(t) \geq -\max_{0 \leq \ell \leq L} \big|F_n\big(\tfrac{\ell}{L}\big) - \tfrac{\ell}{L}\big| - \tfrac{1}{L}\,.$$

Les bornes étant uniformes en t, l'inégalité demandée s'ensuit (le cas $t = 0$ étant immédiat puisque $F_n(0) = 0$, les variables de loi uniforme étant concentrées sur $[0,1]$). d) Il faut prendre soin dans cette question que l'ensemble A n'est pas nécessairement mesurable, puisqu'il met en jeu un supremum (réunion) sur un ensemble non dénombrable. Pour $t_\ell = \frac{\ell}{L}$, $\ell = 0, 1, \ldots, L$, poser $B_L = \bigcap_{\ell=0}^{L} \Omega_\ell$ comme dans la question b), en rappelant que $\mathbb{P}(B_L) = 1$. Puis considérer $B = \bigcap_{L \geq 1} B_L$, avec encore $\mathbb{P}(B) = 1$. L'objectif est de vérifier que $B \subset A$ ce qui répondra à la question. Soit donc $\omega \in B$ fixé ; pour tout $\varepsilon > 0$, soit $L = L(\varepsilon) \geq 1$ tel que $\frac{1}{L} \leq \varepsilon$; comme $\omega \in B$, alors $\omega \in \Omega_L$, et il existe donc $n_0 = n_0(\varepsilon)$ tel que si $n \geq n_0$,

$$\max_{0 \leq \ell \leq L} \left| F_n\big(\tfrac{\ell}{L}\big)(\omega) - \tfrac{\ell}{L} \right| \leq \varepsilon.$$

D'après c) et le choix de L, si $n \geq n_0$,

$$\sup_{t \in [0,1]} \left| F_n(t)(\omega) - t \right| \leq \varepsilon + \tfrac{1}{L} \leq 2\varepsilon.$$

Autrement dit

$$\lim_{n \to \infty} \sup_{t \in [0,1]} \left| F_n(t)(\omega) - t \right| = 0,$$

ce qui exprime que $\omega \in A$. Le résultat est établi.

Remarque. Il est possible d'éviter l'emploi des ε par les limites supérieures de suites. En effet, sur B, et donc sur B_L pour chaque $L \geq 1$, d'après les questions b) et c),

$$\limsup_{n \to \infty} \sup_{t \in [0,1]} \left| F_n(t)(\omega) - t \right| \leq \limsup_{n \to \infty} \max_{0 \leq \ell \leq L} \left| F_n\big(\tfrac{\ell}{L}\big) - \tfrac{\ell}{L} \right| + \tfrac{1}{L} = \tfrac{1}{L}.$$

Comme $L \geq 1$ est arbitraire,

$$\limsup_{n \to \infty} \sup_{t \in [0,1]} \left| F_n(t)(\omega) - t \right| = 0$$

soit le résultat annoncé.

Exercice 7 (*Lois Gamma et Beta*). Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de même loi exponentielle $\mathcal{E}(\alpha)$ de paramètre $\alpha > 0$.

a) Démontrer que la loi de $X_1 + \cdots + X_n$, est une loi Gamma $\gamma(n, \alpha)$ (Exercice 8, Leçon 10), de densité $x \mapsto \frac{1}{\Gamma(n)} \alpha^n x^{n-1} e^{-\alpha x}$ par rapport à la mesure de Lebesgue sur $]0, \infty[$ (rappeler que $\Gamma(n) = (n-1)!$).

Pour plus de simplicité dans la suite, prendre $\alpha = 1$ (le cas général s'en déduisant compte tenu que αX_1 suit une loi exponentielle de paramètre 1 si X_1 est de paramètre α – à vérifier).

b) Démontrer que $\frac{X_1}{X_1 + X_2}$ est indépendant de $X_1 + X_2$, et suit la loi uniforme sur $]0, 1[$.

c*) Généraliser la question précédente en déterminant, pour $k = 1, \ldots, n-1$, la loi de

$$\frac{X_1 + \cdots + X_k}{X_1 + \cdots + X_n},$$

dite Beta et notée $\beta(k, n-k)$. En déduire la valeur de $\int_0^1 x^{k-1}(1-x)^{n-k-1} dx$. (*Indication* : écrire $X_1 + \cdots + X_n = (X_1 + \cdots + X_k) + (X_{k+1} + \cdots + X_n)$ et utiliser la question a).)

À l'image des lois $\gamma(n, \alpha)$, les lois Beta $\beta(k, \ell)$ peuvent être extrapolées pour $k, \ell \in]0, \infty[$.

Corrigé. Outre l'Exercice 8, Leçon 10, la question a) est évoquée dans la Leçon 13 à partir du produit de convolution. La question b) est traitée dans l'Exercice 6, Leçon 12 ; elle est aussi contenue dans la question c*) détaillée ci-dessous avec $n = 2$. c*) Pour $k = 1, \ldots, n-1$ fixé, poser $S_k = X_1 + \cdots + X_k$ de sorte que S_k et $S_n - S_k$ sont indépendantes, S_k de loi $\gamma(k, 1)$ et $S_n - S_k$ de loi $\gamma(n-k, 1)$, et

$$\frac{X_1 + \cdots + X_k}{X_1 + \cdots + X_n} = \frac{S_k}{S_k + (S_n - S_k)}.$$

La loi du couple $(S_k, S_n - S_k)$ est donc le produit des lois $\gamma(k, 1)$ et $\gamma(n-k, 1)$.

Ainsi, si $\phi : \mathbb{R} \to \mathbb{R}$ est borélienne, positive ou bornée,

$$\begin{aligned}
&\mathbb{E}\Big(\phi\Big(\frac{S_k}{S_k + (S_n - S_k)}\Big)\Big) \\
&\qquad = \frac{1}{\Gamma(k)}\,\frac{1}{\Gamma(n-k)} \int_{]0,\infty[^2} \phi\Big(\frac{x}{x+y}\Big) x^{k-1} y^{n-k-1} e^{-(x+y)} d\lambda^2(x,y).
\end{aligned}$$

Le changement de variables

$$h(x,y) = (u,v) = \Big(\frac{x}{x+y}, x+y\Big)$$

de $]0,\infty[^2$ dans $]0,1[\times\,]0,\infty[$, de jacobien $\text{dét}(J_h(x,y)) = -\frac{1}{x+y}$, fournit

$$\begin{aligned}
&\mathbb{E}\Big(\phi\Big(\frac{S_k}{S_k + (S_n - S_k)}\Big)\Big) \\
&\qquad = \frac{1}{\Gamma(k)}\,\frac{1}{\Gamma(n-k)} \int_{]0,\infty[}\int_{]0,1[} \phi(u) u^{k-1} (1-u)^{n-k-1} v^{n-1} e^{-v} d\lambda^2(u,v) \\
&\qquad = \frac{\Gamma(n)}{\Gamma(k)\,\Gamma(n-k)} \int_{]0,1[} \phi(u) u^{k-1} (1-u)^{n-k-1} d\lambda(u)
\end{aligned}$$

après utilisation du théorème de Fubini-Tonelli et de la valeur de la fonction Gamma $\int_{]0,\infty[} v^{n-1} e^{-v} d\lambda(v) = \Gamma(n)$. Ainsi, la variable aléatoire $\frac{S_k}{S_k+(S_n-S_k)}$ a pour densité

$$u \mapsto \frac{\Gamma(n)}{\Gamma(k)\,\Gamma(n-k)}\, u^{k-1}(1-u)^{n-k-1}$$

par rapport à la mesure de Lebesgue sur $]0,1[$, dite loi Beta (de paramètres $(k, n-k)$). En choisissant ϕ la fonction constante égale à 1, il découle des calculs précédents que

$$1 = \frac{\Gamma(n)}{\Gamma(k)\,\Gamma(n-k)} \int_{]0,1[} u^{k-1}(1-u)^{n-k-1} d\lambda(u)$$

ce qui fournit la valeur de $\beta(k, n-k)$.

Si $n = 2$ et $k = 1$, la loi $\beta(k, n-k) = \beta(1,1)$ est la loi uniforme sur $]0,1[$. Comme déjà observé quand $n = 2$ (Exercice 6, Leçon 12), la même démonstration appliquée à

$$\mathbb{E}\left(\phi\left(\frac{S_k}{S_k + (S_n - S_k)}\right)\psi(S_n)\right)$$

pour $\phi, \psi : \mathbb{R} \to \mathbb{R}$ boréliennes, positives ou bornées, assure en outre que $\frac{S_k}{S_n}$ est indépendante de S_n.

Exercice 8*. Soient $X_1, \ldots, X_n, X_{n+1}$ des variables aléatoires indépendantes de même loi exponentielle $\mathcal{E}(1)$ de paramètre 1 ; poser $S_k = X_1 + \cdots + X_k$, $k = 1, \ldots, n+1$.

a) Démontrer que la loi du vecteur aléatoire

$$\left(\frac{X_1}{S_{n+1}}, \frac{X_2}{S_{n+1}}, \ldots, \frac{X_n}{S_{n+1}}\right)$$

est uniforme sur (le simplexe) $\Delta = \left\{y \in]0,\infty[^n ; \sum_{k=1}^n y_k < 1\right\}$.

b) Vérifier que

$$\left(\frac{X_1}{S_{n+1}}, \frac{X_2}{S_{n+1}}, \ldots, \frac{X_{n+1}}{S_{n+1}}\right)$$

est indépendant de S_{n+1} (et donc aussi $\left(\frac{X_1}{S_{n+1}}, \frac{X_2}{S_{n+1}}, \ldots, \frac{X_n}{S_{n+1}}\right)$).

Corrigé. En cette fin de cours (!), la correction se propose de répondre aux deux questions a) et b) simultanément. La méthode est entièrement similaire à celle développée dans la Proposition 3. Pour des fonctions boréliennes $\phi_k : \mathbb{R} \to \mathbb{R}$, $k = 1, \ldots, n+1$, et $\psi : \mathbb{R} \to \mathbb{R}$, positives ou bornées, par le théorème de transport pour la loi de $(X_1, \ldots, X_{n+1})$,

$$\mathbb{E}\left(\phi_1\left(\frac{X_1}{S_{n+1}}\right)\cdots\phi_{n+1}\left(\frac{X_{n+1}}{S_{n+1}}\right)\psi(S_{n+1})\right)$$
$$= \int_{]0,+\infty[^{n+1}} \phi_1\left(\frac{x_1}{\sum_{k=1}^{n+1} x_k}\right)\cdots\phi_{n+1}\left(\frac{x_{n+1}}{\sum_{k=1}^{n+1} x_k}\right)\psi\left(\sum_{k=1}^{n+1} x_k\right) e^{-\sum_{k=1}^{n+1} x_k} d\lambda^{n+1}.$$

Il convient à ce stade, comme dans la Proposition 3, d'effectuer un changement de variables approprié. L'application $h\ :]0,\infty[^{n+1}\to\ \Delta\times]0,\infty[$ définie pour $x\in]0,\infty[^{n+1}$ par

$$h(x)\ =\ \left(\frac{x_1}{\sum_{k=1}^{n+1}x_k},\ldots,\frac{x_n}{\sum_{k=1}^{n+1}x_k},\sum_{k=1}^{n+1}x_k\right)$$

est un difféomorphisme. En effet, Il est aisé de vérifier que h est bien à valeurs dans $\Delta\times\,]0,\infty[$. Pour y dans cet ensemble, ses antécédants, c'est-à-dire les $x\in]0,\infty[^{n+1}$ tels que $y=h(x)$ vérifient

$$y_\ell\ =\ \frac{x_\ell}{\sum_{k=1}^{n+1}x_k},\ \ell=1,\ldots,n,\quad y_{n+1}\ =\ \sum_{k=1}^{n+1}x_k.$$

Cette caractérisation est équivalente à dire que $x_\ell=y_\ell\,y_{n+1}$, $\ell=1,\ldots,n$, et $x_{n+1}=y_{n+1}\big(1-\sum_{\ell=1}^n y_\ell\big)$. Donc chaque y de $\Delta\times\,]0,\infty[$ admet un unique antécédant $x=h^{-1}(y)$ où x est définie par les formules précédentes. L'application h est donc une bijection, et l'application h^{-1} est sa réciproque. Le jacobien de cette dernière est donné par le déterminant

$$\text{dét}\left(\left(\frac{\partial h_k^{-1}}{\partial y_\ell}(y)\right)_{k,\ell}\right)\ =\ \begin{vmatrix} y_{n+1} & 0 & \cdots & 0 & y_1\\ 0 & y_{n+1} & \ddots & \vdots & y_2\\ \vdots & \ddots & \ddots & 0 & \vdots\\ 0 & \cdots & 0 & y_{n+1} & y_n\\ -y_{n+1} & -y_{n+1} & \cdots & -y_{n+1} & 1-\sum_{\ell=1}^n y_\ell \end{vmatrix}.$$

Après avoir ajouté les n premières lignes à la dernière,

$$\text{dét}\left(\left(\frac{\partial h_k^{-1}}{\partial y_\ell}(y)\right)_{k,\ell}\right)\ =\ \begin{vmatrix} y_{n+1} & 0 & \cdots & 0 & y_1\\ 0 & y_{n+1} & \ddots & \vdots & y_2\\ \vdots & \ddots & \ddots & 0 & \vdots\\ 0 & \cdots & 0 & y_{n+1} & y_n\\ 0 & & \cdots & 0 & 1 \end{vmatrix}\ =\ y_{n+1}^n.$$

Par application de ce changement de variables (Leçon 5),

$$\begin{aligned}\mathbb{E}\bigg(\phi_1\bigg(\frac{X_1}{S_{n+1}}\bigg)\cdots\phi_{n+1}\bigg(\frac{X_{n+1}}{S_{n+1}}\bigg)\,\psi(S_{n+1})\bigg)\\ = \int_{\Delta\times\,]0,+\infty[}\phi_1(y_1)\cdots\phi_n(y_n)\phi_{n+1}\bigg(1-\sum_{\ell=1}^{n}y_\ell\bigg)\\ \cdot\,\psi(y_{n+1})\,y_{n+1}^n e^{-y_{n+1}}\,d\lambda^n(y_1,\ldots,y_n)d\lambda^1(y_{n+1}).\end{aligned}$$

La séparation des variables, $y_1,\ldots,y_n$ d'une part, y_{n+1} d'autre part, permet déjà d'assurer l'indépendance entre le vecteur $\big(\frac{X_1}{S_{n+1}},\ldots,\frac{X_{n+1}}{S_{n+1}}\big)$ et S_{n+1} (question b)). La formule autorise également l'identification des lois (question a)). Il peut être rappelé pour commencer que $\int_{]0,\infty[}y_{n+1}^n e^{-y_{n+1}}d\lambda^1 = \Gamma(n+1) = n!$. En choisissant les fonctions ϕ_{n+1} et ψ constantes égales à 1, il vient ainsi

$$\mathbb{E}\bigg(\phi_1\bigg(\frac{X_1}{S_{n+1}}\bigg)\cdots\phi_n\bigg(\frac{X_n}{S_{n+1}}\bigg)\bigg) = n!\int_\Delta \phi_1\cdots\phi_n\,d\lambda^n$$

ce qui exprime que le vecteur $\big(\frac{X_1}{S_{n+1}},\ldots,\frac{X_n}{S_{n+1}}\big)$ est uniforme sur Δ (qui a pour mesure de Lebesgue $\frac{1}{n!}$). En considérant toutes les fonctions ϕ_k, $k=1,\ldots,n+1$, égales à 1, la formule permet aussi de retrouver que S_{n+1} suit une loi Gamma $\gamma(n+1,1)$ de densité $y_{n+1}\mapsto\frac{1}{\Gamma(n+1)}\,y_{n+1}^n e^{-y_{n+1}}$ par rapport à la mesure de Lebesgue sur $]0,\infty[$.