Variables aléatoires

Des lois

Exercice 2. On lance $n = 30$ fois un dé à 6 faces. On note X la variable aléatoire égale au nombre de 6 obtenus. On se pose la question suivante : est-il plus probable d'obtenir un nombre inférieur à $\frac{n}{3}$ ou supérieur à $\frac{n}{4}$?

1. Répondre en s'aidant de Python.
2. Répondre sans s'aider de Python.
   Une façon de faire : on pourra appliquer les inégalités de Markov et de Bienaymé-Tchebytchev.
3. Répondre pour les autres valeurs de $n \in \mathbb{N}$.

1. Le top serait de simuler cette expérience aléatoire. Ici je me contente d'effectuer les calculs de $P\big(X \leqslant \frac{n}{3}\big)$ et de $P\big(X \geqslant \frac{n}{4}\big)$ lorsque X suit une loi binomiale de paramètres $n = 30$ et $p = \frac{1}{6}$. J'utilise donc juste Python comme une calculatrice.
   $P\big(X \leqslant \frac{n}{3}\big) = P\big(X \leqslant 10\big) = \sum\limits_{k=0}^{10} \binom{30}{k}\big(\frac{1}{6}\big)^k\big(\frac{5}{6}\big)^{30-k} \approx 0.993$ (ah quand même).
   $P\big(X \geqslant \frac{n}{4}\big) = P\big(X \geqslant 8\big) = \sum\limits_{k=8}^{30} \binom{30}{k}\big(\frac{1}{6}\big)^k\big(\frac{5}{6}\big)^{30-k} \approx 0.114$ (gloups).
   ```
   >>> sum([factorial(30)/(factorial(k)*factorial(30-k))*(1/6)**k*(5/6)**(30-k) for k in range(0,11)])
   0.9932546114726499
   >>> sum([factorial(30)/(factorial(k)*factorial(30-k))*(1/6)**k*(5/6)**(30-k) for k in range(8,31)])
   0.11368680121359265
   ```
   Dont acte : $P\big(X \leqslant \frac{n}{3}\big) \geqslant P\big(X \geqslant \frac{n}{4}\big)$ pour cette valeur de n.

2. Avec Markov : $P\big(X \geqslant \frac{n}{4}\big) \leqslant \frac{E(X)}{n/4} = \frac{n/6}{n/4} = \frac{2}{3}$ car $X \geqslant 0$. On voit que c'est grossier, mais continuons.
   Avec Bienaymé-Tchebychev : déjà voyons pourquoi on peut l'utiliser. On a $\big(X \leqslant \frac{n}{3}\big) = \big(0 \leqslant X \leqslant \frac{n}{3}\big) = \big(-\frac{n}{6} \leqslant X - \frac{n}{6} \leqslant \frac{n}{6}\big) = \Big(\big|X - \frac{n}{6}\big| \leqslant \frac{n}{6}\Big) = \Big(\big|X - E(X)\big| \leqslant \frac{n}{6}\Big)$. Et donc $P\big(X \leqslant \frac{n}{3}\big) = P\Big(\big|X - E(X)\big| \leqslant \frac{n}{6}\Big) \geqslant P\Big(\big|X - E(X)\big| < \frac{n}{6}\Big)$ par croissance (pour l'inclusion) d'une probabilité. Et donc c'est parti pour un coup de Bienaymé-Tchebychev : $P\Big(\big|X - E(X)\big| < \frac{n}{6}\Big) \geqslant 1 - \frac{V(X)}{\big(n/6\big)^2} = 1 - \frac{5n/36}{n^2/36} = 1 - \frac{5}{n}$.
   Pour $n \geqslant 15$ (en particulier pour $n = 30$), on a donc $P\big(X \geqslant \frac{n}{4}\big) \leqslant P\big(X \geqslant 8\big) = \sum\limits_{k=8}^{30} \binom{30}{k}\big(\frac{1}{6}\big)^k\big(\frac{1}{6}\big)^{30-k} \approx 0.114$

3. Yapuka faire à la main les valeurs de n inférieures à 14 en reprenant le calcul Python précédent. Comme on peut le voir, l'égalité est dans le même sens pour presque toutes (mais pas toutes) les valeurs de n.

Exercice 9. *Lemme des coalitions*
X_1, ..., X_r, X_{r+1}, ..., X_n étant des variables aléatoires réelles mutuellement indépendantes :

1. si les $f_i : \mathbb{R} \to \mathbb{R}$ alors $f_1(X_1)$, ..., $f_n(X_n)$ sont mutuellement indépendantes ;
2. si $f : \mathbb{R}^r \to \mathbb{R}$ et $g : \mathbb{R}^{n-r} \to \mathbb{R}$ alors $f(X_1, \dots, X_r)$ et $g(X_{r+1}, \dots, X_n)$ sont indépendantes.

1. On veut montrer que $f_1(X_1)$, ..., $f_n(X_n)$ sont mutuellement indépendantes. Soient $x_1, \dots, x_n \in \mathbb{R}$.
   Il s'agit de montrer qu'on a $P(f(X_1) = x_1 \text{ et } \dots \text{ et } f(X_n) = x_n) = P(f(X_1) = x_1) \cdots P(f(X_n) = x_n)$.

   On a : $P(f(X_1) = x_1 \text{ et } \dots \text{ et } f(X_n) = x_n)$
   $= P(X_1 \in f_1^{-1}(\{x_1\} \text{ et } \dots \text{ et } X_n \in f_n^{-1}(\{x_n\})$ par définition
   $= P(X_1 \in f_1^{-1}(\{x_1\})) \cdots P(X_n \in f_n^{-1}(\{x_n\}))$ par indépendance mutuelle de $X_1, \dots X_n$
   $= P(f(X_1) = x_1)) \cdots P(f(X_n) = x_n))$ par définition

   Et c'est tout !

2. Notons que deux v.a.r. X et Y sont indépendantes si et seulement si la famille (X, Y) est une famille de v.a.r. mutuellement indépendantes, par définition. Le résultat précédent indique donc que, si X et Y sont des v.a. indépendantes et f, g des fonctions à valeurs réelles alors $f(X)$ et $g(Y)$ sont indépendantes.

   Notons $X = (X_1, \dots, X_r)$ et $Y = (X_{r+1}, .., X_n)$. Montrons que X et Y sont indépendantes. Soit $x = (x_1, \dots, x_r) \in \mathbb{R}^r$ et $y = (y_1, \dots, y_{n-r}) \in \mathbb{R}^{n-r}$. Par définition l'événement $(X = x \text{ et } Y = y)$ n'est autre que l'évévement $(X_1 = x_1 \text{ et } X_2 = x_2 \text{ et } \dots \text{ et } X_r = x_r \text{ et } X_{r+1} = y_1 \text{ et } X_{r+2} = y_2 \text{ et } \dots \text{ et } X_n = y_{n-r})$. Par indépendance mutuelle on a donc : $P(X = x \text{ et } Y = y) = P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_r = x_r)P(X_{r+1} = y_1)P(X_{r+2} = y_2)\cdots P(X_n = y_{n-r})$. Mais de même, on obtient $P(X = x) = P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_r = x_r)$ et $P(Y = y) = P(X_{r+1} = y_1)P(X_{r+2} = y_2)\cdots P(X_n = y_{n-r})$, donc $P(X = x \text{ et } Y = y) = P(X = x)P(Y = y)$.

   Ainsi, X et Y sont indépendantes. D'après ce qui précède, c'est donc aussi le cas de $f(X)$ et $g(Y)$.

Exercice 10. (BCE 2011)

Attention, dans le corrigé il est parfois fait référence à "un exercice traité en cours" dans le but de vous faire ouvrir votre cours. Bien évidemment on ne peut pas rédiger ainsi lors d'un écrit de concours.

Soit n un entier naturel non nul : $n \geqslant 1$ joueurs visent une cible. Chaque joueur effectue deux tirs. À chaque tir, chaque joueur a la probabilité $p \in]0, 1[$ d'atteindre la cible. Les tirs sont indépendants les uns des autres.
On définit la variable aléatoire X égale au nombre de joueurs ayant atteint la cible au premier tir et la variable aléatoire Z égale au nombre de joueurs ayant atteint la cible au moins une fois à l'issue de deux tirs. On note enfin $Y = Z - X$.

L'exercice n'est pas énoncé en termes mathématiques. Il faut donc modéliser le problème avant de commencer la résolution proprement dite. Ici cette modélisation est essentiellement donnée dans le sujet en indication, mais elle ne l'était pas dans l'énoncé originel.

— On **admet** que l'expérience aléatoire dont il est question peut se modéliser par un espace probabilisé fini $(\Omega, \mathcal{P}(\Omega), P)$ sur lequel seront définies toutes les variables aléatoires qui suivent.

— Pour tout $i \in \{1, \dots, n\}$, on note X_i la variable aléatoire qui vaut 1 lorsque le premier tir du joueur i atteint la cible et 0 sinon. De même, on note T_i la variable aléatoire qui vaut 1 lorsque le second tir du joueur i atteint la cible et 0 sinon. On note Z_i la variable aléatoire qui vaut 1 lorsque l'un des deux tirs au moins du joueur i atteint la cible et 0 sinon. On note Y_i la variable aléatoire qui vaut 1 lorsque le premier tir du joueur i manque sa cible et le second l'atteint, et 0 sinon.

— On interprête la deuxième phrase de l'énoncé par le fait qu'on a $\forall i \in \{1, \dots, n\},\ X_i \sim B(p)$ et $T_i \sim B(p)$.

— On interprête la troisième phrase de l'énoncé par le fait que la famille $(X_1, \dots, X_n, T_1, \dots, T_n)$ est une famille de variables aléatoires mutuellement indépendantes.

— Par définition, on a $Z_i = \max(X_i, T_i)$ ce qui se réécrit $Z_i = X_i + T_i - X_iT_i$. De même, par définition, on a $Y_i = Z_i - X_i$ ce qui se réécrit $Y_i = T_i(1 - X_i)$

— D'après l'énoncé, on a $X = \sum_{i=1}^{n} X_i$, $Z = \sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} X_i + T_i - X_iT_i$ et $Y = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} T_i(1 - X_i)$

Comme souvent, maintenant que le problème est convenablement modélisé, les questions qui suivent deviennent très faciles.

1. Quelle est la loi de X ?

   C'est une somme de n variables aléatoires indépendantes suivant toutes une loi de Bernoulli de même paramètre p, donc d'après le cours $X \sim B(n, p)$.

2. Quelle est la loi de Z ?

   On commence par Z_i : Z_i est le max de deux variables aléatoires indépendantes suivant une loi de Bernoulli de même paramètre p, donc d'après un exercice traité en cours, $Z_i \sim B(p(2 - p))$.
   Maintenant Z est une somme de n variables aléatoires indépendantes suivant toutes une loi de Bernoulli de même paramètre $p(2 - p)$, donc d'après le cours $X \sim B(n, p(2 - p))$.

3. Que représente la variable aléatoire Y ? Quelle est sa loi ?

Y compte le nombre de joueurs qui ont échoué au premier tir mais réussit au second tir.
Cherchons la loi de Y_i. Tout d'abord on a $X_i \sim B(p)$ donc $1-X_i \sim B(1-p)$. Ensuite T_i et X_i sont indépendantes donc d'après le cours $f(T_i)$ et $g(X_i)$ le sont quelles que soient les fonctions f et g ; en particulier T_i et $(1-X_i)$ le sont. Finalement Y_i est le produit de deux variables aléatoires indépendantes suivant une loi de Bernoulli, donc d'après un exercice traité en cours, elle suit une loi de Bernoulli de paramètre le produit des paramètres de X_i et $1-T_i$, à savoir $Y_i \sim B(p(1-p))$.
Maintenant Y est une somme de n variables aléatoires indépendantes suivant toutes une loi de Bernoulli de même paramètre $p(1-p)$, donc d'après le cours $Y \sim B(n, p(1-p))$.

4. Les variables X et Y sont-elle indépendantes ?

On calcule :
· $P(X=0 \text{ et } Y=0) = P(\text{"Aucun joueur ne réussit aucun de ses tirs"}) = (1-p)^{2n}$.
· $P(X=0)P(Y=0) = (1-p)^n \big[p(1-p)\big]^n = p^n(1-p)^{2n}$.
Comme on a $p \notin \{0,1\}$, on a donc $P(X=0 \text{ et } Y=0) \neq P(X=0)P(Y=0)$ et les variables X et Y sont indépendantes.
Remarque : on pouvait aussi calculer $P(X=n \text{ et } Y=1) = 0 \neq P(X=n)P(Y=1)$ pour $p \neq \{0,1\}$.

5. Quelle est leur covariance ?

On peut présenter le calcul de différentes façons mais les trois arguments sont la bilinéarité de la covariance (ou la linéarité de l'espérance), le fait que X_i et $T_j(1-X_j)$ sont indépendantes, et le fait que $X_i(1-X_i)$ est nulle.

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n Y_j\right) = \mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n T_j(1-X_j)\right) && \text{d'après les préliminaires}\\
&= \sum_{(i,j)\in\{1,\dots,n\}^2} \mathrm{Cov}\left(X_i, T_j(1-X_j)\right) && \text{par bilinéarité de la covariance}\\
&= \sum_{i\neq j} \mathrm{Cov}\left(X_i, T_j(1-X_j)\right) + \sum_{i=1}^n \mathrm{Cov}\left(X_i, T_i(1-X_i)\right) && \text{en réordonnant les termes}
\end{aligned}$$

Calculons chaque terme séparément.

· Pour $i \neq j$, les variables X_i, T_j et X_j sont mutuellement indépendantes. D'après un théorème du cours les variables X_i et $f(T_j, X_j)$ sont donc indépendantes pour toute fonction f, et en particulier X_i et $T_j(1-X_j)$ sont indépendantes. En particulier leur covariance est nulle.

· Pour $i = j$, on a $\mathrm{Cov}(X_i, T_i(1-X_i)) = E(T_iX_i(1-X_i)) - E(X_i)E(T_i(1-X_i))$ d'après la formule de Kœnig-Hyugens. Or X_i est à valeurs dans $\{0,1\}$ (elle suit une loi de Bernoulli) donc $X_i(1-X_i)$ est nulle. Par ailleurs, les variables T_i et $1-X_i$ sont indépendantes (puisque T_i et X_i le sont) donc d'après le cours $E(T_i(1-X_i)) = E(T_i)E(1-X_i)$. Finalement on a $\mathrm{Cov}(X_i, T_i(1-X_i)) = E(0) - E(X_i)E(T_i)E(1-X_i) = 0 - p^2(1-p)$ car $1-X_i \sim B(1-p)$.

$$\text{Ainsi : } \boxed{\mathrm{Cov}(X,Y) = \sum_{i\neq j} 0 + \sum_{i=1}^n \big(-p^2(1-p)\big) = -np^2(1-p)}.$$