TD7

Couples de variables aléatoires discrètes

Loi d'un couple, lois marginales et conditionnelles

Exercice 7.1 (★)

On considère un couple (X,Y) de variables aléatoires réelles à valeurs dans $\mathbb{N}$ pour lequel il existe un réel a tel que la loi de (X,Y) soit définie par : $\forall (i,j) \in \mathbb{N}^2,\ P([X=i]\cap[Y=j]) = \frac{1}{j!}\frac{a}{2^{i+j}}$.

1. Déterminer a.
2. Déterminer les lois marginales.
3. Les variables X et Y sont-elles indépendantes ?

Exercice 7.2 (★★)

On désigne par n un entier naturel non nul et on considère n urnes numérotées de 1 à n. On suppose que pour tout entier k compris entre 1 et n, l'urne numéro k contient k boules numérotées de 1 à k.
On choisit au hasard une urne parmi les n et dans cette urne on tire au hasard une boule.
On note X la variable aléatoire égale au numéro de l'urne choisie et N la variable aléatoire égale au numéro porté par la boule obtenue.

1. Reconnaitre la loi de X, ainsi que la loi de N sachant $[X=i]$ pour tout $i \in X(\Omega)$.
2. Déterminer la loi du couple (X,N).
3. En déduire la loi de N sous forme de somme.
4. Calculer l'espérance et la variance de N.

Exercice 7.3 (★★★ - QSP HEC 2014)

Soient X et Y deux variables aléatoires définies sur un même espace probabilisé $(\Omega,\mathscr{A},P)$, à valeurs dans $\mathbb{N}$ telles que :

$$\forall (i,j)\in\mathbb{N}^2, \quad P([X=i]\cap[Y=j]) = \frac{a}{(i+j+1)!}.$$

Déterminer le réel a. Les variables X et Y sont-elles indépendantes ?

> Le réel a doit être positif. De plus on doit calculer la série double $\sum P(X=i,Y=j)$ qui doit converger et dont la somme doit valoir 1. Pour cela, étant donné la présence de $i+j$ dans le terme général, on va procéder à une somme suivant les diagonales (comme tout est positif, inutile de mettre des valeurs absolues) :
>
> - Pour tout $n \in \mathbb{N}$, on a :
>
> $$\sum_{i+j=n} P(X=i,Y=j) = \sum_{i+j=n} \frac{a}{(n+1)!} = (n+1)\frac{a}{(n+1)!} = \frac{a}{n!}$$
>
> car, rappelons le, le cardinal de $I_n = \{(i,j)\in\mathbb{N}^2,\ i+j=n\}$ est $n+1$.
>
> - La série $\sum_{n\geq 0} \frac{a}{n!}$ converge car c'est une série exponentielle, et sa somme vaut ae^1.
>
> Par le théorème de sommation suivant les diagonales, la série double $\sum P(X=i,Y=j)$ converge et elle vaut ae^1. Ainsi on a $a=e^{-1}$.

Étudions à présent l'indépendance de X et Y. On a :

$$P(X=0)=\sum_{j=0}^{+\infty}P(X=0,Y=j)=\sum_{j=0}^{+\infty}\frac{e^{-1}}{(j+1)!}=e^{-1}(\sum_{j=1}^{+\infty}\frac{1}{j!})=e^{-1}(e-1)=1-e^{-1}.$$

De même on a $P(Y=0)=1-e^{-1}$, et on a $P(X=0,Y=0)=e^{-1}$. Ainsi on a pas

$$P(X=0,Y=0)=P(X=0)P(Y=0).$$

Donc les variables X et Y ne sont pas indépendantes.

Fonctions de deux variables discrètes

Exercice 7.4 (★)

Soient X et Y deux variables aléatoires indépendantes suivant la même loi de Poisson de paramètre $\lambda>0$.
Justifier que $\dfrac{X}{1+Y}$ possède une espérance et la calculer.

Exercice 7.5 (★★)

Soient X et Y deux variables indépendantes, définies sur le même espace probabilisé $(\Omega,\mathscr{A},P)$ telles que $X\hookrightarrow\mathscr{P}(\lambda)$ et $Y\hookrightarrow\mathscr{G}(p)$, $p\in]0,1[$.
Pour tout $\omega\in\Omega$, on définit une matrice $A(\omega)$ par :

$$A(\omega)=\begin{pmatrix}X(\omega) & Y(\omega)-1\\ Y(\omega)-1 & X(\omega)\end{pmatrix}.$$

Déterminer la probabilité que $A(\omega)$ soit inversible.

$A(\omega)$ n'est pas inversible si et seulement si $\det(A(\omega))=0$, ce qui équivaut à :

$$\begin{aligned}X(\omega)^2-(Y(\omega)-1)^2=0 &\Leftrightarrow (X(\omega)-Y(\omega)+1)(X(\omega)+Y(\omega)-1)=0\\ &\Leftrightarrow X(\omega)-Y(\omega)+1 \text{ ou } X(\omega)+Y(\omega)-1\\ &\Leftrightarrow X(\omega)=Y(\omega)-1 \text{ ou } X(\omega)=-Y(\omega)+1\end{aligned}$$

Or $X\geq 0$ et $-Y+1\leq 0$. Donc on a finalement que $A(\omega)$ est inversible si et seulement si

$$X(\omega)=Y(\omega)-1 \text{ ou } X(\omega)=0=-Y(\omega)+1\Leftrightarrow X(\omega)=Y(\omega)-1$$

Ainsi la probabilité pour que A ne soit pas inversible est égale à $P(X=Y-1)$. Reste à calculer cette probabilité. On utilise le système complet d'évènements $([Y=k])_{k\geq 1}$ et la formule des probabilités totales :

$$\begin{aligned}P(X=Y-1)&=\sum_{k=1}^{+\infty}P(X=Y-1,\ Y=k)=\sum_{k=1}^{+\infty}P(X=k-1,\ Y=k)\\ &=\sum_{k=1}^{+\infty}P(X=k-1)P(Y=k)\quad\text{car les variables sont indépendantes}\\ &=\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}(1-p)^{k-1}p=pe^{-\lambda}\sum_{k=1}^{+\infty}\frac{(\lambda(1-p))^{k-1}}{(k-1)!}=pe^{-\lambda}e^{\lambda(1-p)}=pe^{-\lambda p}\end{aligned}$$

Finalement, la probabilité que A soit inversible est donc égale à $1-pe^{-\lambda p}$.

Exercice 7.6 (★★)

Soient X et Y deux variables aléatoires indépendantes, telles que X suit la loi $\mathscr{P}(\lambda)$ et Y suit la loi $\mathscr{P}(\mu)$. Déterminer la loi de X sachant $[X+Y=n]$.

On a $X([X+Y=n]) = [\![0,n]\!]$. Pour tout $k \in [\![0,n]\!]$, on a :

$$\begin{aligned} P_{[X+Y=n]}(X=k) &= \frac{P([X=k]\cap[X+Y=n])}{P(X+Y=n)} = \frac{P([X=k]\cap[Y=n-k])}{P(X+Y=n)} \\ &= \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)} \end{aligned}$$

car X et Y sont indépendantes. De plus, par stabilité de la loi de Poisson, on a $X+Y \hookrightarrow \mathscr{P}(\lambda+\mu)$. On obtient donc :

$$P_{[X+Y=n]}(X=k) = \frac{\frac{\lambda^k}{k!}e^{-\lambda}\frac{\mu^{n-k}}{(n-k)!}e^{-\mu}}{\frac{(\lambda+\mu)^n}{n!}e^{-(\lambda+\mu)}} = \frac{n!}{k!(n-k)!}\frac{\lambda^k\mu^{n-k}}{(\lambda+\mu)^n} = \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k}$$

Ainsi la loi de X sachant $[X+Y=n]$ est une loi binomiale $\mathscr{B}(n, \frac{\lambda}{\lambda+\mu})$.

Exercice 7.7 (★★)

Soit X et Y deux variables aléatoires indépendantes d'un même espace probabilisé $(\Omega, \mathscr{A}, P)$ et suivant la même loi $\mathscr{G}(p)$.

1. Déterminer la loi de la variable aléatoire $X+Y$. Admet-elle une espérance ?
2. Déterminer la loi de la variable aléatoire $S = \sup(X,Y)$. Admet-elle une espérance ?
3. Déterminer la loi de la variable aléatoire $I = \inf(X,Y)$. Admet-elle une espérance ?
4. Montrer que la variable aléatoire SI admet une espérance et la calculer.

Exercice 7.8 (★★★ - Minimum, maximum de lois uniformes)

On considère une urne contenant n boules numérotées de 1 à n, et on effectue des tirages avec remise dans cette urne. On note X le numéro de la première boule extraite, et Y le numéro de la seconde. On note également I le plus petit des numéros tirés, et S le plus grand.

1. Donner la loi de I et de S.
2. Montrer que $E(S) = \dfrac{(n+1)(4n-1)}{6n}$, $E(I) = \dfrac{(n+1)(2n+1)}{6n}$, puis que $V(S) = V(I) = \dfrac{(n+1)(n-1)(2n^2+1)}{36n^2}$.
3. Déterminer une relation liant X, Y, S et I. En déduire $V(S+I)$, puis $\rho_{S,I} = \dfrac{n^2-1}{2n^2+1}$.

1. On a $X, Y \hookrightarrow \mathscr{U}([\![1,n]\!])$, donc $I(\Omega) = [\![1,n]\!]$.

   Pour tout $k \in [\![1,n]\!]$, on a $[I>k] = [X>k]\cap[Y>k]$ et donc :

   $$P(I>k) \underbrace{=}_{X,Y \text{ indép.}} P(X>k)P(Y>k) \underbrace{=}_{X,Y \text{ de même loi}} P(X>k)^2 = \left(\frac{n-k}{n}\right)^2.$$

   Notons que cette égalité est encore vraie pour $k=0$ car $P(I>0) = 1 = \left(\frac{n-0}{n}\right)^2$. On obtient alors que $k \in [\![1,n]\!]$:

   $$P(I=k) = P(I>k-1) - P(I>k) = \left(\frac{n-k+1}{n}\right)^2 - \left(\frac{n-k}{n}\right)^2 = \frac{2n-2k+1}{n^2}.$$

   La loi de S avait été déterminée en cours, on avait obtenu $S(\Omega) = [\![1,n]\!]$ et :

   $$\forall k \in [\![1,n]\!], \quad P(S=k) = \left(\frac{k}{n}\right)^2 - \left(\frac{k-1}{n}\right)^2 = \frac{2k-1}{n^2}.$$

2. I et S étant des variables aléatoires finies, elles admettent une espérance et une variance. Pour I, on a :

$$\begin{aligned}E(I) &= \sum_{k=1}^{n} kP(I=k) = \sum_{k=1}^{n} k\frac{2n-2k+1}{n^2} = \frac{2n+1}{n^2}\left(\sum_{k=1}^{n} k\right) - \frac{2}{n^2}\left(\sum_{k=1}^{n} k^2\right)\\ &= \frac{2n+1}{n^2}\frac{n(n+1)}{2} - \frac{2}{n^2}\frac{n(n+1)(2n+1)}{6}\\ &= \frac{3(n+1)(2n+1)-2(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6n}\end{aligned}$$

D'autre part, on a :

$$\begin{aligned}E(I^2) &= \sum_{k=1}^{n} k^2P(I=k) = \sum_{k=1}^{n} k\frac{2n-2k+1}{n^2} = \frac{2n+1}{n^2}\left(\sum_{k=1}^{n} k^2\right) - \frac{2}{n^2}\left(\sum_{k=1}^{n} k^3\right)\\ &= \frac{2n+1}{n^2}\frac{n(n+1)(2n+1)}{6} - \frac{2}{n^2}\frac{n^2(n+1)^2}{4}\\ &= \frac{6n(n+1)(2n+1)^2-18n^2(n+1)^2}{36n^2} = \frac{n(n+1)\left[6(4n^2+4n+1)-18(n^2+n)\right]}{36n^2}\\ &= \frac{n(n+1)\left[6n^2+6n+6\right]}{36n^2}\end{aligned}$$

Par la formule de Huygens, on obtient donc :

$$\begin{aligned}V(I) &= E(I^2)-E(I)^2 = \frac{n(n+1)\left[6n^2+6n+6\right]}{36n^2} - \frac{(n+1)^2(2n+1)^2}{36n^2}\\ &= (n+1)\frac{6n^3+6n^2+6n-4n^3-8n^2-5n-1}{36n^2} = (n+1)\frac{2n^3-2n^2+n-1}{36n^2}\\ &= \frac{(n+1)(n-1)(2n^2+1)}{36n^2}\end{aligned}$$

On procède de manière similaire pour S, je vous laisse le faire en exercice.

3. On a $S+I = X+Y$, et puisque X et Y sont indépendantes :

$$V(S+I) = V(X+Y) = V(X)+V(Y) = \frac{n^2-1}{12} + \frac{n^2-1}{12} = \frac{n^2-1}{6}.$$

On obtient alors la covariance :

$$\begin{aligned}\mathrm{Cov}(S,I) &= \frac{1}{2}\left(V(S+I)-V(S)-V(I)\right) = \frac{1}{2}\left(\frac{n^2-1}{6} - \frac{(n^2-1)(2n^2+1)}{18n^2}\right)\\ &= \frac{n^2-1}{2}\times\frac{3n^2-2n^2-1}{18n^2} = \frac{(n^2-1)^2}{36n^2}\end{aligned}$$

On en déduit que :

$$\rho_{I,S} = \frac{\mathrm{Cov}(I,S)}{\sqrt{V(I)V(S)}} = \frac{\mathrm{Cov}(I,S)}{V(I)} = \frac{n^2-1}{2n^2+1}.$$

Exercice 7.9 (★★★ - Somme de lois uniformes)
Soit $n \geq 1$, et X, Y des variables indépendantes suivant la loi $\mathscr{U}(\llbracket 1, n\rrbracket)$.
Déterminer la loi de $Z = X+Y$.
Pour le calcul $P(X+Y=k)$, on pourra distinguer les cas $k \leq n+1$ et $k > n+1$.

Exercice 7.10 (★★★ - QSP HEC 2012)
Les variables aléatoires sont définies sur un espace probabilisé $(\Omega, \mathscr{A}, P)$.
Soit X une variable aléatoire qui suit la loi de Poisson de paramètre $\lambda > 0$, et soit Y une variable aléatoire indépendante de X telle que : $Y(\Omega) = \{1,2\}$, $P(Y=1) = P(Y=2) = \dfrac{1}{2}$. On pose : $Z = XY$.

1. Déterminer la loi de Z.

2. Quelle est la probabilité que Z prenne des valeurs paires ?

1. On a $X(\Omega)=\mathbb{N}$ et $Y(\Omega)=\{1,2\}$, donc $Z(\Omega)=\mathbb{N}$. Soit $k\in\mathbb{N}$, on a deux cas possibles :
   - k est impair, soit de la forme $k=2p+1$ avec $p\in\mathbb{N}$: $([Y=1],[Y=2])$ est un système complet d'évènements. Par la formule des probabilités totales, on a :

$$\begin{aligned}P(Z=2p+1)&=P(Y=1,XY=2p+1)+P(Y=2,XY=2p+1)\\&=P(Y=1,X=2p+1)+P(Y=2,2X=2p+1)\\&=P(Y=1)P(X=2p+1)+P(Y=2)P(2X=2p+1)\qquad\text{par indépendance}\\&=\frac{1}{2}\frac{\lambda^{2p+1}}{(2p+1)!}e^{-\lambda}+\frac{1}{2}\times 0=\frac{1}{2}\frac{\lambda^k}{k!}e^{-\lambda}\end{aligned}$$

   - k est pair, soit de la forme $k=2p$ avec $p\in\mathbb{N}$: toujours avec le SCE $([Y=1],[Y=2])$, on a :

$$\begin{aligned}P(Z=2p)&=P(Y=1,XY=2p)+P(Y=2,XY=2p)=P(Y=1,X=2p)+P(Y=2,2X=2p)\\&=P(Y=1)P(X=2p)+P(Y=2)P(X=p)\qquad\text{par indépendance}\\&=\frac{1}{2}\frac{\lambda^{2p}}{(2p)!}e^{-\lambda}+\frac{1}{2}\frac{\lambda^p}{p!}e^{-\lambda}=\frac{e^{-\lambda}}{2}\left(\frac{\lambda^k}{k!}+\frac{\lambda^{k/2}}{(k/2)!}\right)\end{aligned}$$

2. On cherche la probabilité de $\cup_{p\in\mathbb{N}}[Z=2p]$. Puisque ces évènements sont incompatibles, cette probabilité est égale à :

$$\sum_{p=0}^{+\infty}P(Z=2p)=\frac{e^{-\lambda}}{2}\sum_{p=0}^{+\infty}\left(\frac{\lambda^{2p}}{(2p)!}+\frac{\lambda^p}{p!}\right)$$

La série $\sum_{p\geq 0}\frac{\lambda^p}{p!}$ converge et sa somme vaut e^{λ}.

On étudie l'autre série $\sum_{p\geq 0}\frac{\lambda^{2p}}{(2p)!}$: il s'agit de "la série exponentielle à laquelle on ne garde que les termes pairs". En d'autres termes, il s'agit de la partie paire de l'exponentielle. En effet on se souvient que toute fonction f de $\mathbb{R}$ dans $\mathbb{R}$ se décompose de manière unique en la somme d'une fonction paire $p:x\mapsto\frac{f(x)+f(-x)}{2}$ et d'une fonction impaire $i:x\mapsto\frac{f(x)-f(-x)}{2}$ (on peut le faire par Analyse-Synthèse). On est donc amené à considérer :

$$\frac{e^x+e^{-x}}{2}=\frac{1}{2}\sum_{k=0}^{+\infty}\left(\frac{x^k}{k!}+(-1)^k\frac{x^k}{k!}\right).$$

Or on a $\dfrac{\frac{x^k}{k!}+(-1)^k\frac{x^k}{k!}}{2}=\begin{cases}0 & \text{si } k \text{ impair}\\ \frac{x^k}{k!} & \text{si } k \text{ pair}\end{cases}$. On obtient donc que :

$$\frac{e^x+e^{-x}}{2}=\sum_{p=0}^{+\infty}\frac{\lambda^{2p}}{(2p)!}.$$

Ainsi la série $\sum_{p\geq 0}\frac{\lambda^{2p}}{(2p)!}$ converge et sa somme vaut $\frac{e^x+e^{-x}}{2}$.

Finalement, comme tout converge, on peut écrire :

$$\frac{e^{-\lambda}}{2}\left(\sum_{p=0}^{+\infty}\frac{\lambda^{2p}}{(2p)!}+\sum_{p=0}^{+\infty}\frac{\lambda^p}{p!}\right)=\frac{e^{-\lambda}}{2}\left(\frac{e^{\lambda}+e^{-\lambda}}{2}+e^{\lambda}\right)=\frac{3}{4}+\frac{e^{-2\lambda}}{4}$$

Ainsi la probabilité que Z soit paire est de $\frac{3}{4} + \frac{e^{-2\lambda}}{4}$.

Covariance, corrélation linéaire

Exercice 7.11 (★)
Soit un couple (X, Y) de variables aléatoires dont la loi conjointe est donnée par :

$X \backslash Y$	b_1	b_2	b_3
a_1	0	$\frac{1}{5}$	$\frac{1}{5}$
a_2	$\frac{1}{5}$	$\frac{1}{5}$	α

,

avec $a_i \neq a_j$ et $b_i \neq b_j$ si $i \neq j$.

1. Que vaut α ? Déterminer les lois marginales de X et de Y.
2. Montrer que $\text{Cov}(X, Y) = \frac{(a_2 - a_1)(2b_1 - b_2 - b_3)}{25}$.
3. Les variables aléatoires X et Y sont-elles indépendantes ?

Exercice 7.12 (★ - ✍)
Soit (X, Y) un couple de variables aléatoires de Bernoulli.

1. Montrer que $\text{Cov}(X, Y) = P([X = 1] \cap [Y = 1]) - P([X = 1])P([Y = 1])$.
2. Montrer que deux variables aléatoires de Bernoulli sont indépendantes si et seulement si elles sont non corrélées.

a) Notons tout d'abord que XY uniquement les valeurs 0 et 1, et suit donc une loi de Bernoulli de paramètre $p = P(XY = 1) = P([X = 1] \cap [Y = 1])$. Ainsi on a :

$$E(XY) = p = P([X = 1] \cap [Y = 1]).$$

Par la formule de Huygens, on a donc (toutes les variables sont finies, donc elles admettent bien des espérances et variances) :

$$Cov(X, Y) = E(XY) - E(X)E(Y) = P([X = 1] \cap [Y = 1]) - P(X = 1)P(Y = 1).$$

b) On sait déjà que deux variables indépendantes sont non corrélées, c'est à dire que $Cov(X, Y) = 0$. Mais la réciproque est fausse en générale !!! On va montrer que c'est cependant vrai si les variables suivent une loi de Bernoulli toutes les deux. Supposons donc que $Cov(X, Y) = 0$. Avec le calcul précédent, on obtient donc l'égalité :

$$P([X = 1] \cap [Y = 1]) = P(X = 1)P(Y = 1).$$

La famille $([X = 1], [Y = 1])$ d'évènement est donc indépendante. Par le cours, on sait qu'il en est alors de même de toute famille construite à partir de celle-ci en remplaçant l'un des évènement par son évènement contraire. Ce qui donne donc que pour tout $(i, j) \in \{0, 1\}^2$:

$$P([X = i] \cap [Y = j]) = P(X = i)P(Y = j).$$

D'où l'indépendance de X et Y.

Exercice 7.13 (★★)

Une urne contient r boules rouges, v boules vertes et b boules bleues. On tire n boules dans l'urne successivement et avec remise. On note R (resp. V, B) la variable aléatoire égale au nombre de boules rouges (resp. vertes, bleues) tirées.

1. Reconnaître les lois de R, V, B. Déterminer une relation liant R, V et B.
2. Calculer $V(R+V)$, puis montrer que $\rho_{R,V} = -\sqrt{\dfrac{rv}{(b+r)(b+v)}}$. Que peut-on en déduire ?
3. Que peut-on dire si $b = 0$? Était ce prévisible ?

Exercice 7.14 (★★)

Soit $p \in]0,1[$, et soit (X,Y) un couple de variables aléatoires à valeurs dans $\mathbb{N}$ dont la loi conjointe est donnée par :

$$\forall (n,k) \in \mathbb{N}^2, \quad P([X=n] \cap [Y=k]) = \begin{cases} \lambda(1-p)^k & \text{si } k \geq n \\ 0 & \text{sinon} \end{cases}$$

1. Déterminer la valeur de λ.
2. Déterminer les lois marginales du couple (X,Y). Quelle est la loi de $X+1$? En déduire $E(X)$ et $V(X)$.
3. Montrer que X et $Y-X$ suivent la même loi.
4. Montrer que X et $Y-X$ sont indépendantes. En déduire $\operatorname{Cov}(X,Y)$. X et Y sont-elles indépendantes ?

1. Notons pour tout $(n,k) \in \mathbb{N}^2$, $p_{n,k} = \begin{cases} \lambda(1-p)^k & \text{si } k \geq n \\ 0 & \text{sinon} \end{cases}$. Notons pour commencer que $\lambda \geq 0$ puisqu'une probabilité est nécessairement positive, et que les $p_{n,k}$ sont donc tous positif. On va montrer que la famille $(p_{n,k})$ est sommable (ce qui est sous-entendu dans l'énoncé puisque $([X=n] \cap [Y=k])$ est un SCE) et que sa somme vaut 1. Appliquons pour cela le théorème de Fubini :

   - Pour tout $k \geq 0$, $\sum\limits_{n \geq 0} p_{n,k}$ converge car $p_{n,k} = 0$ si $n > k$. Il y a donc un nombre fini de termes non nuls dans cette série, et sa somme vaut donc :
   $$\sum_{n=0}^{+\infty} p_{n,k} = \sum_{n=0}^{k} \lambda(1-p)^k = \lambda(k+1)(1-p)^k.$$
   - La série $\sum\limits_{k \geq 0} \lambda(k+1)(1-p)^k$ est une série géométrique dérivée, avec $1-p \in]-1,1[$. Elle converge donc également.

   On peut donc conclure que la famille $(p_{n,k})$ est sommable, et sa somme vaut :
   $$\sum_{n=0}^{+\infty}\sum_{k=0}^{+\infty} p_{n,k} = \lambda \sum_{k=0}^{+\infty}(k+1)(1-p)^k = \lambda \sum_{k=1}^{+\infty} k(1-p)^{k-1} = \frac{\lambda}{(1-(1-p))^2}.$$
   Ainsi on a $\lambda = p^2$.

2. Calculons la loi marginale de X, à l'aide de la FPT et du SCE $([Y=k])_{k \in \mathbb{N}}$. Pour tout $n \in X(\Omega) = \mathbb{N}$, on a :
   $$\begin{aligned} P(X=n) &= \sum_{k=0}^{+\infty} P(X=n,\ Y=k) = \sum_{k=n}^{+\infty} P(X=n,\ Y=k) \\ &= \lambda \sum_{k=n}^{+\infty}(1-p)^k = \lambda(1-p)^n \frac{1}{1-(1-p)} = (1-p)^n p. \end{aligned}$$

En particulier, si on pose $Z = X + 1$, on a $Z(\Omega) = \mathbb{N}^*$ et pour tout $n \in \mathbb{N}^*$:

$$P(Z = n) = P(X = n - 1) = (1 - p)^{n-1}p.$$

Donc Z suit une loi $\mathscr{G}(p)$. On en déduit en particulier que $X = Z + 1$ admet une espérance et une variance et qu'on a :

$$E(X) = E(Z + 1) = E(Z) + 1 = \frac{1}{p} + 1 \quad \text{et} \quad V(X) = V(Z + 1) = V(Z) = \frac{1 - p}{p^2}.$$

Déterminons à présent la loi de Y. On a $Y(\Omega) = \mathbb{N}$ et pour tout $k \in \mathbb{N}$ (à l'aide de la FPT avec le SCE $([X = n])$) :

$$P(Y = k) = \sum_{n=0}^{+\infty} P(X = n,\ Y = k) = \sum_{n=0}^{k} P(X = n,\ Y = k) = (k + 1)p^2(1 - p)^k.$$

3. A priori, on a $(Y - X)(\Omega) \subset \mathbb{Z}$. Soit $k \in \mathbb{Z}$, $k < 0$, on a à l'aide de la FPT :

$$P(Y - X = k) = \sum_{n=0}^{+\infty} P(X = n,\ Y - X = k) = \sum_{n=0}^{+\infty} P(X = n,\ Y = k + n) = \sum_{n=0}^{+\infty} p_{n,k+n} = 0$$

car $k + n < n$ pour tout $n \in \mathbb{N}$. Supposons à présent $k \geq 0$. En reprenant le calcul précédent, on obtient :

$$\begin{aligned} P(Y - X = k) &= \sum_{n=0}^{+\infty} P(X = n,\ Y = k + n) \\ &= \lambda \sum_{n=0}^{+\infty} (1 - p)^{k+n} = p^2(1 - p)^n \frac{1}{1 - (1 - p)} = p(1 - p)^k \end{aligned}$$

On constate que X et $Y - X$ suivent bien la même loi.

4. Pour tout $(n, k) \in \mathbb{N}^*$, on a d'une part que :

$$P(X = n,\ Y - X = k) = P(X = n,\ Y = k + n) = p^2(1 - p)^{k+n},$$

et d'autre part :

$$P(X = n)P(Y - X = k) = (1 - p)^n p(1 - p)^k p = p(1 - p)^{k+n}.$$

Ces deux quantités sont donc égales, de sorte que X et $Y - X$ sont indépendantes. On en déduit que :

$$0 = \mathrm{Cov}(X, Y - X) = \mathrm{Cov}(X, Y) - \mathrm{Cov}(X, X)$$

par linéarité à droite de la covariance. D'où $\mathrm{Cov}(X, Y) = V(X) = \dfrac{1 - p}{p^2} \neq 0$, et X et Y ne sont pas indépendantes.

Exercice 7.15 (★★)
Soit $n \in \mathbb{N}$, $n \geq 2$. Une urne contient n boules numérotées de 1 à n. On tire deux boules successivement et sans remise dans cette urne. Soit X la variable aléatoire égale au premier numéro obtenu et Y la variable aléatoire égale au deuxième numéro obtenu.

1. Déterminer la loi du couple (X, Y) ainsi que ses lois marginales.
2. Montrer que $E(XY) = \dfrac{(n + 1)(3n + 2)}{12}$.
3. En déduire la covariance et le cœfficient de corrélation linéaire de X et Y.
4. Exprimer sous forme factorisée la variance $V(X + Y)$.

Exercice 7.16 (★★)

Soit $p \in]0;1[$ et $n > 2$. On considère n joueurs de basket-ball qui tirent chacun deux lancers francs. On considère qu'à chaque lancer, un joueur à une probabilité p de marquer, et que les deux lancers sont indépendants. On note X la variable aléatoire égale au nombre de joueurs ayant marqué leur premier lancer franc, et Z la variable aléatoire égale au nombre de joueurs ayant marqué au moins un lancer franc.

1. Déterminer la loi de X.
2. Montrer que Z suit une loi binomiale, donner son espérance et sa variance.
3. On pose $Y = Z - X$. Que représente la variable aléatoire Y ? Déterminer sa loi.
4. Les variables X et Y sont-elles indépendantes ? Calculer $\text{Cov}(X, Y)$.

> 1. X représente le nombre de succès dans une succession d'épreuves de Bernoulli identiques (réussir ou non son premier lancer franc) et indépendantes (les joueurs lancent de façon indépendante les uns des autres). Elle suit donc une loi binomiale $\mathscr{B}(n, p)$.
> 2. La probabilité qu'un joueur rate ses deux lancers est $(1-p)^2$ (par indépendance des deux lancers), et donc la probabilité qu'il marque au moins un des deux lancers est $1-(1-p)^2$. On en déduit que $Z \hookrightarrow \mathscr{B}(n, 1-(1-p)^2)$.
> 3. Y représente le nombre de joueurs ayant marqué uniquement leur second lancer franc. Comme pour chaque joueur, ceci se produit avec probabilité $p(1-p)$, on en déduit que $Y \hookrightarrow \mathscr{B}(n, p(1-p))$.
> 4. Les variables X et Y ne sont pas indépendantes, puisque par exemple $P(X = n)P(Y = n) \neq 0$ alors que :
> $$P(X = n,\ Y = n) = 0$$
> puisqu'on ne peut pas avoir à la fois n joueurs qui réussissent peur premier lancer franc, et n joueurs qui réussissent uniquement leur second lancer franc.
>
> On connait $V(X + Y) = V(Z) = n(1-(1-p)^2)(1-p)^2$. On peut obtenir la covariance par la formule :
> $$\begin{aligned}\text{Cov}(X, Y) &= \frac{1}{2}(V(X+Y) - V(X) - V(Y)) \\ &= \frac{n}{2}((1-(1-p)^2)(1-p)^2 - p(1-p) - p(1-p)(1-p(1-p)) \\ &= \frac{n(1-p)}{2}(2p - p^2 - 2p^2 + p^3 - p - p + p^2 - p^3) = -n(1-p)p^2\end{aligned}$$
>
> **Remarques.**
>
> - On aurait pu commencer par calculer la covariance, constater qu'elle est non nulle, et en déduire que X et Y ne sont pas indépendantes.
> - On a une covariance négative. Cela signifie qu'en moyenne, quand X augmente, Y a tendance à diminuer, et vice-versa. Cela semble conforme à l'intuition : plus le nombre de joueurs marquant leur premier panier est important, plus le nombre joueurs manquant le premier et réussissant le second est faible.

Exercice 7.17 (★★)

On lance n dés équilibrés à 6 faces. On note X le nombre de numéros distincts qui sont sortis lors des n lancers, et pour tout $i \in \{1, \cdots, 6\}$, on note X_i la variable de Bernoulli qui vaut 1 si et seulement si le numéro i est apparu.

1. Déterminer la loi des variables X_i.
2. Déterminer l'espérance de X.
3. Pour tout $(i, j) \in [\![1, 6]\!]^2$, déterminer la loi de X_iX_j. En déduire $\text{Cov}(X_i, X_j)$. Les variables aléatoires X_i et X_j sont-elles indépendantes ?

4. Déterminer $V(X)$.

Exercice 7.18 (★★)

On suppose que le nombre de personnes qui se présentent à l'entrée d'un cinéma en une heure est une variable aléatoire X qui suit une loi de Poisson de paramètre $\lambda > 0$. Le cinéma comporte $N \geq 3$ caisses, et on suppose que chaque personne choisit au hasard sa caisse parmi les N. Pour $i \in [\![1, N]\!]$, on note X_i le nombre de personnes ayant choisi la caisse numéro i.

1. Déterminer la loi de X_i conditionnellement à l'événement $[X = n]$, puis la loi de X_i.
2. Déterminer, sans nouveaux calculs la loi de $X_1 + X_2$.
3. En déduire la covariance de X_1 et X_2.
4. Déterminer le coefficient de corrélation linéaire entre X_1 et X.
5. (★) Les variables aléatoires X_1 et X_2 sont-elles indépendantes ?

1. La loi de X_i conditionnellement à $[X = n]$ est une loi binomiale $\mathscr{B}(n, \frac{1}{N})$. En effet, chacune des n personnes arrivant au cinéma choisit de façon indépendante et équiprobable l'un des N guichets. Chaque personne a donc une probabilité de $\dfrac{1}{N}$ de choisir la caisse i.

   On a $X_i(\Omega) = \mathbb{N}$. Les évènements $[X = n]$ pour $n \in \mathbb{N}$ forment un SCE. Par la formule des probabilités totales, on a donc pour tout $k \in \mathbb{N}$:

$$\begin{aligned}
P(X_i = k) &= \sum_{n=0}^{+\infty} P(X_i = k, X = n) = \sum_{n=k}^{+\infty} P(X_i = k, X = n) \\
&= \sum_{n=k}^{+\infty} P(X = n) P_{[X=n]}(X_i = k) = \sum_{n=k}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{k} (\frac{1}{N})^k (\frac{N-1}{N})^{n-k} \\
&= \frac{e^{-\lambda}}{N^k} \sum_{n=k}^{+\infty} \frac{\lambda^n}{n!} \frac{n!}{k!(n-k)!} (\frac{N-1}{N})^{n-k} = \frac{e^{-\lambda}}{k! N^k} \sum_{n=k}^{+\infty} \frac{\lambda^n}{(n-k)!} (\frac{N-1}{N})^{n-k} \\
&= \frac{\lambda^k e^{-\lambda}}{k! N^k} \sum_{n=k}^{+\infty} \frac{\lambda^{n-k}}{(n-k)!} (\frac{N-1}{N})^{n-k} = \frac{\lambda^k e^{-\lambda}}{k! N^k} \sum_{n=0}^{+\infty} \frac{\lambda^n}{(n)!} (\frac{N-1}{N})^n \\
&= \frac{\lambda^k e^{-\lambda}}{k! N^k} e^{\lambda \frac{N-1}{N}} = \frac{(\lambda/N)^k}{k!} e^{-\frac{\lambda}{N}}
\end{aligned}$$

   Ainsi la variable X_i suit une loi de Poisson de paramètre λ/N.

2. En procédant comme précédemment, on montre que la loi de $X_1 + X_2$ sachant $[X = n]$ est une loi $\mathscr{B}(n, 2/N)$, et avec le même calcul que précédemment, on montre alors que $X_1 + X_2$ suit une loi $\mathscr{P}(2\lambda/N)$.

   Mise en garde.

   On ne peut pas utiliser la stabilité de la loi de Poisson car on ne sait pas si les variables X_1 et X_2 sont indépendantes ou pas.

3. On a la formule :

$$Cov(X_1, X_2) = \frac{1}{2}(V(X_1 + X_2) - V(X_1) - V(X_2))$$

   d'où avec les calculs précédents, et en se souvenant de la variance d'une loi de Poisson :

$$Cov(X_1, X_2) = \frac{1}{2} \left(\frac{2N}{\lambda} - \frac{N}{\lambda} - \frac{N}{\lambda} \right) = 0$$

4. On a $X = X_1 + \cdots + X_n$. D'où par linéarité à droite de la covariance :
$$Cov(X_1, X) = \sum_{i=1}^{n} Cov(X_1, X_i) = V(X_1) + \underbrace{\sum_{i=2}^{n} Cov(X_1, X_i)}_{=0} = V(X_1) = \frac{\lambda}{N}.$$
Ainsi on obtient que :
$$\rho_{X_1,X} = \frac{Cov(X_1, X)}{\sigma(X_1)\sigma(X)} = \frac{\lambda/N}{\sqrt{(\lambda/N)\lambda}} = \frac{1}{\sqrt{N}}.$$

5. La covariance entre X_1 et X_2 est nulle, mais cela n'assure pas que ces variables sont indépendantes, et ne nous dispense donc pas du calcul. On a pour tout $(i, j) \in \mathbb{N}^2$, en utilisant que $([X = n], n \in \mathbb{N})$ est un SCE :
$$\begin{aligned} P(X_1 = i, X_2 = j) &= \sum_{n=0}^{+\infty} P(X = n) P_{[X=n]}(X_1 = i, X_2 = j) \\ &= \sum_{n=i+j}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{i} \binom{n-i}{j} \frac{1}{N^i} \frac{1}{N^j} \left(\frac{N-2}{N}\right)^{n-i-j} \end{aligned}$$
car il faut choisir i personnes parmi n prenant la caisse 1, et donc j parmi $n - i$ prenant la caisse 2. Et alors les $n - i - j$ autres personnes choisissent une caisse parmi les $N - 2$ autres. D'où :
$$\begin{aligned} P(X_1 = i, X_2 = j) &= \frac{e^{-\lambda}}{N^{i+j}} \sum_{n=i+j}^{+\infty} \frac{\lambda^n}{n!} \frac{n!}{i!(n-i)!} \frac{(n-i)!}{j!(n-i-j)!} \left(1 - \frac{2}{N}\right)^{n-i-j} \\ &= \frac{\lambda^{i+j} e^{-\lambda}}{i! j! N^{i+j}} \sum_{n=i+j}^{+\infty} \frac{\lambda^{n-i-j}}{(n-i-j)!} \left(1 - \frac{2}{N}\right)^{n-i-j} \\ &= \frac{\lambda^{i+j} e^{-\lambda}}{i! j! N^{i+j}} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \left(1 - \frac{2}{N}\right)^{n} \\ &= \frac{\lambda^{i+j} e^{-\lambda}}{i! j! N^{i+j}} e^{\lambda - 2\frac{\lambda}{N}} = \frac{(\lambda/N)^i}{i!} e^{-\frac{\lambda}{N}} \frac{(\lambda/N)^j}{j!} e^{-\frac{\lambda}{N}} = P(X = i) P(Y = j) \end{aligned}$$
Ainsi les variables X_1 et X_2 sont bien indépendantes.

Exercice 7.19 (★★★★ - Oral HEC 2014)

On lance indéfiniment un dé équilibré et, pour tout $n \in \mathbb{N}^*$, on note X_n le numéro sorti au n-ième tirage. Les variables aléatoires X_n, définies sur un espace probabilisé $(\Omega, \mathscr{A}, P)$, sont donc supposées indépendantes et de même loi uniforme sur $[\![1, 6]\!]$.
Pour tout $i \in [\![1, 6]\!]$, on note T_i le temps d'attente de la sortie du numéro i.

1. (a) Donner la loi de T_1 ainsi que son espérance et sa variance.
   (b) Trouver l'espérance des variables aléatoires $\text{Inf}(T_1, T_2)$ et $\text{Sup}(T_1, T_2)$.
2. Justifier l'existence de la covariance de T_1 et de T_2, que l'on notera $\text{Cov}(T_1, T_2)$.
3. (a) Établir, pour tout $i \in [\![2, 6]\!]$, la relation : $E(T_1 \mid [X_1 = i]) = 7$.
   (b) Montrer que pour tout $i \in [\![3, 6]\!]$, on a : $E(T_1 T_2 \mid [X_1 = i]) = E((1 + T_1)(1 + T_2))$.
   (c) Calculer $E(T_1 T_2)$.
   (d) En déduire $\text{Cov}(T_1, T_2)$ ainsi que le coefficient de corrélation linéaire de T_1 et T_2.
4. (a) Trouver un réel α tel que les variables T_1 et $T_2 + \alpha T_1$ soient non corrélées.
   (b) Calculer l'espérance conditionnelle $E(T_2 + \alpha T_1 \mid [T_1 = 1])$.
   (c) Les variables aléatoires T_1 et $T_2 + \alpha T_1$ sont-elles indépendantes ?

1. (a) T_1 suit une loi géométrique de paramètre $p = \frac{1}{6}$. On a $E(T_1) = \frac{1}{p} = 6$ et $V(T_1) = \frac{1-p}{p^2} = 36 - 6 = 30$.

   (b) $I = Inf(T_1, T_2)$ représente le temps d'attente de la sortie du numéro 1 ou du numéro 2. Cette variable suit donc aussi une loi géométrique de paramètre $\frac{2}{6} = \frac{1}{3}$. On a donc $E(I) = 3$.

   On a d'autre part, en notant $S = Sup(T_1, T_2)$, l'égalité :

$$T_1 + T_2 = I + S \quad \Rightarrow \quad S = T_1 + T_2 - I.$$

   Comme toutes les variables I, T_1, T_2 admettent une espérance, il en est de même de S, et on a :

$$E(S) = -E(I) + E(T_1) + E(T_2) = 6 + 6 - 3 = 9.$$

2. T_1 et T_2 admettent une variance puisqu'elles suivent des lois géométriques, donc $Cov(T_1, T_2)$ existe bien.

3. (a) Soit $i \in [\![2, 6]\!]$. Si $[X_1 = i]$ est réalisé, on a donc obtenu un numéro différent de 1 au premier tirage. Le temps d'attente pour obtenir le numéro 1 sachant cet évènement réalisé est donc $1 + T_1$, c'est à dire le temps d'attente pour obtenir le numéro 1 auquel il faut ajouté le premier lancé qui n'a pas donné le numéro 1. Ainsi la loi de T_1 sachant $[X_1 = i]$ est égale à la loi de $T_1 + 1$. Donc $E(T_1 \mid [X_1 = i])$ existe et vaut $E(T_1 + 1) = E(T_1) + 1 = 7$.

   (b) Soit $i \in [\![2, 6]\!]$. Par le même raisonnement, on a que la loi de T_1 sachant $[X_1 = i]$ est la même que celle de $T_1 + 1$, et de même pour T_2. Les espérances en jeux existant bien puisque l'espérance de T_1T_2 existe, on a :

$$E(T_1T_2 \mid [X_1 = i]) = E(1 + T_1)(1 + T_2))$$

   (c) Le système $([X_1 = i], i = 1, \dots, 6)$ est complet, et l'espérance $E(T_1T_2)$ existe. Par la formule de l'espérance totale, on a :

$$\begin{aligned}
E(T_1T_2) &= \sum_{i=1}^{6} E(T_1T_2 \mid [X_1 = i])P(X_1 = i) \\
&= \frac{1}{6}E(T_1T_2 \mid [X_1 = 1]) + \frac{1}{6}E(T_1T_2 \mid [X_1 = 2]) + \frac{1}{6}\sum_{i=3}^{6} E(T_1T_2 \mid [X_1 = i]) \\
&= \frac{1}{6}E(T_2 \mid [X_1 = 1]) + \frac{1}{6}E(T_1 \mid [X_1 = 2]) + \frac{1}{6}\sum_{i=3}^{6} E((1 + T_1)(1 + T_2)) \\
&= \frac{1}{6}(7 + 7 + 4E((1 + T_1)(1 + T_2))) = \frac{1}{6}(14 + 4(1 + E(T_1) + E(T_2) + 4E(T_1T_2)) \\
&= \frac{1}{6}(18 + 8 \times 6 + E(T_1T_2)) = 11 + \frac{2}{3}E(T_1T_2)
\end{aligned}$$

   On obtient donc que $E(T_1T_2) = 33$.

   (d) On obtient donc que $Cov(T_1, T_2) = E(T_1T_2) - E(T_1)E(T_2) = 33 - 36 = -3$ et que :

$$\rho_{T_1,T_2} = \frac{Cov(T_1, T_2)}{\sigma(T_1)\sigma(T_2)} = \frac{-3}{30} = -\frac{1}{10}.$$

4. (a) Soit $\alpha \in \mathbb{R}$. On a par linéarité à droite de la covariance :

$$Cov(T_1, T_2 + \alpha T_1) = Cov(T_1, T_2) + \alpha V(T_1) = -3 + 30\alpha.$$

   Ainsi $Cov(T_1, T_2 + \alpha T_1) = 0$ si et seulement si $\alpha = \frac{1}{10}$.

(b) On a par linéarité de l'espérance, en remarquant que $[T_1 = 1] = [X_1 = 1]$, que :

$$\begin{aligned} E(T_2 + \alpha T_1 \mid [T_1 = 1]) &= E(T_2 + \alpha \mid [T_1 = 1]) = E(T_2 \mid [T_1 = 1]) + \alpha \\ &= E(T_2 \mid [X_1 = 1]) + \alpha = 7 + \frac{1}{10} = \frac{71}{10} \end{aligned}$$

(c) Ces deux variables sont non corrélées, mais cela n'implique pas qu'elles soient indépendantes. Pour montrer qu'elles ne le sont pas, on calcule $E(T_2 + \alpha T_1)$. On a par linéarité de l'espérance :

$$E(T_2 + \alpha T_1) = E(T_2) + \alpha E(T_1) = 6 + \frac{6}{10} = \frac{66}{10}$$

Puisque $E(T_2 + \alpha T_1) \neq E(T_2 + \alpha T_1 \mid [T_1 = 1])$, on peut en déduire que T_1 et $T_2 + \alpha T_1$ ne sont pas indépendantes.