Série d'exercices n° 3. Probabilité conditionnelle, indépendance

Une étoile désigne un exercice important.

Probabilité conditionnelle

⋆ **Exercice 3.1.** *Faux positifs.* Une maladie M affecte une personne sur 1000 dans une population donnée. Un test sanguin permet de détecter cette maladie avec une fiabilité de 99% (lorsque cette maladie est effectivement présente). En revanche, pour un individu sain, la probabilité que le test soit positif est de $0,1\%$ (on dit que $0,1\%$ est le taux de faux positifs). Si un test est positif, quelle est la probabilité que l'individu soit réellement malade ?

Solution. On choisit comme espace des états Ω l'ensemble des individus de la population. On appelle M l'événement "l'individu est malade", et P l'événement "le test est positif". Alors, l'énoncé nous dit que

$$\mathbb{P}(M) = \frac{1}{1000}; \text{ d'où } \mathbb{P}(M^c) = \frac{999}{1000}; \quad \mathbb{P}(P \mid M) = \frac{99}{100}; \quad \mathbb{P}(P \mid M^c) = \frac{1}{1000}.$$

On cherche

$$\mathbb{P}(M \mid P) = \frac{\mathbb{P}(M \cap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P \mid M)\mathbb{P}(M)}{\mathbb{P}(P \mid M)\mathbb{P}(M) + \mathbb{P}(P \mid M^c)\mathbb{P}(M^c)} = \frac{990}{990 + 999} \approx 0,5.$$

Si le test est positif, la probabilité que l'individu soit réellement malade est $\approx 0,5$. Donc attention avant de tirer des conclusions trop hâtives.

Exercice 3.2. Une secrétaire donne n appels téléphoniques ($n \geq 1$ est fixé). A chacun de ces appels, la probabilité qu'elle parvienne à joindre son correspondant est p ($p \in]0,1[$ est fixé). On suppose que les résultats de tous ces appels sont indépendants. Après cette première série d'essais, elle tente, le lendemain de rappeler les correspondants qu'elle n'a pas réussi à joindre. Les hypothèses sur ses chances de réussite sont les mêmes. On note X le nombre de personnes jointes dès le premier jour et Y le nombre de personnes jointes l'un ou l'autre jour.

1. Quelle est la loi de X ?
2. Pour $h \leq k \leq n$, que vaut $\mathbb{P}(Y = k | X = h)$?
3. En déduire la loi de Y. Retrouver ce résultat par un argument direct.

Solution.

1. Chaque personne est jointe le premier jour avec probabilité p. Donc X est une binômiale $B(n,p)$ de paramètre n et p.
2. Sachant que $X = h$, il reste $n - h$ personnes à joindre le 2e jour. Chaque personne est jointe avec probabilité p le deuxème jour, donc le nombre de personnes jointes le 2e jour (qui est $Y - X$) suit une loi binômiale $B(n-h,p)$ de paramètre $n - h$ et p. Donc,

$$\mathbb{P}(Y = k \mid X = h) = \mathbb{P}(Y - X = k - h \mid X = h) = \binom{n-h}{k-h} p^{k-h}(1-p)^{n-k}.$$

3. On veut calculer $\mathbb{P}(Y = k)$. On discute selon les valeurs de X (on se restreint à $X \in \{0, \ldots, k\}$ car on doit avoir $X \leq Y$)

$$\mathbb{P}(Y = k) = \sum_{h=0}^{k} \mathbb{P}(Y = k, X = h) = \sum_{h=0}^{k} \mathbb{P}(Y = k \mid X = h)\mathbb{P}(X = h).$$

On a $\mathbb{P}(X=h)=\binom{n}{h}p^h(1-p)^{n-h}$ donc

$$\begin{aligned}\mathbb{P}(Y=k) &= \sum_{h=0}^{k}\binom{n-h}{k-h}p^{k-h}(1-p)^{n-k}\binom{n}{h}p^h(1-p)^{n-h}\\ &= p^k(1-p)^{2(n-k)}\sum_{h=0}^{k}\binom{n-h}{k-h}\binom{n}{h}(1-p)^{k-h}.\end{aligned}$$

On remarque que

$$\begin{aligned}\sum_{h=0}^{k}\binom{n-h}{k-h}\binom{n}{h}(1-p)^{k-h} &= \sum_{h=0}^{k}\frac{n!}{(n-k)!(k-h)!h!}(1-p)^{k-h}\\ &= \binom{n}{k}\sum_{h=0}^{k}\binom{n}{h}(1-p)^{k-h} = \binom{n}{k}(2-p)^k.\end{aligned}$$

On a donc $\mathbb{P}(Y=k)=\binom{n}{k}\left(p(2-p)\right)^k(1-p)^{2(n-k)}$. Donc Y suit une loi binômiale $B(n,p(2-p))$de paramètres n et $p(2-p)$. On aurait pu le voir directement : chaque personne est jointe (le 1er ou 2e jour) avec probabilité $p+p(1-p)$. En effet, une personne est jointe le 1er jour avec probabilité p et le 2e jour avec probabilité $(1-p)p$. Y suit donc une binômiale de paramètres n et $p+(1-p)p=p(2-p)$.

Indépendance

Exercice 3.3. Soit X et Y deux v.a. independantes, de loi de Poisson de paramètres respectifs λ_1 et λ_2. Calculer la loi de $X+Y$.

Solution. On calcule

$$\begin{aligned}\mathbb{P}(X+Y=n) &= \sum_{k=0}^{n}\mathbb{P}\big(X=k,Y=n-k\big) \overset{\text{indép.}}{=} \sum_{k=0}^{n}\Big(\frac{\lambda_1^k}{k!}e^{-\lambda_1}\Big)\times\Big(\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}\Big)\\ &= \frac{1}{n!}e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} = \frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}.\end{aligned}$$

Exercice 3.4. Soit n un entier, et soient X et Y deux variables aléatoires indépendantes de loi uniforme sur $\{1,\dots,n\}$:

$$\forall k\in\{1,\cdots,n\},\ \mathbb{P}(X=k)=\mathbb{P}(Y=k)=\frac{1}{n}.$$

Calculer $\mathbb{P}(X=Y)$ et $\mathbb{P}(X\geq Y)$. Déterminer la loi de $X-Y$.

Solution. L'événement $\{X=Y\}$ se décompose de la manière suivante : $\{X=Y\}=\cup_{k=1}^n\{X=k,Y=k\}$. De plus les événements $\{X=k,Y=k\}$, $k=1,\cdots,n$ sont disjoints et pour tout k, $\{X=k\}$ et $\{Y=k\}$ sont indépendants. On a donc : $\mathbb{P}(X=Y)=\sum_{k=1}^n\mathbb{P}(X=k)\mathbb{P}(Y=k)=1/n$. L'espace probabilisé Ω se décompose ainsi : $\Omega=\{X>Y\}\cup\{X<Y\}\cup\{X=Y\}$. Par symétrie, on a $\mathbb{P}(X>Y)=\mathbb{P}(X<Y)$, d'où l'équation : $2\mathbb{P}(X>Y)+\mathbb{P}(X=Y)=1$. On en déduit que $\mathbb{P}(X>Y)=(n-1)/2n$ et $\mathbb{P}(X\geq Y)=\mathbb{P}(X>Y)+\mathbb{P}(X=Y)=1/2+1/2n$. On aurait pu aussi calculer cette probabilité directement :

$$\mathbb{P}(X\geq Y)=\sum_{i=1}^{n}\mathbb{P}(Y=i,X\geq i)=\sum_{i=1}^{n}\mathbb{P}(Y=i,X\geq i)=\sum_{i=1}^{n}\mathbb{P}(Y=i)\mathbb{P}(X\geq i)$$

par l'indépendance de X et Y. On calcule que $\mathbb{P}(X \geq i) = \sum_{j=i}^{n} \mathbb{P}(X = j) = \frac{(n-i+1)}{n}$. On obtient donc (en utilisant aussi $\mathbb{P}(Y = i) = 1/n$)

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

$X-Y$ est une v.a. symétrique à valeurs dans l'ensemble $\{-(n-1), -(n-2), \cdots, -1, 0, 1, \cdots, (n-2), (n-1)\}$. Pour déterminer sa loi, il suffit de calculer $\mathbb{P}(X - Y = k)$, pour $k = 0, 1, \cdots, n-1$:

$$\begin{aligned} \mathbb{P}(X - Y = k) &= \sum_{i=1}^{n-k} \mathbb{P}(X = k+i, Y = i) \\ &= \sum_{i=1}^{n-k} \mathbb{P}(X = k+i)\mathbb{P}(Y = i) = \frac{n-k}{n^2}, \end{aligned}$$

où la seconde égalité vient de l'indépendance entre X et Y.

Exercice 3.5. Montrer qu'une v.a. X est indépendante d'elle-même si et seulement si elle est p.s. constante : a. en la supposant de carré intégrable et en calculant $\mathrm{Var}\,(X)$, b. plus généralement en déterminant sa fonction de répartition.

Solution.

a - Var(X)= $\mathbb{E}[X^2] - \mathbb{E}[X]^2 = 0$, si X est indépendante d'elle même. D'autre part Var(X)= $\mathbb{E}[(X - \mathbb{E}[X])^2]$, donc $X - \mathbb{E}[X] = 0$ p.s., et X est égale à $\mathbb{E}[X]$ p.s.

(*On a utilisé la propriété suivante : Si Y est une variable aléatoire* **positive** *et si $\mathbb{E}[Y] = 0$ alors $Y = 0$ p.s.*)

b - $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{E}[1_{X \leq x}] = \mathbb{E}[1^2_{X \leq x}] = \mathbb{E}[1_{X \leq x}] \times \mathbb{E}[1_{X \leq x}] = (F_X(x))^2$

Donc la fonction de répartition vaut soit 0 soit 1. C'est la fonction de répartition d'une mesure de Dirac, c'est-à-dire que X est constante p.s.

⋆ **Exercice 3.6.** *(Indépendance et indépendance deux à deux).* On suppose données, sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ deux variables de Bernoulli ε_1 et ε_2, indépendantes, à valeurs dans $\{-1, +1\}$ avec

$$\mathbb{P}(\varepsilon_i = +1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}, \ (i = 1,\ 2).$$

1. Montrer que la variable aléatoire $\varepsilon_1\varepsilon_2$ est indépendante d'une part de ε_1, et d'autre part de ε_2.
2. La variable aléatoire $\varepsilon_1\varepsilon_2$ est-elle indépendante du couple $(\varepsilon_1, \varepsilon_2)$?

Solution.

1. On vérifie que $\mathbb{P}(\varepsilon_1 = \eta) = \mathbb{P}(\varepsilon_2 = \eta) = \mathbb{P}(\varepsilon_1\varepsilon_2 = \eta) = 1/2$ pour $\eta = 1$ ou -1.On vérifie aussi que $\mathbb{P}(\varepsilon_1\varepsilon_2 = \eta_1, \varepsilon_1 = \eta_2) = 1/4$ pour $(\eta_1, \eta_2) = (1, 1); (1, -1); (-1, 1)$ ou $(-1; -1)$. Donc $\varepsilon_1\varepsilon_2$ est indépendante de ε_1. De même, $\varepsilon_1\varepsilon_2$ est indépendante de ε_2.
2. On remarque que $\mathbb{P}(\varepsilon_1\varepsilon_2 = 1, (\varepsilon_1, \varepsilon_2) = (-1, 1)) = 0$. Or $\mathbb{P}(\varepsilon_1\varepsilon_2 = 1) = 1/2$ et $\mathbb{P}((\varepsilon_1, \varepsilon_2) = (-1, 1)) = 1/4$ et $\frac{1}{2}\frac{1}{4} \neq 0$ donc $\varepsilon_1\varepsilon_2$ n'est pas indépendante de $(\varepsilon_1, \varepsilon_2)$ dans ce cas.

Exercice 3.7. Soient X, Y et Z trois variables aléatoires discrètes. Vrai ou faux ? (si vrai le prouver, si faux donner un contre exemple) :

1. Si X et Y sont indép., et si X et Z sont indép., alors X est indép. de (Y, Z).
2. Si (X, Y) et Z sont indép., alors Y est indép. de Z et X est indép. de Z.
3. Si X et Y sont indép. et (X, Y) est indép. de Z, alors X est indép. de (Y, Z).

Solution.

1. FAUX : contre-exemple prendre $\varepsilon_1, \varepsilon_2$ indépendantes telles que $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_2 = 1) = 1/2$ et $\mathbb{P}(\varepsilon_1 = -1) = \mathbb{P}(\varepsilon_2 = -1) = 1/2$ puis poser $Y = \varepsilon_1$, $Z = \varepsilon_2$ et $X = \varepsilon_1\varepsilon_2$. D'après l'exercice précédent, X et Y sont indépendantes, et X et Z sont indépendantes mais X et (Y, Z) ne sont pas indépendantes.
2. VRAI : En discutant sur toutes les valeurs y que peut prendre Y, on a $\mathbb{P}(X = x, Z = z) = \sum_y \mathbb{P}((X, Y) = (x, y), Z = z)$. Comme Z et (X, Y) sont indépendantes, on a $\mathbb{P}((X, Y) = (x, y), Z = z) = \mathbb{P}((X, Y) = (x, y))\mathbb{P}(Z = z)$ donc $\mathbb{P}(X = x, Z = z) = \sum_y \mathbb{P}((X, Y) = (x, y))\mathbb{P}(Z = z) = \mathbb{P}(X = x)\mathbb{P}(Z = z)$ car $\mathbb{P}(X = x) = \sum_y \mathbb{P}((X, Y) = (x, y))$.
3. VRAI : $\mathbb{P}(X = x, Y = y, Z = z) = \mathbb{P}(X = x, Y = y)\mathbb{P}(Z = z)$ car Z indépendante de (X, Y). Puis, $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ car X et Y sont indépendantes. Donc $\mathbb{P}(X = x, Y = y, Z = z) = \mathbb{P}(X = x)\mathbb{P}(Y = y)\mathbb{P}(Z = z) = \mathbb{P}(X = x)\mathbb{P}(Y = y, Z = z)$ car Z est indépendante de (X, Y), donc en particulier Z est indépendante de Y.

$\star$ **Exercice 3.8.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $p \in]0, 1[$. On pose

$$Y_n := X_n X_{n+1},\ S_n := X_1 + \cdots + X_n,\ V_n := Y_1 + \cdots + Y_n.$$

1. Calculer $\mathbb{E}(S_n)$, $\mathbb{E}(V_n)$.
2. Calculer $\mathrm{Var}(S_n)$, $\mathrm{Var}(V_n)$ et $\mathrm{Cov}(S_n, V_n)$.

Solution.

1. On a $\mathbb{E}(X_n) = \sum_{k\geq 0} k\mathbb{P}(X_n = k) = 0\times\mathbb{P}(X_n = 0) + 1\times\mathbb{P}(X_n = 1) = p$. De plus, $\mathbb{E}(Y_n) = \mathbb{E}(X_n)\mathbb{E}(X_{n+1}) = p^2$ puisque X_n et X_{n+1} sont indépendantes. Donc, par linéarité de l'espérance, $\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}[X_i] = np$ et $\mathbb{E}(V_n) = \sum_{i=1}^n \mathbb{E}[Y_i] = np^2$. (On aurait aussi pu remarquer que S_n suit une loi binômiale de paramètres n et p donc $\mathbb{E}(S_n) = np$).

2. Puisque les v.a. $(X_n)_{n\in\mathbb{N}}$ sont indépendantes et de même loi, on a $\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\mathrm{Var}(X_1) = np(1-p)$. (*Si U et V sont indépendantes, alors* $\mathrm{Var}(U + V) = \mathrm{Var}(U) + \mathrm{Var}(V)$).

   Les v.a. $(Y_n)_{n\in\mathbb{N}}$ ne sont pas indépendantes, mais on a

   $$\mathrm{Var}(V_n) = \mathrm{Var}(Y_1 + \cdots + Y_n) = \sum_{i=1}^n \mathrm{Var}(Y_i) + 2\sum_{1\leq i<j\leq n} \mathrm{Cov}(Y_i, Y_j).$$

   Or $\mathrm{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(X_i^2)\mathbb{E}(X_{i+1}^2) - p^4 = p^2 - p^4 = p^2(1-p^2)$ puisque X_i et X_{i+1} sont indépendantes. Si $j > i+1$, on a Y_i et Y_j qui sont indépendantes (en effet, on a alors (X_i, X_{i+1}) et (X_j, X_{j+1}) qui sont indépendantes donc X_iX_{i+1} et X_jX_{j+1} sont indépendantes). Ainsi, si $j > i+1$, $\mathrm{Cov}(Y_i, Y_j) = 0$ (*si X et Y sont indépendantes alors* $\mathrm{Cov}(X, Y) = 0$*, mais la réciproque est fausse !*). Si $j = i+1$, on a

   $$\mathrm{Cov}(Y_i, Y_{i+1}) = \mathbb{E}(X_iX_{i+1}^2X_{i+2}) - \mathbb{E}(Y_i)\mathbb{E}(Y_{i+1}) = p^3 - p^4 = p^3(1-p).$$

   Donc

   $$\mathrm{Var}(V_n) = np^2(1-p^2) + 2(n-1)p^3(1-p) = p^2(1-p)(n + 3np - 2p).$$

   Par bilinéarité de la covariance, on a $\mathrm{Cov}(S_n, V_n) = \sum_{1\leq i,j\leq n} \mathrm{Cov}(X_i, Y_j)$. Or $\mathrm{Cov}(X_i, Y_j) = 0$ pour j différent de i et de $i-1$ (car dans ce cas X_i et Y_j sont indépendantes) et $\mathrm{Cov}(X_i, Y_i) = \mathbb{E}(X_i^2X_{i+1}) - \mathbb{E}(X_i)\mathbb{E}(Y_i) = p^2(1-p)$ et $\mathrm{Cov}(X_i, Y_{i-1}) = \mathbb{E}(X_i^2X_{i-1}) - \mathbb{E}(X_i)\mathbb{E}(Y_{i-1}) = p^2(1-p)$. Donc

   $$\mathrm{Cov}(S_n, V_n) = \sum_{i=1}^n \mathrm{Cov}(X_i, Y_i) + \sum_{i=2}^n \mathrm{Cov}(X_i, Y_{i-1}) = (2n-1)p^2(1-p).$$

Exercice 3.9. Sur un espace de probabilité Ω on se donne une suite $(X_n)_{n\geq 1}$ de variables aléatoires de Bernoulli de paramètre p, $(0 < p < 1)$, indépendantes.

1. Soit $A_n = \{\omega \in \Omega : X_n(\omega) \neq X_{n-1}(\omega)\}$, $n \geq 2$. Calculer $\mathbb{P}(A_n \cap A_{n+1})$ pour $n \geq 2$. Donner une condition nécessaire et suffisante pour que les A_n soient indépendants.
2. Soit $\nu(\omega) = \inf\{n \geq 2 : \omega \in A_n\}$, avec $\inf \emptyset = +\infty$. Montrer que ν est une variable aléatoire. Quelle est la loi de ν ? Montrer que $\mathbb{P}(\nu = +\infty) = 0$.

Solution.

1. L'intersection de A_n avec A_{n+1} se décompose ainsi :

$$\begin{aligned} A_n \cap A_{n+1} &= \{X_n \neq X_{n-1}\} \cap \{X_n \neq X_{n+1}\} \\ &= \{X_{n-1} = 0, X_n = 1, X_{n+1} = 0\} \cup \{X_{n-1} = 1, X_n = 0, X_{n+1} = 1\}, \end{aligned}$$

où les deux événements de la dernière égalité sont disjoints. Les v.a. X_n, $n \geq 1$ étant indépendantes, on a $P(A_n \cap A_{n+1}) = (1-p)^2 p + p^2(1-p) = p(1-p)$. Si les A_n, $n \geq 1$ sont indépendants, alors on a nécessairement $P(A_n \cap A_{n+1}) = P(A_n)P(A_{n+1})$. Avec

$$\begin{aligned} P(A_n) &= P(X_n \neq X_{n-1}) \\ &= P(X_{n-1} = 0, X_n = 1) + P(X_{n-1} = 1, X_n = 0) \\ &= 2p(1-p), \end{aligned}$$

ceci entraîne que $p(1-p) = 4p^2(1-p)^2$, soit $p = 1/2$.
Montrons que cette condition est suffisante. Supposons donc $p = 1/2$ de sorte que $\mathbb{P}(X_1 = 0) = \mathbb{P}(X_1 = 1) = 1/2$ et $\mathbb{P}(A_1) = 1/2$. On rappelle que des évènements $(A_i, i \in I)$ sont indépendants si et seulement si pour toute famille finie d'indices $n_1 < n_2 < \ldots < n_k$ de I, on a

$$\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k}) = \prod_{i=1}^{k} \mathbb{P}(A_{n_i}).$$

On veut donc montrer que pour toute suite d'entiers $1 \leq n_1 < n_2 < \ldots < n_k$,

$$\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k}) = \left(\frac{1}{2}\right)^k.$$

On le montre par récurrence sur k. Si $k = 1$, l'identité est triviale car $\mathbb{P}(A_n) = 1/2$ pour tout n. Supposons que l'identité est vraie au rang k. Montrons qu'elle est vraie au rang $k+1$. On a

$$\begin{aligned} \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_{k+1}}) &= \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} \neq X_{n_{k+1}+1}\}) \qquad (2) \\ &= \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\} \cap \{X_{n_{k+1}+1} = 1\}) \\ &\quad + \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 1\} \cap \{X_{n_{k+1}+1} = 0\}). \end{aligned}$$

[**Notation**. On note $\sigma(Y)$ la tribu engendrée par la variable aléatoire Y.] Comme la suite $(X_\ell)_{\ell\geq 1}$ est indépendante, la tribu $\sigma(X_1, \ldots, X_{n_{k+1}})$ et la tribu $\sigma(X_{n_{k+1}+1})$ sont indépendantes (Ceci veut dire que tout évènement de la tribu $\sigma(X_1, \ldots, X_{n_{k+1}})$ et tout évènement de la tribu $\sigma(X_{n_{k+1}+1})$ sont indépendants). Or l'évènement $A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\}$ appartient à $\sigma(X_1, \ldots, X_{n_{k+1}})$ tandis que $\{X_{n_{k+1}+1} = 1\} \in \sigma(X_{n_{k+1}+1})$. Ainsi

$$\begin{aligned} &\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\} \cap \{X_{n_{k+1}+1} = 1\}) \\ &\qquad = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\})\mathbb{P}(X_{n_{k+1}+1} = 1) \\ &\qquad = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\})\frac{1}{2}. \end{aligned}$$

De même,

$$
\begin{aligned}
&\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 1\} \cap \{X_{n_{k+1}+1} = 0\}) \\
&\qquad = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 1\})\mathbb{P}(X_{n_{k+1}+1} = 0) \\
&\qquad = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 1\})\frac{1}{2}.
\end{aligned}
$$

Donc, en revenant à (2), on obtient

$$\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_{k+1}}) = \big(\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\}) + \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 1\})\big)\,\frac{1}{2}.$$

On remarque que

$$\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 0\}) + \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k} \cap \{X_{n_{k+1}} = 1\}) = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k}).$$

Donc,

$$\mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_{k+1}}) = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_k})\frac{1}{2}.$$

Il suffit d'appliquer l'hypothèse de récurrence pour terminer la récurrence.

2. Pour tout $n \geq 2$, $\{\nu = n\} = \cap_{i=1}^{n-2}\{X_i = X_{i+1}\} \cap \{X_{n-1} \neq X_n\}$. Puisque les X_n sont des v.a., les ensembles $\{X_i = X_{i+1}\}$, $i \leq n$ et $\{X_{n-1} \neq X_n\}$ appartiennent tous à la tribu $\mathcal{F}$, c'est donc aussi le cas de $\{\nu = n\}$. Ceci montre que ν est une variable aléatoire.

   Pour tout $n \geq 2$, on a

$$
\begin{aligned}
\mathbb{P}(\nu = n) &= P(X_1 = X_2 \cdots = X_{n-1} = 1)\mathbb{P}(X_n = 0) + \\
&\quad\ \mathbb{P}(X_1 = X_2 \cdots = X_{n-1} = 0)\mathbb{P}(X_n = 1) \\
&= (1-p)^{n-1}p + p^{n-1}(1-p)\,,
\end{aligned}
$$

   ce qui caractérise la loi de ν. Calculons la probabilité de l'événement complémentaire : $\{\nu < +\infty\}$.

$$
\begin{aligned}
\mathbb{P}(\nu < +\infty) &= \mathbb{P}(\cup_{n=2}^{\infty}\{\nu = n\}) \\
&= \sum_{n=2}^{\infty}((1-p)^{n-1}p + p^{n-1}(1-p)) \\
&= p\left(\frac{1}{1-(1-p)} - 1\right) + (1-p)\left(\frac{1}{1-p} - 1\right) = 1\,,
\end{aligned}
$$

   d'où $\mathbb{P}(\nu = +\infty) = 0$).

Fonctions génératrices

Exercice 3.10. *(Un rappel ?)* Soient X et Y deux variables aléatoires indépendantes, suivant une loi de Poisson, de paramètres respectifs λ_1 et λ_2.

1. Calculer la fonction génératrice de X.
2. Calculer la fonction génératrice de $X + Y$. Qu'en déduisez-vous ?

Solution.

1. On écrit

$$\mathbb{E}[r^X] = \sum_{k \geq 0} r^k \mathbb{P}(X = k) = \sum_{k \geq 0} r^k \frac{\lambda_1^k}{k!} e^{-\lambda_1} = e^{r\lambda_1}e^{-\lambda_1}.$$

2. Par indépendance, on a $\mathbb{E}[r^{X+Y}] = \mathbb{E}[r^X]\mathbb{E}[r^Y] = e^{r\lambda_1}e^{-\lambda_1}e^{r\lambda_2}e^{-\lambda_2} = e^{(r-1)(\lambda_1+\lambda_2)}$. On reconnaît la fonction génératrice d'une loi de Poisson de paramètre $\lambda_1 + \lambda_2$. Comme la fonction génératrice caractérise la loi d'une variable aléatoire, on obtient que $X + Y$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

⋆ **Exercice 3.11.** Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes et N une variable aléatoire à valeurs entières indépendante de la suite (X_n). On définit S_N sur Ω par $S_N(\omega) = 0$ si $N(\omega) = 0$ et $S_N(\omega) = \sum_{n=1}^{N(\omega)} X_n(\omega)$ si $N(\omega) \geq 1$.

1. Montrer que S_N est une variable aléatoire.
2. On suppose que les X_n sont à valeurs entières et ont même loi. Déterminer la fonction génératrice de S_N en fonction de celle de N et de X_1.
3. En déduire l'espérance et la variance de S_N.
4. Trouver la loi de S_N lorsque les X_n suivent une loi de Bernoulli de paramètre $p \in]0, 1[$ et que N suit une loi géométrique de paramètre $a \in]0, 1[$.

Solution.

1. On a, pour tout borélien A de $\mathbb{R}$ les égalités suivantes :

$$\{\omega\,; S_N(\omega) \in A\} = \bigcup_{k\in\mathbb{N}}\{\omega\,; \sum_{i=1}^{N(\omega)} X_i(\omega) \in A \text{ et } N(\omega) = k\} = \bigcup_{k\in\mathbb{N}}\{\omega\,; \sum_{i=1}^{k} X_i(\omega) \in A \text{ et } N(\omega) = k\}.$$

Puisque N et $\sum_{i=1}^k X_i$ sont des variables aléatoires, les événements $\{\sum_{i=1}^k X_i \in A\}$ et $\{N = k\}$ sont dans la tribu $\mathcal{F}$; leur intersection y est donc aussi et la réunion dénombrable d'éléments de $\mathcal{F}$ est dans $\mathcal{F}$. On en déduit que, pour tout borélien A de $\mathbb{R}$ l'événement $\{S_N \in A\}$ est dans $\mathcal{F}$, ce qui prouve que S_N est une variable aléatoire.

2. Par définition, la fonction génératrice de S_N est, pour $s \in [0, 1]$,

$$G_{S_N}(s) = \mathbb{E}(s^{S_N})$$

On a $1 = \sum_{k\geq 0} \mathbf{1}_{\{N=k\}}$ donc

$$G_{S_N}(s) = \mathbb{E}(\sum_{k\geq 0} \mathbf{1}_{\{N=k\}} s^{S_N}) = \sum_{k\geq 0} \mathbb{E}(\mathbf{1}_{\{N=k\}} s^{S_N}) = s^0\mathbb{P}(N = 0) + \sum_{k\geq 1} \mathbb{E}(\mathbf{1}_{\{N=k\}} s^{\sum_{i=1}^k X_i}).$$

Puisque N et la suite (X_i) sont indépendantes, on a

$$G_{S_N}(s) = \mathbb{P}(N = 0) + \sum_{k\geq 1} \mathbb{E}\left(s^{\sum_{i=1}^k X_i}\right)\mathbb{P}(N = k).$$

Enfin, puisque les X_i sont indépendantes et de même loi, on a

$$\mathbb{E}\left(s^{\sum_{i=1}^k X_i}\right) = \prod_{i=1}^k G_{X_i}(s) = \big(G_{X_1}(s)\big)^k.$$

Ceci vaut 1 pour $k = 0$. Donc

$$G_{S_N}(s) = \sum_{k\geq 0}\big(G_{X_1}(s)\big)^k\mathbb{P}(N = k) = G_N\big(G_{X_1}(s)\big) = G_N \circ G_{X_1}(s).$$

3. On a $G'_{S_N}(1) = \mathbb{E}(S_N)$ et $G''_{S_N}(1) = \mathbb{E}(S_N(S_N-1))$. On déduit que $\mathbb{E}(S_N) = G'_{X_1}(1)G'_N(G_{X_1}(1)) = G'_{X_1}(1)G'_N(1) = \mathbb{E}(X_1)\mathbb{E}(N)$. (on a utilisé que $G_Y(1) = 1$ pour toute fonction génératrice G_Y). De même,

$$G''_{S_N}(1) = G'_N\big(G_{X_1}(1)\big)G''_{X_1}(1) + G''_N\big(G_{X_1}(1)\big)\big(G'_{X_1}(1)\big)^2 = \mathbb{E}(N)\mathbb{E}\big(X_1(X_1-1)\big) + \mathbb{E}\big(N(N-1)\big)\mathbb{E}(X_1)^2.$$

On en déduit, puisque $\mathbb{E}(S_N) = \mathbb{E}(N)\mathbb{E}(X_1)$, que $\mathbb{E}(S_N^2) = \mathbb{E}(N)\mathbb{E}\big(X_1(X_1-1)\big) + \mathbb{E}(N^2)\mathbb{E}(X_1)^2$. D'où

$$\mathrm{Var}(S_N) = \mathbb{E}(N)\mathrm{Var}(X_1) + \mathrm{Var}(N)\mathbb{E}(X_1)^2.$$

4. Si les X_n suivent une loi de Bernoulli de paramètre p, on a $G_{X_1}(s) = 1-p+ps$. Si N suit une loi géométrique de paramètre a, on a $G_N(s) = \frac{1-a}{1-as}$. Donc

$$G_{S_N}(s) = G_N \circ G_{X_1}(s) = \frac{1-a}{1-a(1-p+ps)} = \frac{\frac{1-a}{1-a(1-p)}}{1-\frac{ap}{1-a(1-p)}s}.$$

On en déduit, puisque la fonction génératrice caractérise la loi, que S_N suit une loi géométrique de paramètre $\frac{ap}{1-a(1-p)}$.

Exercice 3.12. Soient X et Y deux variables aléatoires indépendantes. X suit une loi de Bernoulli de paramètre p, $(0 < p < 1)$ et Y suit une loi de Poisson de paramètre λ, $\lambda > 0$. Soit Z la variable aléatoire égale à 0 si $X = 0$ et à Y si $X = 1$.

1. Calculer la loi de Z.
2. Quelle est la fonction génératrice de Z, son espérance et sa variance ?
3. Que vaut la probabilité conditionnelle de $X = 0$, respectivement, $X = 1$, sachant que $Z = 0$?

Solution.

1. On a $\mathbb{P}(Z=k) = \mathbb{P}(Z=k, X=1) + \mathbb{P}(Z=k, X=0) = \mathbb{P}(Y=k, X=1) + \mathbb{P}(0=k, X=0)$. On en déduit que si $k \geq 1$, $\mathbb{P}(Z=k) = pe^{-\lambda}\frac{\lambda^k}{k!}$ et si $k=0$, $\mathbb{P}(Z=0) = 1-p+pe^{-\lambda}$.
2. La fonction génératrice de Z est

$$G_Z(s) = \sum_{k=0}^{\infty} s^k \mathbb{P}(Z=k) = 1-p+pe^{-\lambda} + pe^{-\lambda}\sum_{k=1}^{\infty}\frac{(\lambda s)^k}{k!} = 1-p+pe^{\lambda(s-1)}.$$

   On peut obtenir les moments de Z à l'aide des dérivées de sa fonction génératrice : comme $G_Z(s) = \mathbb{E}(s^Z)$, on a $\mathbb{E}(Z) = G_Z'(1)$ et $\mathbb{E}(Z(Z-1)) = G_z''(1)$. On trouve $\mathbb{E}(Z) = \lambda p$ et $\mathbb{E}(Z(Z-1)) = \lambda^2 p$ puis on en déduit $\mathrm{Var}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \lambda p + \lambda^2 p - \lambda^2 p^2$.
3. On a

$$\mathbb{P}(X=0 \mid Z=0) = \frac{\mathbb{P}(X=0, Z=0)}{\mathbb{P}(Z=0)} = \frac{\mathbb{P}(X=0)}{\mathbb{P}(Z=0)} = \frac{1-p}{1-p+pe^{-\lambda}}$$

   et

$$\mathbb{P}(X=1 \mid Z=0) = 1 - \mathbb{P}(X=0 \mid Z=0) = \frac{pe^{-\lambda}}{1-p+pe^{-\lambda}}.$$