Contrôle

Documents et calculatrice sont autorisés.
Le barème est donné à titre indicatif.

Une loi Bêta

Le but de cet exercice est d'étudier quelques propriétés de la loi Bêta de paramètres (2,1), notée $\boldsymbol{\beta}$. Sa densité est définie sur $[0,1]$ par

$$f(x) = 2x.$$

Il s'agit d'une des lois Bêta les plus simples, mais les propriétés que nous allons étudier ici s'étendent aux lois Bêta les plus complexes.

Nous considérons dans la suite une variable U suivant la loi $\boldsymbol{\beta}$ ci-dessus.

1. **Caractéristiques de la loi Bêta** (2 pts)
   (a) Calculer l'espérance et la variance de U.
   $\mathbf{E}(U) = \int_0^1 x \cdot 2x\mathrm{d}x = \frac{2}{3}$ et $\mathbf{V}(U) = \int_0^1 (x - \frac{2}{3})^2 \cdot 2x\mathrm{d}x = \frac{1}{18}$.
   (b) Déterminer la médiane de U, le nombre m pour lequel $\mathbf{P}(U \leqslant m) = \mathbf{P}(U \geqslant m)$.
   On cherche m tel que $\mathbf{P}(U \leqslant m) = \frac{1}{2}$, donc $\int_0^m 2x\mathrm{d}x = \frac{1}{2}$. On obtient $m^2 = \frac{1}{2}$, donc $m = \sqrt{\frac{1}{2}}$.

2. **Logarithme de la loi Bêta** (3 pts)
   On définit $V = -\ln(U)$. On souhaite déterminer la loi de V.
   (a) Quelles sont les valeurs prises par V ?
   U est à valeurs dans $[0,1]$, donc $\ln(U)$ est à valeurs dans $]-\infty, 0]$ et V est à valeurs dans $\mathbf{R}_+$.
   Notons que si $U = 0$, V n'est pas défini, mais la probabilité de cet évènement est nulle.
   (b) Soit t une de ces valeurs. Déterminer la valeur de la fonction de répartition de V en t : $F_V(t) = \mathbf{P}(V \leqslant t)$.
   Soit $t \in \mathbf{R}_+$. Alors
   $$F_V(t) = \mathbf{P}(V \leqslant t) = \mathbf{P}(-\ln(U) \leqslant t) = \mathbf{P}(U \geqslant \mathrm{e}^{-t}) = \int_{\mathrm{e}^{-t}}^1 2x\mathrm{d}x = 1 - \mathrm{e}^{-2t}.$$
   (c) En déduire que V suit la loi exponentielle de paramètre 2 : $V \sim \mathcal{E}(2)$.
   On en déduit que la densité de V est donnée par :
   $$f_V(t) = F_V'(t) = 2\mathrm{e}^{-2t}.$$
   On reconnaît la densité de la loi exponentielle $\mathcal{E}(2)$. Donc $V \sim \mathcal{E}(2)$.

3. **Théorème central limite** (6 pts)

On souhaite estimer une probabilité pour un produit de variables de même loi Bêta.

(a) Soient V_k des variables indépendantes de même loi exponentielle $\mathcal{E}(2)$ et $S_n = \sum_{k=1}^n V_k$. Énoncer le théorème central limite pour la variable S_n.

La variable $\frac{S_n - n\mathbf{E}(V)}{\sigma(V)\sqrt{n}} = \frac{S_n - n/2}{\sqrt{n}/2}$ converge en loi vers la loi normale $\mathcal{N}(0,1)$. Concrètement, cela signifie que pour n assez grand ($n \geqslant 30$), on peut considérer que $\frac{S_n - n/2}{\sqrt{n}/2} \sim \mathcal{N}(0,1)$.

(b) Estimer ainsi $\mathbf{P}(S_{50} \leqslant 28)$.

$$\mathbf{P}(S_{50} \leqslant 28) = \mathbf{P}\left(\frac{S_{50} - 25}{\sqrt{50}/2} \leqslant \frac{28 - 25}{\sqrt{50}/2}\right) \approx \mathbf{P}\left(\frac{S_{50} - 25}{\sqrt{50}/2} \leqslant 0,85\right).$$

D'après le TCL, on peut approcher cette probabilité par

$$F_{\mathcal{N}}(0,85) \approx 0,80.$$

Soient maintenant U_k des variables indépendantes de même loi Bêta et $P_n = \prod_{k=1}^n U_k$.

(c) Estimer $\mathbf{P}\left((\frac{7}{4})^{50} P_{50} \geqslant 1\right)$.

Indication : passer au logarithme et utiliser les résultats des questions 2-c et 3-b.

$$\mathbf{P}\left((\tfrac{7}{4})^{50} P_{50} \geqslant 1\right) = \mathbf{P}\left(\ln\left((\tfrac{7}{4})^{50} P_{50}\right) \geqslant 0\right) = \mathbf{P}(50 \ln(7/4) + \sum_{k=1}^{50} \ln(U_k) \geqslant 0)$$

$$\approx \mathbf{P}(\sum_{k=1}^{50}(-\ln(U_k)) \leqslant 28).$$

Si on note $V_k = -\ln(U_k)$, alors d'après 2-c, les variables V_k sont indépendantes de même loi $\mathcal{E}(2)$. Cette probabilité est alors celle que nous avons calculé précédemment, donc

$$\mathbf{P}\left((\tfrac{7}{4})^{50} P_{50} \geqslant 1\right) \approx 0,80.$$

(d) Déterminer une valeur de n à partir de laquelle $\mathbf{P}\left((\frac{7}{4})^n P_n \geqslant 1\right) \approx 0,95$.

On reprend le raisonnement ci-dessus avec n quelconque :

$$\mathbf{P}\left((\tfrac{7}{4})^n P_n \geqslant 1\right) = \mathbf{P}(\sum_{k=1}^{n}(-\ln(U_k)) \leqslant n\ln(7/4)) = \mathbf{P}(\sum_{k=1}^{n} V_k \leqslant n\ln(7/4))$$

$$= \mathbf{P}\left(\frac{S_n - n/2}{\sqrt{n}/2} \leqslant \frac{n\ln(7/4) - n/2}{\sqrt{n}/2}\right).$$

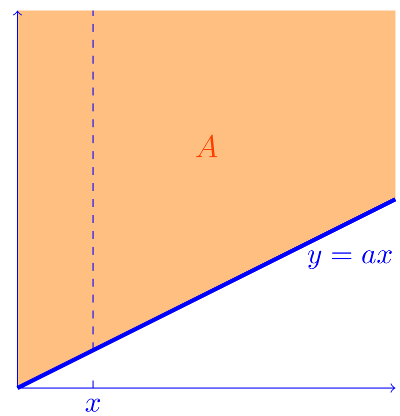
D'après 2-a, la variable $\frac{S_n - n/2}{\sqrt{n}/2}$ suit approximativement la loi normale centrée réduite. Ainsi, cette dernière probabilité vaut approximativement $F_{\mathcal{N}}((2\ln(7/4) - 1)\sqrt{n})$. Pour qu'elle soit égale à $0,95$, il faut que $(2\ln(7/4) - 1)\sqrt{n} \approx 1,64$ et on obtient ainsi $n \approx 190$.

4. **Couple de variables** (5 pts)

Soient X et Y deux variables indépendantes à valeurs dans $[0, +\infty[$. La variable X suit la loi d'Erlang de densité $f_X(x) = x\mathrm{e}^{-x}$ et Y la loi exponentielle de densité $f_Y(y) = \mathrm{e}^{-y}$.

On pose $Z = \dfrac{X}{X+Y}$. On souhaite déterminer la loi de Z.

(a) Soit $a > 0$.
Représenter dans le quart de plan $\mathbf{R}_+ \times \mathbf{R}_+$ le domaine $A = \{(x, y) \mid ax \leqslant y\}$.



(b) Calculer à l'aide d'une intégrale double la probabilité $\mathbf{P}(aX \leqslant Y)$.

L'ensemble A est paramétré par $A = \{(x, y)\ ;\ x \in [0, +\infty[, y \in [ax, +\infty[\}$. Ainsi :

$$\mathbf{P}(aX \leqslant Y) = \iint_A f_X(x) f_Y(y) \mathrm{d}x\mathrm{d}y = \int_{x=0}^{+\infty} \int_{y=ax}^{+\infty} x\mathrm{e}^{-x}\mathrm{e}^{-y} \mathrm{d}y\ \mathrm{d}x$$

$$= \int_{x=0}^{+\infty} x\mathrm{e}^{-x} \left[-\mathrm{e}^{-y}\right]_{ax}^{+\infty} \mathrm{d}x = \int_{x=0}^{+\infty} x\mathrm{e}^{-x}\mathrm{e}^{-ax} \mathrm{d}x$$

$$= \left[-\tfrac{1}{a+1} x\mathrm{e}^{-(a+1)x} - \tfrac{1}{(a+1)^2}\mathrm{e}^{-(a+1)x}\right]_0^{+\infty} = \frac{1}{(a+1)^2}.$$

(c) Quelles sont les valeurs prises par la variable Z ?

Les variables X et Y sont à valeurs positives, donc Z l'est également, mais de plus $X \leqslant X + Y$, donc $Z \leqslant 1$. On peut alors vérifier facilement que Z est à valeurs dans $[0, 1]$.

(d) Déterminer, à l'aide de la question b, la fonction de répartition de Z.

Soit $t \in [0, 1]$. Alors

$$F_Z(t) = \mathbf{P}\left(\tfrac{X}{X+Y} \leqslant t\right) = \mathbf{P}\left(\tfrac{1-t}{t} X \leqslant Y\right).$$

D'après le calcul précédent, on déduit avec $a = \frac{1-t}{t}$:

$$F_Z(t) = \frac{1}{(\frac{1-t}{t} + 1)^2} = \frac{1}{(\frac{1}{t})^2} = t^2.$$

(e) En déduire que $Z \sim \boldsymbol{\beta}$.

Ainsi la densité de Z est donnée par

$$\forall t \in [0,1], \quad f_Z(t) = F_Z'(t) = 2t.$$

On reconnaît la densité de la loi $\boldsymbol{\beta}$, donc $Z \sim \boldsymbol{\beta}$.

5. **Variable discrète** (6 pts)

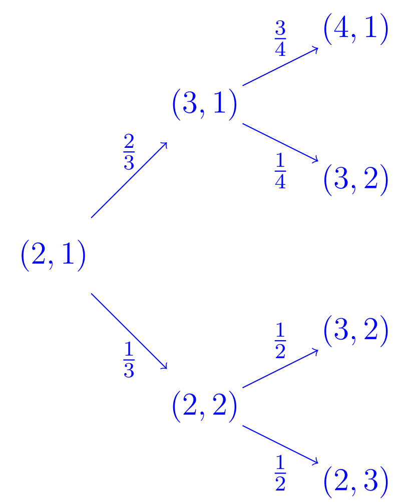
On considère une urne contenant initialement deux boules rouges et une boule bleue. À chaque étape, on tire une boule au hasard, puis on la remet dans l'urne en ajoutant une boule de la même couleur.

Par exemple, si on tire une boule rouge à la première étape, on la remet avec une autre boule rouge et l'urne contient alors trois boules rouges et une boule bleue. Puis on tire une nouvelle boule et ainsi de suite.

On note X_n la proportion de boules rouges dans l'urne au bout de n étapes. (Par exemple, X_0 est la variable constante égale à $\frac{2}{3}$.)

(a) Représenter, à l'aide d'un arbre de probabilités les deux premières étapes du processus.

Dans l'arbre, (r, b) désigne les nombres r de boules rouges et b de boules bleues à l'étape n.



(b) En déduire les lois de X_1 et X_2 et représenter leurs densités.

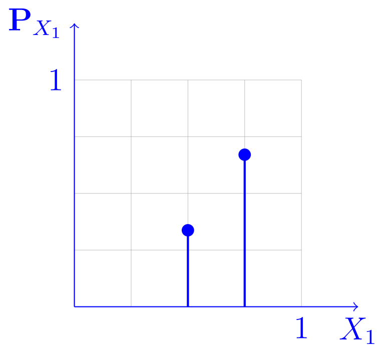
Ainsi X_1 vaut soit $\frac{3}{4}$ (3 boules rouges sur 4 boules) avec une probabilité $\frac{2}{3}$, soit $\frac{2}{4}$ avec une probabilité $\frac{1}{3}$.

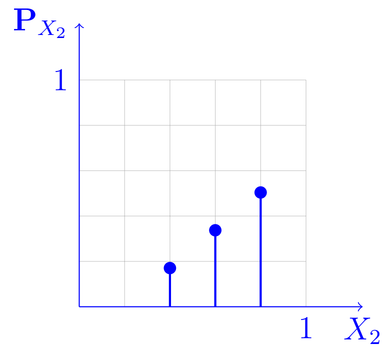
Et, en suivant les différentes branches de l'arbre, X_2 vaut :

$\frac{4}{5}$ avec une probabilité $\frac{2}{3} \times \frac{3}{4} = \frac{1}{2}$,

$\frac{3}{5}$ avec une probabilité $\frac{2}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{3}$,

$\frac{2}{5}$ avec une probabilité $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$.





On constate que la répartition des probabilités ressemble (en version discrète) à la densité de la loi $\boldsymbol{\beta}$. C'est cohérent avec la propriété limite donnée plus bas.

(c) Justifier que pour tout n et tout $k \leqslant n$,

$$\mathbf{P}\left(X_{n+1} = \frac{k+2}{n+4}\right) = \frac{n-k+1}{n+3}\mathbf{P}\left(X_n = \frac{k+2}{n+3}\right) + \frac{k+1}{n+3}\mathbf{P}\left(X_n = \frac{k+1}{n+3}\right).$$

Cette formule vient de l'idée suivante. À la fin de l'étape $n+1$, il y a $n+4$ boules dans l'urne. Pour qu'il y ait parmi elles $k+2$ boules, il y a deux possibilités : il y en avait $k+1$ à l'étape n et on a tiré une boule rouge, ou il y avait déjà $k+2$ boules rouges à l'étape n et on a tiré une boule bleue. Autrement dit, on suit dans l'arbre de probabilités, les deux branches qui aboutissent à l'évènement $X_{n+1} = \frac{k+2}{n+4}$, et on obtient la formule voulue.

Cette formule permet de déterminer récursivement la loi de X_n. Nous n'allons pas le faire car la loi obtenue est difficile à utiliser dans les calculs. Nous allons admettre que lorsque n tend vers l'infini, la loi de X_n converge vers la loi $\boldsymbol{\beta}$.

(d) Estimer ainsi la probabilité, qu'après un très grand nombre de tirages, plus de la moitié des boules de l'urne soient rouges.

Si n est assez grand, la proportion de boules rouges suit la loi $\boldsymbol{\beta}$. La probabilité demandée est alors estimée par :

$$\mathbf{P}(X_n \geqslant \frac{1}{2}) = \mathbf{P}_{\boldsymbol{\beta}}([\tfrac{1}{2}, 1]) = \int_{\frac{1}{2}}^{1} 2x\mathrm{d}x = \frac{3}{4}.$$

(e) Expliquer brièvement pourquoi il n'est pas possible d'appliquer ici le théorème central limite pour montrer que la loi des variables X_n converge vers une loi normale.

Les variables X_n ne sont ni des sommes, ni des moyennes, elles ne suivent pas la même loi et ne sont pas du tout indépendantes (la valeur de X_n est fortement liée à celle de X_{n-1} ; par exemple, si $X_1 = \frac{3}{4}$, il est impossible que $X_2 = \frac{2}{5}$).
Nous ne sommes donc absolument pas dans le cadre du TCL et cela n'aurait aucun sens de l'appliquer ici.

Rappels

- *La loi exponentielle de paramètre λ, notée $\mathcal{E}(\lambda)$ a pour densité, espérance et variance :*

$$\forall t \in \mathbf{R}_+,\ f(t) = \lambda \mathrm{e}^{-\lambda t}, \qquad \mathbf{E} = \frac{1}{\lambda}, \qquad \mathbf{V} = \frac{1}{\lambda^2}.$$

- *Table de la loi normale centrée réduite : pour* $X \sim \mathcal{N}(0,1)$

α	*0*	*0.25*	*0.52*	*0.85*	*1.28*	*1.64*	*1.96*	*2.33*
$\mathbf{P}(X \leqslant \alpha)$	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*0.95*	*0.975*	*0.99*

- *Soit* $c \in \mathbf{R}$. *On donne la primitive suivante :*

$$\int x\mathrm{e}^{-cx}\mathrm{d}x = -\frac{1}{c}x\mathrm{e}^{-cx} - \frac{1}{c^2}\mathrm{e}^{-cx}.$$