SÉANCE 4 : FONCTIONS DE VARIABLES ALÉATOIRES ; CORRIGE

Exercice 1

Soit X une variable aléatoire de fonction de répartition continue F. Calculez la distribution de

(a) X^2 (b) $\sqrt{X}$
(c) $|X|$ (d) $\lfloor X \rfloor$
(e) $X^+ = \max\{0, X\}$.

Solution

L'idée de l'exercice consiste à exprimer les fonctions de répartition des nouvelles variables aléatoires en fonction de F. Ce genre de manipulations est tout à fait standard en probabilités, et sera désormais considéré comme "vu, connu et acquis".

(a)

$$P(X^2 \le x) = \begin{cases} 0 & \text{si } x \le 0 \\ F(\sqrt{x}) - F(-\sqrt{x}) & \text{sinon.} \end{cases}$$

(b) Notons qu'il faut que la variable aléatoire X soit positive, donc $F(0) = 0$.

$$P(\sqrt{X} \le x) = \begin{cases} 0 & \text{si } x \le 0 \\ F(x^2) & \text{sinon.} \end{cases}$$

(c)

$$P(|X| \le x) = \begin{cases} 0 & \text{si } x \le 0 \\ F(x) - F(-x) & \text{sinon.} \end{cases}$$

(d) Ce cas-ci étant discret, nous cherchons à déterminer $P(\lfloor X \rfloor = n)$ pour tout n entier. On voit facilement que

$$P(\lfloor X \rfloor = n) = P(n \le X < n+1) = F(n+1) - F(n).$$

Notez toutefois que si jamais F n'était pas continue (mais juste continue à droite), cette expression serait d'une forme légèrement différente (laquelle ?).

(e) Voir TP3.

Exercice 2

Soient F et G des fonctions de répartition. Soient $\lambda \in [0, 1]$ et r un entier positif. Montrez que les fonctions suivantes sont encore des fonctions de répartition.

(a) $\lambda F(x) + (1 - \lambda)G(x)$ (b) $F(x)G(x)$
(c) $F^r(x)$ (d) $1 - (1 - F(x))^r$
(e) $F(e^x)$ (f) $F(\alpha x + b)$

Donnez, dans les trois derniers cas, une variable aléatoire dont la loi est donnée par cette fonction.

Solution

Pour chaque exemple, il s'agit de montrer que la fonction proposée est continue à droite, croissante, qu'elle vaut 0 en sa plus petite valeur et 1 en sa plus grande valeur. Chouette petit exercice... donc nous n'allons pas le détailler ici...mais passer directement à la question supplémentaire, plus intéressante !

(d) Supposons que nous ayons un échantillon de r observations i.i.d. de loi F. La loi demandée est alors celle de $\min\{X_1, X_2, \ldots, X_r\}$. A quelle variable aléatoire associerait-on maintenant la loi sous (c) ?

(e) Soit X une variable aléatoire positive de fonction de répartition F. La loi recherchée est celle de $\ln(X)$.

(f) Soit X une variable aléatoire quelconque de fonction de répartition F. Que diriez-vous de la loi de $\frac{X-b}{\alpha}$ quand $\alpha \neq 0$? Pour $\alpha = 0$, la fonction proposée ne serait plus une fonction de répartition...

Exercice 3

Soit X une variable aléatoire de fonction de répartition F, et soient $a < b$ des réels. Dessinez l'allure des fonctions de répartition des *variables aléatoires tronquées* Y et Z définies par

$$Y = \left\{ \begin{array}{ll} a & \text{si } X \leq a, \\ X & \text{si } a \leq X \leq b, \\ b & \text{si } X > b. \end{array} \right. , \quad Z = \left\{ \begin{array}{ll} X & \text{si } |X| \leq b, \\ 0 & \text{si } \ |X| > b. \end{array} \right.$$

Que se passe-t-il lorsque $a \to -\infty, b \to \infty$?

Solution

Cet exercice-ci est CAPITAL pour une bonne compréhension des fonctions de répartition ! Nous n'allons pas faire de dessin ici, mais détailler tous les calculs qui y mènent.

- Pour la variable aléatoire Y, une première observation de la situation révèle qu'elle prend ses valeurs dans $[a, b]$, ce qui donne déjà que $P(Y < a) = 0$ et $P(Y \leq b) = 1$. Maintenant, nous constatons que a et b jouent des rôles particuliers ; en effet, on assigne à chacun une probabilité strictement positive, ce qui montre déjà que Y est de loi mixte avec des atomes en a et en b, menant à des sauts de discontinuité de la fonction de répartition en ces points. Clairement, $P(Y = a) = P(X \leq a) = F(a)$ et $P(Y = b) = P(X > b) = 1 - F(b)$. Finalement, pour $a < x < b$, nous avons $P(Y \leq x) = P(Y = a) + P(a < Y \leq x) = F(x)$. La fonction de répartiton $F^Y(\cdot)$ prend donc l'allure suivante : elle est nulle jusqu'en a, où elle subit un saut de discontinuité de hauteur $F(a)$. Entre a et b, elle suit exactement la fonction F, et en b elle saute directement en 1 (saut de discontinuité de $1 - F(b)$). Il est facile de voir que les conditions évoquées au début sont vérifées.
- Pour la variable aléatoire Z, on devine maintenant que le point de discontinuité doit forcément se trouver en 0. De plus, Z prend ses valeurs dans l'intervalle $[-b, b]$. Comme précédemment, $P(Z < -b) = 0$ et $P(Z \leq b) = 1$. Par ailleurs, $P(Z = 0) = P(|X| > b) = F(-b) + 1 - F(b)$. Pour $-b \leq x < 0$, nous avons $P(Z \leq x) = P(-b \leq Z \leq x) = P(-b \leq X \leq x) = F(x) - F(-b)$. Si par contre x varie entre 0 (exclus) et $b\,(> 0)$, alors $P(Z \leq x) = P(-b \leq X \leq 0) + P(Z = 0) + P(0 \leq X \leq x) = F(0) - F(-b) + F(-b) + 1 - F(b) + F(x) - F(0) = 1 + F(x) - F(b)$. On en tire que $F^Z(\cdot)$ est nulle jusqu'en $-b$ (inclus), que sur $[-b, 0)$ $F^Z(x) = F(x) - F(-b)$, qu'il y a un saut de discontinuité en 0 de $F(-b) + 1 - F(b)$, et qu'après, sur $(0, b]$, $F^Z(x) = 1 + F(x) - F(b)$. On vérifie aisément que les conditions en $-b$ et b énoncées préalablement sont satisfaites.

Exercice 4

Supposons que vous disposiez d'un algorithme permettant d'engendrer une suite de nombres réels $x_1, x_2 \ldots$ qui sont distribués uniformément dans $(0, 1)$. Expliquez comment vous pouvez utiliser cet algorithme pour engendrer une suite de nombres réels $y_1, y_2 \ldots$ distribués selon une loi exponentielle négative de paramètre $\lambda > 0$.

Solution

Comme vu à l'exercice 2 du TP 3, nous savons que $F^{-1}(U) \sim F$ si U est uniforme sur [0,1]. Un fait à exploiter ! Dès lors, il suffit de déterminer l'inverse de la fonction de répartition d'une variable aléatoire de loi exponentielle négative pour obtenir une telle suite de nombres réels.

D'un point de vue algorithmique, on aurait donc :

- begin
- for i=1 :n, $u_i = unif[0,1]$, end
- for i=1 :n, $y_i = F_{\exp}^{-1}(u_i)$, end
- end

Exercice 5

Soit Y une variable aléatoire suivant une loi dont la fonction de densité est notée g. Soit U une deuxième variable aléatoire, de loi uniforme sur $[0,1]$ et indépendante de Y. Soit f une fonction de densité continue pour laquelle il existe $c \in \mathbb{R}$ tel que $\dfrac{f(y)}{g(y)} \leq c$ pour tout $y \in \mathbb{R}$. Quelle est la fonction de densité de la variable aléatoire

$$X \stackrel{\mathcal{D}}{=} Y \,|\, U \leq \frac{f(Y)}{cg(Y)}?$$

Solution

A première vue, un tel exercice peut paraître choquant si on n'en a jamais rencontré avant... or, on en a déjà fait un du même tonneau au TP3, donc nous sommes à l'aise. Faisons alors les calculs tranquillement :

$$\begin{aligned}
P(X \leq x) &= P\left(Y \leq x \,|\, U \leq \frac{f(Y)}{cg(Y)}\right) \\
&= \frac{P\left(Y \leq x \cap U \leq \frac{f(Y)}{cg(Y)}\right)}{P\left(U \leq \frac{f(Y)}{cg(Y)}\right)} \\
&= \frac{\int_{-\infty}^{\infty} P\left(Y \leq x \cap U \leq \frac{f(Y)}{cg(Y)} \,|\, Y = y\right) g(y)dy}{\int_{-\infty}^{\infty} P\left(U \leq \frac{f(Y)}{cg(Y)} \,|\, Y = y\right) g(y)dy} \\
&= \frac{\int_{-\infty}^{x} P\left(U \leq \frac{f(y)}{cg(y)}\right) g(y)dy}{\int_{-\infty}^{\infty} P\left(U \leq \frac{f(y)}{cg(y)}\right) g(y)dy} \\
&= \frac{\int_{-\infty}^{x} \frac{f(y)}{cg(y)} g(y)dy}{\int_{-\infty}^{\infty} \frac{f(y)}{cg(y)} g(y)dy} \\
&= \frac{\int_{-\infty}^{x} f(y)dy}{\int_{-\infty}^{\infty} f(y)dy} \\
&= \int_{-\infty}^{x} f(y)dy
\end{aligned}$$

et il s'ensuit que la variable aléatoire X suit une loi de fonction de densité f (assez inattendu non ?).

Exercice 6

Vous avez su transformer votre algorithme de l'exercice 4 de telle manière que vous savez maintenant engendrer des variables aléatoires distribuées selon une loi exponentielle négative. Quelqu'un vous demande à présent d'engendrer des variables aléatoires de fonction de répartition F.

a) Dans un premier temps, vous restez modestes et lui proposez de le faire via l'algorithme engendrant des lois uniformes. Facile maintenant, et efficace. Quelle ligne rajoutez-vous allègrement ?

b) Une deuxième personne vient vous demander la même chose, mais pour une fonction de répartition G. La modestie fait vite place à une grande assurance, et d'un grand sourire vous déclarez que vous savez le faire à partir de la loi F que vous venez d'engendrer. Comment vous y prenez-vous ?

c) Jamais deux sans trois, surtout quand on est notoirement reconnu comme spécialiste en programmation : une troisième personne vient vous voir pour les mêmes raisons, cependant sa loi-cible à elle ne possède pas de fonction de répartition "simple", et elle ne vous donne que la densité. Comment vous y prenez-vous pour sauver votre honneur ?

Solution

a) La ligne à rajouter est quasi la même que pour l'exercice 4, à savoir : for i=1 :n, $x_i = F^{-1}(u_i)$, end.

b) De par l'exercice 2 du TP3, nous savons que $F^{-1}(U) \sim F$ et que $F(X) \sim U$ si $X \sim F$. Dès lors, si $Y \sim F$, alors $G^{-1}(F(Y)) \sim G$ par ce qui précède. La ligne à rajouter est désormais claire : for i=1 :n, $z_i = G^{-1}(F(x_i))$, end.

c) N'étant pas en possession d'une fonction de répartition, notre méthode basée sur l'interaction avec une loi uniforme ne s'applique plus. Que faire ? Ah, nous avons vu un résultat intéressant à l'exercice 5 : on est parti d'une densité g, et grâce à des lois uniformes nous sommes tombés sur une loi de densité f. Voilà la solution du problème ! Partons donc ici d'une variable aléatoire Y de loi G et de densité g (par exemple celle obtenue au point précédent). Déterminons une constante c telle que $f(y) \leq cg(y)$ en toute valeur de y. Engendrons ensuite des variables aléatoires Y et U (uniforme sur [0,1]) jusqu'à ce que la condition $u \leq \frac{f(y)}{cg(y)}$ soit satisfaite (on utilise des minuscules ici puisqu'on parle de réalisations des variables aléatoires). Alors, nous savons que la valeur obtenue pour Y, à savoir y, représente un élément de la population de loi de densité f. Dans les lignes-code à rajouter doit donc figurer une condition "while" (comment le feriez-vous ?).

Exercice 7

Soient k un entier et $\lambda > 0$. Soient $X_1, X_2, \ldots$ une suite de variables aléatoires indépendantes et identiquement distribuées, de loi exponentielle négative de paramètre λ. Alors la variable aléatoire $Z = \sum_{i=1}^{k} X_i$ est distribuée selon une loi Erlang de paramètres k et λ (on note $Z \sim \text{Erlang}(\lambda, k)$). Sa densité est donnée par

$$f_{\lambda,k}(z) = \begin{cases} \frac{\lambda^k z^{k-1} e^{-\lambda z}}{(k-1)!} & \text{si } z > 0, \\ 0 & \text{sinon.} \end{cases}$$

a) Calculez le moment d'ordre n de Z, pour $n \in \mathbb{N}$.

b) Soient $Z_1 \sim \text{Erlang}(\lambda, k_1)$ et $Z_2 \sim \text{Erlang}(\lambda, k_2)$. Si Z_1 et Z_2 sont indépendantes, quelle sera la loi de $Z_1 + Z_2$? (Justifiez avec précision !)

Solution

Pour des raisons évidentes (voir exercice qui suit), le point a) de cet exercice ne sera pas corrigé ici.

Venons-en à présent au point b) du présent exercice. De par la définition d'une loi Erlang, $Z_1 = \sum_{i=1}^{k_1} X_i$ et $Z_2 = \sum_{j=1}^{k_2} Y_j$, où les X_i et Y_j sont i.i.d. de loi exponentielle négative de paramètre λ. De plus, l'indépendance entre Z_1 et Z_2 entraîne que X_i est indépendant de Y_j pour tout $i = 1, \ldots, k_1$ et $j = 1, \ldots, k_2$. Ainsi, $Z_1 + Z_2$ peut s'écrire comme somme de $k_1 + k_2$ variables aléatoires i.i.d. de loi exponentielle négative de paramètre λ. Le résultat en découle.

Exercice 8

La généralisation de la loi Erlang-k au cas où $k \notin \mathbb{N}$ est connue sous le nom de loi Gamma de paramètres $a > 0$ et $b > 0$ (on note $Z \sim \Gamma(a, b)$). Sa densité est donnée par

$$f_{a,b}(z) = \begin{cases} \frac{a^b}{\Gamma(b)} z^{b-1} e^{-az} & \text{si } z > 0 \\ 0 & \text{sinon,} \end{cases}$$

où $\Gamma(b) = \int_0^{+\infty} x^{b-1} e^{-x} dx$.

a) Montrez que lorsque $b = k \in \mathbb{N}_0$ on retrouve la loi Erlang(a, k).

b) Montrez que le moment d'ordre k de Z est donné par

$$E(Z^k) = \frac{\Gamma(b+k)}{\Gamma(b)} a^{-k}.$$

c) Soient $Z \sim \Gamma(a, b)$ et $t \in \mathbb{R}$. Quelle sera la loi de $Y \stackrel{\mathcal{D}}{=} tZ$?

Solution

a) Si $b = k \in \mathbb{N}_0$, alors $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx = [-x^{k-1} e^{-x}]_0^\infty + (k-1) \int_0^\infty x^{k-2} e^{-x} dx$ après intégration par parties. Il est bien connu que $\lim_{x\to\infty} \frac{x^{k-1}}{e^x} = 0$. Nous obtenons donc $\Gamma(k) = (k-1) \int_0^\infty x^{k-2} e^{-x} dx$. En appliquant le même raisonnement encore $k-2$ fois, nous obtenons

$$\Gamma(k) = (k-1)! \int_0^\infty e^{-x} dx = (k-1)!.$$

La loi Gamma de paramètres a et $k \in \mathbb{N}_0$ devient

$$f_{a,b}(z) = \begin{cases} \frac{a^k}{(k-1)!} z^{k-1} e^{-az} & \text{si } z > 0 \\ 0 & \text{sinon,} \end{cases}$$

qui est bien la loi Erlang de paramètres (a, k).

b) Calculons :

$$\begin{aligned} E(Z^k) &= \int_{-\infty}^\infty z^k f_{a,b}(z) dz \\ &= \int_{-\infty}^\infty z^k \frac{a^b}{\Gamma(b)} z^{b-1} e^{-az} \mathbf{I}_{\mathbb{R}_0^+}(z) dz \\ &= \int_0^\infty \frac{a^b}{\Gamma(b)} z^{b+k-1} e^{-az} dz \\ &= a^{-k} \frac{\Gamma(b+k)}{\Gamma(b)} \int_0^\infty \frac{a^{b+k}}{\Gamma(b+k)} z^{b+k-1} e^{-az} dz \\ &= a^{-k} \frac{\Gamma(b+k)}{\Gamma(b)}, \end{aligned}$$

où le passage de l'avant-dernière à la dernière ligne suit du fait que cette intégrale vaut 1 en tant qu'intégrale d'une fonction de densité sur tout son domaine.

c) Trois cas sont à envisager pour $Y \stackrel{\mathcal{D}}{=} tZ$:

- si $t = 0$, alors $Y \stackrel{\mathcal{D}}{=} 0$, donc la loi de Y est $P(Y = 0) = 1$;
- si $t > 0$, alors $\forall y \in \mathbb{R}$, $P(Y \leq y) = P(Z \leq \frac{y}{t})$, donc la densité f_Y de Y est donnée par

$$f_Y(y) = f_{a,b}(\frac{y}{t})\frac{1}{t} = \frac{a^b}{\Gamma(b)}\left(\frac{y}{t}\right)^{b-1} e^{-a\frac{y}{t}}\mathbf{I}_{\mathbb{R}_0^+}(\frac{y}{t})\frac{1}{t} = \frac{(\frac{a}{t})^b}{\Gamma(b)}y^{b-1}e^{-\frac{a}{t}y}\mathbf{I}_{\mathbb{R}_0^+}(y) = f_{\frac{a}{t},b}(y),$$

  donc $Y \sim \Gamma(\frac{a}{t}, b)$;
- si $t < 0$, alors nous nous ramenons au cas précédent avec $t = -|t|$ et voyons que $Y \stackrel{\mathcal{D}}{=} -Y'$ où $Y' \sim \Gamma(\frac{a}{|t|}, b)$.