Variables aléatoires : loi et espérance.

1. Dans une population de n oiseaux, on en capture m que l'on bague puis que l'on relâche. Un peu plus tard, on en capture à nouveau m.

a) Soit $k \in \{0, \ldots, n\}$. Quelle est la probabilité que parmi les m oiseaux capturés, k soient bagués ?

b) Pour quelle valeur de k la probabilité calculée ci-dessus est-elle maximale ?

Solution de l'exercice 1.

a) Sans tenir compte de l'ordre de capture, il y a C_n^m manières de capturer les m oiseaux, que l'on suppose équiprobables. Capturer m animaux dont k bagués revient à en capturer k parmi les m bagués (C_m^k possibilités) et $m-k$ parmi les $n-m$ non bagués (C_{n-m}^{m-k} possibilités, en supposant $m-k \leq n-m$, autrement dit $k \geq 2m-n$, car sinon la probabilité est nulle), le nombre de manières de la faire est donc :

$$C_m^k C_{n-m}^{m-k} = \frac{m!(n-m)!}{k!(m-k)!^2(n-2m+k)!}.$$

Et la probabilité correspondante vaut :

$$P(n,m,k) = \frac{C_m^k C_{n-m}^{m-k}}{C_n^m} = \frac{m!^2(n-m)!^2}{n!k!(m-k)!^2(n-2m+k)!}.$$

b) Si k est tel que $\max(0, 2m-n) \leq k < m$, on obtient, en mettant le dénominateur sous forme canonique (en la variable k) :

$$\frac{P(n,m,k+1)}{P(n,m,k)} = \frac{(m-k)^2}{(k+1)(n-2m+k+1)} = \frac{(m-k)^2}{(n/2+1-(m-k))^2-(n/2-m)^2}.$$

Ce ratio est décroissant en k. Pour $k = m-1$ (dernière valeur), le numérateur vaut 1 et donc le ratio est strictement inférieur à 1. $P(n,m,k)$ est donc croissant en k jusqu'à un certain $\max(0, 2m-n) \leq k_0 \leq m-2$, et ensuite décroissant pour $k \geq k_0$ (voire décroissant depuis le début si $k_0 = \max(0, 2m-n)$). Il reste à déterminer k_0 qui va réaliser le maximum de la probabilité. D'après la formule précédente, on a :

$$\frac{P(n,m,k+1)}{P(n,m,k)} > 1 \Leftrightarrow (m-k)^2 < (n/2+1-(m-k))^2 - (n/2-m)^2.$$

En développant, cela se simplifie en :

$$P(n,m,k+1) > P(n,m,k) \Leftrightarrow (n+2)k < m^2 + 2m - n - 1.$$

Ainsi $P(n,m,k)$ est maximal pour

$$k = k_0 := 1 + \left\lfloor \frac{m^2+2m-n-1}{n+2} \right\rfloor = \left\lfloor \frac{(m+1)^2}{n+2} \right\rfloor .$$

Si $\frac{(m+1)^2}{n+2}$ est entier, la valeur maximale est réalisée aussi atteinte pour $k_0 - 1$, sinon elle n'est atteinte que pour k_0. On n'est pas surpris de constater que la valeur de k maximisant $P(n,m,k)$ est telle que la proportion d'oiseaux bagués k/m capturés la seconde fois est très proche de celle dans la population m/n.

2. Soient $\alpha, \beta \in]0,1[$ deux réels. Pour tout $(i,j) \in \mathbb{N}^2$, on pose $p_{i,j} = \alpha\beta(1-\alpha)^i(1-\beta)^j$.

a) Montrer qu'en posant $\mathbb{P}(\{(i,j)\}) = p_{ij}$ pour tout $(i,j) \in \mathbb{N}^2$, on définit une mesure de probabilités sur $\mathbb{N}^2$.

Pour tout $(i,j) \in \mathbb{N}^2$, on pose $X((i,j)) = i$ et $Y((i,j)) = j$.

b) Déterminer la loi de X et la loi de Y.

c) Calculer $\mathbb{P}(X < Y)$, $\mathbb{P}(X = Y)$ et $\mathbb{P}(X > Y)$.

Solution de l'exercice 2.

a) Il s'agit de vérifier que $\forall (i,j) \in \mathbb{N}^2$, $p_{i,j} \geq 0$ et $\sum_{(i,j)\in\mathbb{N}^2} p_{i,j} = 1$. La positivité des $p_{i,j}$ est évidente d'après leur définition. En utilisant $\sum_{i\in\mathbb{N}} x^i = 1/(1-x)$ pour $x = 1-\alpha$ et $x = 1-\beta$, on obtient :

$$\sum_{(i,j)\in\mathbb{N}^2} p_{i,j} = \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} \alpha\beta(1-\alpha)^i(1-\beta)^j = \sum_{i\in\mathbb{N}} \alpha(1-\alpha)^i = 1.$$

b) Soit $k \in \mathbb{N}$. En utilisant encore $\sum_{j\in\mathbb{N}} \beta(1-\beta)^j = 1$, on obtient :

$$\mathbb{P}(X = k) = \sum_{(i,j)\in\mathbb{N}^2} p_{i,j} \mathbb{1}_{\{k\}}(i) = \sum_{j\in\mathbb{N}} p_{k,j} = \alpha(1-\alpha)^k.$$

Ainsi, X suit une loi géométrique de paramètre α. De même, $\mathbb{P}(Y = k) = \beta(1-\beta)^k$, autrement dit Y suit une loi géométrique de paramètre β.

c)

$$\mathbb{P}(X < Y) = \sum_{\substack{(i,j)\in\mathbb{N}^2 \\ i<j}} p_{i,j} = \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} p_{i,j}.$$

Comme pour tout $i \in \mathbb{N}$, $\sum_{j=i+1}^{\infty} \beta(1-\beta)^j = (1-\beta)^{i+1}$, cela donne

$$\mathbb{P}(X < Y) = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i(1-\beta)^{i+1} = \sum_{i=0}^{\infty} \alpha(1-\beta)\left((1-\alpha)(1-\beta)\right)^i = \frac{\alpha(1-\beta)}{1-(1-\alpha)(1-\beta)} = \frac{\alpha-\alpha\beta}{\alpha+\beta-}$$

De même (en échangeant les rôles de X et Y, donc de α et β), on obtient

$$\mathbb{P}(X > Y) = \frac{\beta - \alpha\beta}{\alpha + \beta - \alpha\beta}.$$

On peut calculer $\mathbb{P}(X = Y)$ directement en sommant les $p_{i,i}$, ou alors on peut utiliser le fait que les parties $\{X = Y\}$, $\{X < Y\}$ et $\{X > Y\}$ forment une partition de $\mathbb{N}^2$ et les résultats précédents :

$$\mathbb{P}(X = Y) = 1 - \mathbb{P}(X > Y) - \mathbb{P}(X < Y) = \frac{\alpha + \beta - \alpha\beta - (\beta - \alpha\beta) - (\alpha - \alpha\beta)}{\alpha + \beta - \alpha\beta} = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}.$$

3. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Pour tout $n \geq 0$, on note $p_n = \mathbb{P}(X = n)$ et on suppose $p_n > 0$. Soit $\lambda > 0$ un réel. Montrer que les deux assertions suivantes sont équivalentes.
1. La variable aléatoire X suit la loi de Poisson de paramètre λ.
2. Pour tout $n \geq 1$, on a $\dfrac{p_n}{p_{n-1}} = \dfrac{\lambda}{n}$.

Solution de l'exercice 3. La première assertion s'écrit aussi $\forall n \in \mathbb{N}$, $p_n = \exp(-\lambda)\frac{\lambda^n}{n!}$. Il est clair que cela entraine la seconde assertion.

Supposons maintenant la seconde assertion vraie. Par récurrence sur n, on obtient immédiatement

$$p_n = p_0 \frac{\lambda^n}{n!}.$$

Il reste à déterminer p_0. Grâce à la formule précédente, l'égalité $\sum_{n\in\mathbb{N}} p_n = 1$ s'écrit $p_0 \sum_{n\in\mathbb{N}} \frac{\lambda^n}{n!} = 1$, ou encore $p_0 \exp(\lambda) = 1$. D'où le résultat.

4. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Montrer qu'on a l'égalité

$$\mathbb{E}[X] = \sum_{n\geq 1} \mathbb{P}(X \geq n).$$

Solution de l'exercice 4. Calculons le membre de droite, en utilisant le fait que pour toute famille à double indice $(a_{m,n})_{m,n\geq 1}$ de réels positifs, on a

$$\sum_{n\geq 1} \sum_{m\geq 1} a_{m,n} = \sum_{m\geq 1} \sum_{n\geq 1} a_{m,n}.$$

On trouve

$$\begin{aligned}\sum_{n\geq 1}\mathbb{P}(X\geq n) &= \sum_{n\geq 1}\sum_{m\geq n}\mathbb{P}(X=m)\\ &= \sum_{n\geq 1}\sum_{m\geq 1}\mathbb{1}_{\{m\geq n\}}\mathbb{P}(X=m)\\ &= \sum_{m\geq 1}\sum_{n\geq 1}\mathbb{1}_{\{m\geq n\}}\mathbb{P}(X=m)\\ &= \sum_{m\geq 1} m\mathbb{P}(X=m)\\ &= \mathbb{E}[X],\end{aligned}$$

ce qui est le résultat attendu.

5. Soient $X, Y : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{N}$ deux variables à valeurs dans $\mathbb{N}$. Montrer de deux façons différentes que

$$\sum_{n\geq 0} n\mathbb{P}(X=n) + \sum_{n\geq 0} n\mathbb{P}(Y=n) = \sum_{n\geq 0} n\mathbb{P}(X+Y=n).$$

Solution de l'exercice 5. Le membre de gauche est $\mathbb{E}(X)+\mathbb{E}(Y)$ et le membre de droite est $\mathbb{E}(X+Y)$. Par linéarité de l'espérance, ils sont donc égaux.

Donnons une deuxième démonstration de cette égalité. Calculons le membre de droite. On trouve

$$\begin{aligned}\sum_{n\geq 0} n\mathbb{P}(X+Y=n) &= \sum_{n\geq 0}\sum_{k=0}^{n} n\mathbb{P}(X=k \text{ et } Y=n-k)\\ &= \sum_{n\geq 0}\sum_{k\geq 0}\mathbb{1}_{\{k\leq n\}} n\mathbb{P}(X=k \text{ et } Y=n-k)\\ &= \sum_{k\geq 0}\sum_{n\geq 0}\mathbb{1}_{\{k\leq n\}} n\mathbb{P}(X=k \text{ et } Y=n-k)\\ &= \sum_{k\geq 0}\sum_{n\geq k} n\mathbb{P}(X=k \text{ et } Y=n-k)\\ &= \sum_{k\geq 0}\sum_{l\geq 0}(k+l)\mathbb{P}(X=k \text{ et } Y=l)\\ &= \sum_{k\geq 0} k\sum_{l\geq 0}\mathbb{P}(X=k \text{ et } Y=l) + \sum_{l\geq 0} l\sum_{k\geq 0}\mathbb{P}(X=k \text{ et } Y=l).\end{aligned}$$

On a utilisé plusieurs fois la règle d'interversion rappelée à l'exercice précédent, et on a fait un changement d'indice $l = n-k$ à l'avant-dernière ligne. Dans le premier terme de la dernière ligne, pour tout $k\geq 0$, la somme sur l est la somme des probabilités d'événements deux à deux disjoints dont la réunion vaut $\{X = k\}$. De même, dans le deuxième terme, pour tout $l \geq 0$, la somme sur k est la somme des probabilités d'événements deux à deux disjoints dont la réunion vaut $\{Y = l\}$. Ainsi, on trouve

$$\sum_{n \geq 0} n\mathbb{P}(X + Y = n) = \sum_{k \geq 0} k\mathbb{P}(X = k) + \sum_{l \geq 0} l\mathbb{P}(Y = l),$$

ce qui, au nom des indices près, est la formule voulue.

6. Soit $(\Omega, \mathscr{F}, \mathbb{P})$ un espace de probabilités. Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire telle que $X(\Omega) \subset \mathbb{N}$.

a) Montrer que pour tout réel $s \in [0, 1]$, la fonction s^X est une variable aléatoire intégrable. *On rappelle que par convention,* $0^0 = 1$.

On appelle *fonction génératrice* de X la fonction

$$\begin{aligned} G_X : [0,1] &\longrightarrow \mathbb{R} \\ s &\longmapsto \mathbb{E}[s^X]. \end{aligned}$$

b) Montrer que G_X est une fonction positive croissante. Calculer ses valeurs en 0 et en 1.

c) Calculer la fonction G_X lorsque X suit l'une des lois suivantes :

(i) Bernoulli de paramètre $p \in [0, 1]$,

(ii) binomiale de paramètres $n \geq 0$ et $p \in [0, 1]$,

(iii) géométrique de paramètre $p \in]0, 1[$,

(iv) Poisson de paramètre $\lambda > 0$.

d) Que peut-on dire de deux variables aléatoires à valeurs entières qui ont la même fonction génératrice ?

Solution de l'exercice 6.

a) Soit $s \in (0, 1]$. L'hypothèse $X(\Omega) \subset \mathbb{N}$ est inutile pour la mesurabilité, et pour l'intégrabilité on peut se contenter de $X(\Omega) \subset \mathbb{R}^+$. Comme la fonction $x \in \mathbb{R} \mapsto s^x \in \mathbb{R}$ est continue, le fait que X est une variables aléatoire entraine que s^X aussi (proposition 2.2.2b du poly). Comme elle est bornée (par 1), il s'agit d'une variable aléatoire intégrable. Pour $s = 0$, on a $s^X = \mathbb{1}_{\{X=0\}}$, qui est aussi une variable aléatoire, elle aussi bornée donc intégrable.

b) Soient s et t des réels tels que $0 \leq s \leq t \leq 1$. Pour tout $\omega \in \Omega$, on a $0 \leq s^{X(\omega)} \leq t^{X(\omega)} \leq 1$. Par positivité de l'espérance (théorème 2.3.1b), on en déduit que $0 \leq G_X(s) \leq G_X(t) \leq 1$. G_X est donc positive et croissante. Enfin, $G_X(0) = \mathbb{P}(X = 0)$ et $G_X(1) = 1$.

c) Soit X une variable aléatoire de loi :

(i) Bernoulli de paramètre $p \in [0,1]$. La fonction génératrice de X est

$$G_X(s) = \mathbb{E}[s^X] = s^0\mathbb{P}(X=0) + s^1\mathbb{P}(X=1) = 1-p+ps.$$

(ii) binomiale de paramètres $n \geq 0$ et $p \in [0,1]$. La fonction génératrice de X est

$$G_X(s) = \sum_{k=0}^{n} \binom{n}{k} s^k p^k (1-p)^{n-k} = (1-p+sp)^n.$$

(iii) géométrique de paramètre $p \in]0,1[$. La fonction génératrice de X est

$$G_X(s) = (1-p)\sum_{n\geq 0} s^n p^n = \frac{1-p}{1-sp}.$$

Notons que son rayon de convergence est égal à $\frac{1}{p}$.

(iv) Poisson de paramètre $\lambda > 0$. La fonction génératrice de X est

$$G_X(s) = e^{-\lambda}\sum_{n\geq 0} s^n \frac{\lambda^n}{n!} = e^{-\lambda+s\lambda} = e^{\lambda(1-s)}.$$

d) Deux variables aléatoires à valeurs entières qui ont la même fonction génératrice ont même loi (et la réciproque est triviale). Pour le voir, on remarque que ces fonctions sont C^∞ et que la dérivée d'ordre n de G_X est $G_X^{(n)}(s) = \mathbb{E}\left[\frac{X!}{(X-n)!}s^{X-n}\mathbb{1}_{\{X\geq n\}}\right]$. En particulier, $G_X^{(n)}(0) = n!\mathbb{P}(X=n)$, ce qui permet de retrouver la loi à partir de la fonction génératrice.

7. Soit $X : (\Omega, \mathscr{F}, \mathbb{P}) \to \mathbb{R}$ une variable aléatoire. Soit $a > 0$ un nombre réel.

a) Montrer que $\mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}(|X|)}{a}$.

b) Que peut-on dire de la proportion de la population qui gagne plus de dix fois le salaire moyen ?

L'inégalité démontrée au a) s'appelle l'*inégalité de Markov.*

Solution de l'exercice 7.

a) On vérifie facilement que pour tout $\omega \in \Omega$, $a\mathbb{1}_{\{|X(\omega)|\geq a\}} \leq |X(\omega)|$. Par positivité de l'espérance, on obtient en intégrant :

$$a\mathbb{P}(|X| \geq a) \leq \mathbb{E}(|X|).$$

Il suffit de diviser par a pour obtenir l'inégalité demandée.

b) En considérant que X est le salaire, et en choisissant $a = 10\mathbb{E}(|X|)$, on déduit de l'inégalité de Markov que la proportion de la population qui gagne plus de dix fois le salaire moyen est inférieure à $1/10$.

8. Soit Ω un ensemble. Soit $f : \Omega \to \mathbb{R}$ une fonction. Montrer qu'il existe une unique paire de fonctions (g, h) de Ω dans $\mathbb{R}_+$ telles que $f = g - h$ et qu'on ait à la fois $g \leq |f|$ et $h \leq |f|$. Montrer que si $(\tilde{g}, \tilde{h})$ est une autre paire de fonctions positives telles que $f = \tilde{g} - \tilde{h}$, alors on a $g \leq \tilde{g}$ et $h \leq \tilde{h}$.

Solution de l'exercice 8. Soit x un nombre réel. Soit $x = a - b$ une écriture de x comme différence de deux nombres positifs. tels que $a \leq |x|$ et $b \leq |x|$. Si $x \geq 0$, on a $a = x + b \geq x = |x|$, donc $a = |x| = x$ et $b = 0$. Si $x \leq 0$, on a $b = a - x \geq -x = |x|$, donc $b = |x|$ et $a = 0$. Ces deux cas sont pris en compte par les formules $a = \max(x, 0)$ et $b = \max(-x, 0)$. Les nombres a et b ainsi définis ont bien les propriétés voulues. De plus, si $\tilde{a}$ et $\tilde{b}$ sont deux nombres positifs dont la différence vaut x, alors le même raisonnement que ci-dessus montre que $\tilde{a} \geq a$ et $\tilde{b} \geq b$.

On note $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$.

Les fonctions g et h qui à tout $\omega \in \Omega$ associent respectivement $g(\omega) = f(\omega)^+$ et $h(\omega) = f(\omega)^-$ ont toutes les propriétés voulues.

9. On effectue des lancers successifs et indépendants d'une pièce qui tombe sur *pile* avec probabilité p et sur *face* avec probabilité $1 - p$.

a) Décrire le modèle probabiliste utilisé pour modéliser cette situation.

b) On appelle T_1 le numéro du premier lancer où l'on obtient *pile*. Déterminer la loi de T_1.

c) Pour tout $i \geq 1$, on appelle T_i le numéro du lancer où l'on obtient *pile* pour la i-ième fois. Déterminer la loi de T_i pour tout $i \geq 1$.

Solution de l'exercice 9.

a) ? ? ?

b) Soit $k \geq 1$ un entier. On a clairement $\{T_1 = k\} = \{X_1 = 0, X_2 = 0, \ldots, X_{k-1} = 0, X_k = 1\}$. Ainsi $\mathbb{P}(T_1 = k) = (1 - p)^{k-1}p$. Autrement dit, T_1 suit une loi de Bernouilli de paramètre p.

c) c'est encore pareil que dans l'exo5 de la feuille 2.