# Feuille d'exercices n$^o$26 : Études statistiques de variables aléatoires

**Exercice 1 [Minimum et maximum]**

On note $U, V$ les montants de la première et la seconde boule respectivement.

Pour le tirage avec remise : $U$ et $V$ sont indépendantes, chacune de loi uniforme sur $[\![1;n]\!]$, et donc :

$$\forall i,j \in [\![1;n]\!],\ \mathbb{P}((X,Y)=(i,j)) = \begin{cases} 0 & \text{si} \quad i<j \\ \dfrac{1}{n^2} & \text{si} \quad i=j \\ \dfrac{2}{n^2} & \text{si} \quad i>j \end{cases}$$

en notant l'égalité des événements :

$$((X,Y)=(i,j)) = \begin{cases} \emptyset & \text{si} \quad i<j \\ (U=i)\cap(V=j) & \text{si} \quad i=j \\ ((U=i)\cap(V=j))\cup((U=j)\cap(V=i)) & \text{si} \quad i>j \text{ (et cette union est disjointe)} \end{cases}$$

On conclut en sommant que :

$$\forall i \in [\![1;n]\!],\ \mathbb{P}(X=i) = \sum_{j=1}^{n} \mathbb{P}((X,Y)=(i,j)) = \frac{1}{n^2} + \frac{2(i-1)}{n^2} = \frac{2i-1}{n^2}$$

$$\forall j \in [\![1;n]\!],\ \mathbb{P}(Y=j) = \sum_{i=1}^{n} \mathbb{P}((X,Y)=(i,j)) = \frac{1}{n^2} + \frac{2(n-j)}{n^2} = \frac{2n-2j+1}{n^2}$$

et dans les deux cas on vérifie que la somme fait 1.

Pour le tirage sans remise : les montants sont distincts, $U$ suit une loi uniforme sur $[\![1;n]\!]$ et $V$ une loi uniforme sur $[\![1;n]\!]\setminus\{u\}$ (pour $u$ la valeur prise par $U$). On a alors :

$$\forall i,j \in [\![1;n]\!],\ \mathbb{P}((X,Y)=(i,j)) = \begin{cases} 0 & \text{si} \quad i\leqslant j \\ \dfrac{2}{n(n-1)} & \text{si} \quad i>j \end{cases}$$

en notant l'égalité des événements :

$$((X,Y)=(i,j)) = \begin{cases} \emptyset & \text{si} \quad i\leqslant j \\ ((U=i)\cap(V=j))\cup((U=j)\cap(V=i)) & \text{si} \quad i>j \text{ (et cette union est disjointe)} \end{cases}$$

On conclut en sommant que :

$$\forall i \in [\![1;n]\!],\ \mathbb{P}(X=i) = \sum_{j=1}^{n} \mathbb{P}((X,Y)=(i,j)) = \frac{2(i-1)}{n(n-1)}$$

$$\forall j \in [\![1;n]\!],\ \mathbb{P}(Y=j) = \sum_{i=1}^{n} \mathbb{P}((X,Y)=(i,j)) = \frac{2(n-j)}{n(n-1)}$$

et dans les deux cas on vérifie que la somme fait 1.

Dans les deux cas : on note une probabilité $\mathbb{P}(X=i)$ d'autant plus grande que $i$ est grand, et l'inverse pour $Y$, ce qui est rassurant comme $X$ étant le maximum, il a plus de chances d'être grand (et l'inverse pour $Y$).

Pour les espérances et les variances : on fait apparaître des sommes cubiques, quadratiques ou arithmétiques (par linéarité) et on déduit après simplification :

— pour un tirage avec remise :

$$E(X) = \frac{(n+1)(4n-1)}{6n},\ E(Y) = \frac{(n+1)(2n+1)}{6n} \text{ et } V(X) = V(Y) = \frac{(n+1)(n-1)(2n^2+1)}{36n^2}$$

— pour un tirage sans remise :

$$E(X) = \frac{2(n+1)}{3},\ E(Y) = \frac{n+1}{3} \text{ et } V(X) = V(Y) = \frac{(n+1)(n-2)}{18}$$

et on a dans les deux cas $E(X+Y) = n+1$ (rassurant car les montants des boules valent en moyenne $\frac{n+1}{2}$), et pour les variances on retrouve une variance nulle si $n = 1$ (avec remise) ou $n = 2$ (sans remise) : rassurant car alors il y a une seule possibilité pour $X$ et $Y$ (donc variance nulle). Et le fait de trouver des variances égales pour $X$ et $Y$ est rassurant comme on passe de $X$ à $Y$ en échangeant maximum et minimum, donc les mêmes dispersions.

**Exercice 2 [Minimum]**

On considère $X_1, \ldots, X_n$ variables aléatoires indépendantes qui suivent chacune une loi uniforme sur $[\![1;n]\!]$. On pose $M = \min(X_1, \ldots, X_n)$.

1. Notons déjà que $M(\Omega) = [\![1;n]\!]$.
   On a par définition d'un minimum pour tout $k \in [\![1;n]\!]$ :
   $$(M \geqslant k) \Leftrightarrow (\forall i \in [\![1;n]\!],\ X_i \geqslant k)$$
   et par indépendance on déduit :
   $$\mathbb{P}(M \geqslant k) = \prod_{i=1}^{n} \mathbb{P}(X_i \geqslant k) = \left(\frac{n+1-k}{n}\right)^n.$$
   Et la formule reste valable pour $k = n+1$.
   Mais on a également l'union disjointe : $(M \geqslant k) = (M = k) \cup (M \geqslant k+1)$. Et ainsi :
   $$\forall k \in [\![1;n]\!],\ \mathbb{P}(M = k) = \mathbb{P}(M \geqslant k) - \mathbb{P}(M \geqslant k+1).$$
   Et comme la formule est valable pour sur $[\![1;n+1]\!]$, pour $k \in [\![1;n]\!]$ on peut l'appliquer pour calculer $\mathbb{P}(M \geqslant k)$ et $\mathbb{P}(M \geqslant k+1)$. Ce qui donne :
   $$\mathbb{P}(M = k) = \left(\frac{n+1-k}{n}\right)^n - \left(\frac{n-k}{n}\right)^n = \left(1 - \frac{k-1}{n}\right)^n - \left(1 - \frac{k}{n}\right)^n.$$
2. L'événement $(\exists i \in [\![1;n]\!],\ X_i = 1)$ correspond à l'événement $(M = 1)$.
   Par la question précédente, on a donc :
   $$\mathbb{P}\left(\exists i \in [\![1;n]\!],\ X_i = 1\right) = 1 - \left(1 - \frac{1}{n}\right)^n$$
   mais on a aussi :
   $$\left(1 - \frac{1}{n}\right)^n = \exp\left(n \ln(1 - 1/n)\right) \leqslant \exp(n \cdot (-1/n)) = \frac{1}{e}$$
   par concavité de ln. Et finalement : $\mathbb{P}\left(\exists i \in [\![1;n]\!],\ X_i = 1\right) \geqslant 1 - \frac{1}{e}$.

**Exercice 3 [Tirage dans une urne]**

On a directement $X(\Omega) = [\![0;n]\!]$ et :

$$\forall k \in [\![0;n]\!],\ \mathbb{P}(X=k) = \frac{\binom{n}{k}\binom{n}{n-k}}{\binom{2n}{n}}.$$

Pour l'espérance et la variance, on utilise que :

$$\forall n,m,p \in \mathbb{N},\ \sum_{k=0}^{n} \binom{n}{k}\binom{m}{p-k} = \binom{n+m}{p}$$

qu'on peut montrer ou bien directement par combinatoire (on considère $n$ boules blanches et $m$ noires dans une urne, et on regarde tous les tirages possibles), ou bien en regardant le coefficient de degré $p$ de $(1+x)^n(1+x)^m = (1+x)^{n+m}$.

On déduit par formule du capitaine :

$$E(X) = \frac{1}{\binom{2n}{n}} \sum_{k=0}^{n} k\binom{n}{k}\binom{n}{n-k} = \frac{n}{\binom{2n}{n}} \sum_{k=0}^{n} \binom{n-1}{k-1}\binom{n}{n-k} = \frac{n\binom{2n-1}{n}}{\binom{2n}{n}} = \frac{n}{2}$$

qu'on peut aussi trouver en notant que $X$ et $n-X$ ont même loi (donc même espérance), et cette espérance commune $\alpha$ vérifie donc : $\alpha = n-\alpha$ donc $\alpha = E(X) = \dfrac{n}{2}$.

Pour la variance on procède de même :

$$E(X^2) = \frac{1}{\binom{2n}{n}} \sum_{k=0}^{n} \binom{n}{k}^2 k^2 = \frac{n^2}{\binom{2n}{n}} \sum_{k=0}^{n} \binom{n-1}{k-1}^2 = \frac{n^2\binom{2n-2}{n-1}}{\binom{2n}{n}} = \frac{n^3}{2(2n-1)}$$

puis : $V(X) = \dfrac{n^2}{4(2n-1)}$.

**Exercice 4 [Un problème de chaussures]**

1. La variable $X_i$ suit une loi de Bernoulli de paramètre $p = \mathbb{P}(X_i = 1) = \dfrac{\binom{2N-2}{2n-2}}{\binom{2N}{2n}} = \dfrac{n(2n-1)}{N(2N-1)}$.

   Son espérance vaut $p$.
2. On a par définition $X = \sum_{i=1}^{N} X_i$. Donc par linéarité :

$$E(X) = \sum_{i=1}^{N} E(X_i) = \frac{n(2n-1)}{2N-1}.$$

**Exercice 5 [Télescopage dans la variance]**

Comme $X$ est à valeurs dans $[\![0;N]\!]$, on a :

$$E(X) = \sum_{k=0}^{N} k\mathbb{P}(X=k) = \sum_{k=1}^{N} k\mathbb{P}(X=k)$$

mais pour $k \in [\![1;N]\!]$, on a par union disjointe :

$$\mathbb{P}(X=k) = \mathbb{P}(X>k-1) - \mathbb{P}(X>k)$$

qu'on réinjecte dans l'expression de l'espérance :

$$E(X) = \sum_{k=1}^{N} k \left(\mathbb{P}(X > k-1) - \mathbb{P}(X > k)\right) = \sum_{k=1}^{N} k\mathbb{P}(X > k-1) - \sum_{k=1}^{N} k\mathbb{P}(X > k)$$

$$= \sum_{k=0}^{N-1} (k+1)\mathbb{P}(X > k) - \sum_{k=1}^{N} k\mathbb{P}(X > k) = \mathbb{P}(X > 0) + \sum_{k=1}^{N-1} (k+1-k)\mathbb{P}(X > k) + N\mathbb{P}(X > N) = \sum_{k=0}^{N} \mathbb{P}(X > k)$$

en rajoutant dans la somme le terme $\mathbb{P}(X > 0)$ et en notant que $\mathbb{P}(X > N) = 0$.

**Exercice 6 [Auto-indépendance (le retour)]**

Pour $X$ une telle variable aléatoire : elle est décorrelée d'elle-même, donc de variance nulle, donc de loi certaine.

Et on vérifie facilement qu'une variable aléatoire de loi certaine est indépendante d'elle-même. Elle est même indépendante de toute variable aléatoire.

**Exercice 7 [Variable aléatoire et coefficients binomiaux]**

Notons $\alpha > 0$ tel que :

$$\forall k \in [\![0; n]\!],\ \mathbb{P}(X = k) = \alpha \binom{n}{k}.$$

En sommant toutes ces probabilités, on doit trouver 1 et donc $\alpha = \dfrac{1}{2^n}$, et donc la loi de $X$ est donnée par :

$$\forall k \in [\![0; n]\!],\ \mathbb{P}(X = k) = \frac{\binom{n}{k}}{2^n} = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k}$$

Donc $X$ suit une loi binomiale de paramètres $n$ et $p = 1/2$.

Son espérance est $np = \dfrac{n}{2}$ et sa variance est $np(1-p)$.

**Exercice 8 [Transfert dans l'inégalité de Markov]**

Comme $g$ est strictement croissante, on a l'égalité des événements :

$$(|X| \geqslant a) = (g(|X|) \geqslant g(a))$$

et, comme $a > 0$ et que $g$ est strictement croissante, on déduit $g(a) > g(0) \geqslant 0$ donc $g(a) > 0$. Et de plus la variable aléatoire $Y = g(|X|)$ est positive, donc on peut lui appliquer l'inégalité de Markov, qui donne directement :

$$\mathbb{P}\left(Y \geqslant g(a)\right) \leqslant \frac{E(Y)}{g(a)}$$

qui est l'inégalité voulue avec les remarques ci-dessus.

**Exercice 9 [Tirage jusqu'à épuisement]**

Dans le meilleur des cas, on ne tire que des boules blanches. Dans le pire des cas, on tire toutes les boules noires avant d'avoir sorti la dernière boule blanche. Tous les résultats intermédiaires sont possibles, et ainsi : $X(\Omega) = [\![n; 2n]\!]$.

Pour $k \in [\![0; n]\!]$, l'événement $X = n + k$ revient à tirer $k$ boules noires, et toutes les boules blanches, en terminant par une boule blanche. On a donc :

$$\binom{n+k-1}{k} n!n!$$

tirages correspondants : on veut que $k$ boules noires exactement soient parmi les $n+k-1$ première boules (les $n+k$ premières sauf la dernière). Et une fois ces places choisies, il y a $n!$ manières de choisir les boules noires, et pareil pour les blanches.

Comme en tout il y a $(2n)!$ tirages, on déduit :

$$\forall k \in [\![0;n]\!],\ \mathbb{P}(X=n+k) = \frac{\binom{n+k-1}{k}}{\binom{2n}{n}}.$$

Pour l'espérance, on a :

$$E(X) = \sum_{k=0}^{n} (n+k)\frac{\binom{n+k-1}{k}}{\binom{2n}{n}} = \frac{1}{\binom{2n}{n}} \sum_{k=0}^{n} n\binom{n+k}{k}$$

par formule du capitaine. Et on simplifie par triangle de Pascal généralisé :

$$\sum_{k=0}^{n} \binom{n+k}{k} = \binom{2n+1}{n}$$

qui donne : $E(X) = \dfrac{n(2n+1)}{n+1}$.

Pour la variance, on calcule $E(X^2)$ (on peut calculer $E(X(X+1))$ pour simplifier les calculs) et on déduit :

$$V(X) = \frac{n^2(2n+1)}{(n+2)(n+1)^2}.$$

**Exercice 10 [Intérêt d'un jeu d'argent]**

On calcule l'espérance de gain, et on veut savoir pour quelles valeurs de $p$ elle est positive.

Si on note $Y$ le nombre de fois que l'on tombe sur pile, et $X$ le gain, on a : $X = 3Y - 20$ et $Y$ suit une loi binomiale de paramètres 12 et $p$.

Donc $E(Y) = 12p$ puis $E(X) = 36p - 20$ et :

$$E(X) \geqslant 0 \Leftrightarrow 36p \geqslant 20 \Leftrightarrow p \geqslant \frac{20}{36} = \frac{5}{9}$$

donc le jeu est intéressant pour le joueur dès que $p \geqslant \dfrac{5}{9}$.

**Exercice 11 [Variables aléatoires indépendantes de même loi]**

On utilise le système complet d'événement associé à $X$ : par formule des probabilités totales :

$$\mathbb{P}(X \neq Y) = \sum_{i=1}^{n} \mathbb{P}(X=a_i)\mathbb{P}_{X=a_i}(Y \neq X) = \sum_{i=1}^{n} \mathbb{P}(X=a_i)\mathbb{P}_{X=a_i}(Y \neq a_i)$$

mais par définition des $a_i$ on a $\mathbb{P}(X=a_i) = p_i$. Et par indépendance : $\mathbb{P}_{X=a_i}(Y \neq a_i) = \mathbb{P}(Y \neq a_i) = 1 - p_i$.

On a bien la somme en réinjectant.

**Exercice 12 [Dés indépendants et loi uniforme]**

On veut montrer qu'on ne peut créer deux dés (éventuellement différents) à $n$ faces tels que la somme des montants du lancer indépendant des deux dés donne une loi uniforme sur $[\![2;2n]\!]$.

Par l'absurde, on suppose deux tels dés à disposition, on note $X$ et $Y$ les variables aléatoires des montants des dés, et on note :

$$\forall k \in [\![1;n]\!],\ \mathbb{P}(X=k) = p_k \text{ et } \mathbb{P}(Y=k) = q_k.$$

1. Le seul moyen de faire 2 est un double 1 ; le seul moyen de faire $2n$ est un double $n$. Et donc :

$$\mathbb{P}(X+Y=2)=\mathbb{P}(X=1,Y=1)=\mathbb{P}(X=1)\mathbb{P}(Y=1)=p_1q_1$$

$$\mathbb{P}(X+Y=2n)=\mathbb{P}(X=n,Y=n)=\mathbb{P}(X=n)\mathbb{P}(Y=n)=p_nq_n$$

par indépendance des lancers à chaque fois. Mais la loi de $X+Y$ étant uniforme sur $[\![2;2n]\!]$, on a aussi :

$$\mathbb{P}(X+Y=2)=\mathbb{P}(X+Y=2n)=\frac{1}{2n-1}.$$

Et finalement : $p_1q_1=p_nq_n=\dfrac{1}{2n-1}$.

2. On regarde l'événement $X+Y=n+1$. Par formule des probabilités totales associée au système complet d'événements associé à $X$, on a :

$$\mathbb{P}(X+Y=n+1)=\sum_{k=1}^{n}\mathbb{P}(X=k)\mathbb{P}_{X=k}(X+Y=n+1)=\sum_{k=1}^{n}\mathbb{P}(X=k)\mathbb{P}(Y=n+1-k)$$

en utilisant la définition de la probabilité conditionnelle, et l'indépendance de $X$ et $Y$.
Et finalement en ne gardant que les termes correspondant à $k=1$ et $k=n$ :

$$\frac{1}{2n-1}=\mathbb{P}(X+Y=n+1)\geqslant p_1q_n+p_nq_1$$

et ce sont bien deux termes différents qui apparaissent dans la somme comme $n\geqslant 2$.

3. En combinant les deux questions, on déduit :

$$\frac{p_1}{p_n}+\frac{p_n}{p_1}\ leq1$$

ce qui est impossible car pour tout $x\in\mathbb{R}_+^*$ on a : $x+\dfrac{1}{x}=\dfrac{x^2+1}{x}=2+\dfrac{(x-1)^2}{x}\geqslant 2$ (avec égalité si, et seulement si, $x=1$).

**Exercice 13 [Application de Bienaymé–Tchebychev]**

On considère une usine $A$ qui fabrique 100 pièces, dont chacune a 5% de chance (de manière indépendante) d'être défectueuse, et une usine $B$ qui en fabrique 400, de manière indépendante, et dont 10% sont défectueuses. On note $X$ (resp. $Y$) le nombre de pièces défectueuses fabriquées par $A$ (resp. $B$).

1. $X$ et $Y$ suivent des lois binomiales, de paramètres respectifs $n=100$ et $p=0.05$, et $m=400$ et $q=0.1$. On a donc :

$$E(X)=np=5,\ V(X)=np(1-p)=\frac{19}{4},\ E(Y)=mq=40,\ V(Y)=36.$$

Pour $Z$ : $Z(\Omega)=[\![0,500]\!]$ avec par formule des probabilités totales appliquée au système complet d'événements associé à $X$ :

$$\forall k\in[\![0,500]\!],\mathbb{P}(Z=k)=\sum_{l=0}^{100}\mathbb{P}(X=l)\mathbb{P}(Y=k-l)=\sum_{l=0}^{1}00\binom{n}{l}\binom{m}{k-l}p^l(1-p)^{n-l}q^{k-l}(1-q)^{m-k+l}$$

qui ne se simplifie pas tellement.
Mais par linéarité de l'espérance, et indépendance de $X$ et $Y$, on a tout de même :

$$E(Z)=E(X)+Z(Y)=45\text{ et }V(Z)=V(X)+V(Y)=\frac{163}{4}.$$

2. L'inégalité de Bienaymé–Tchebychev donne que, pour tout $a > 0$ :

$$\mathbb{P}(|Z - E(Z)| \geqslant a) \leqslant \frac{V(Z)}{a^2}$$

Et pour un tel $a$, on a :

$$|Z - E(Z)| \geqslant a \Leftrightarrow (Z - E(Z) \geqslant a \text{ ou } Z - E(Z) \leqslant -a)$$

donc par union disjointe :

$$\mathbb{P}(Z - E(Z) \geqslant a) = \mathbb{P}(|Z - E(Z)| \geqslant a) - \mathbb{P}(Z - E(Z) \leqslant -a) \leqslant \mathbb{P}(|Z - E(Z)| \geqslant a) \leqslant \frac{V(Z)}{a^2}.$$

Et finalement, pour $c > E(Z) = 45$, on a :

$$\mathbb{P}(Z \geqslant c) = \mathbb{P}(Z - E(Z) \geqslant c - 45) \leqslant \frac{V(Z)}{(c-45)^2}$$

et comme $V(Z) = \frac{163}{4}$, on cherche $c$ tel que : $\frac{V(Z)}{(c-45)^2} \leqslant 0.05$, c'est-à-dire :

$$c \geqslant \sqrt{\frac{V(Z)}{0.05}} + 45 = \sqrt{815} + 45 \simeq 73.4$$

donc $c = 74$ convient.