Probabilités discrètes et Statistiques descriptives **L2 Université Lyon 1 2018**

Corrigé du CC2

Exercice 1 :

1. On utilise soit la formule de Bayes : $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)P(A)}{\mathbb{P}(B|A)P(A)+\mathbb{P}(B|A^c)P(A^c)}$ soit la définition de la probabilité conditionnelle $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$ et la formule des probabilités totales $\mathbb{P}(B) = \mathbb{P}(B|A)P(A) + \mathbb{P}(B|A^c)P(A^c)$, pour tous évènements A et B.
2. La probabilité que le test soit positif pour une personne donnée est

$$\begin{aligned}\mathbb{P}(Positif) &= \mathbb{P}(Malade)\mathbb{P}(Positif|Malade) + \mathbb{P}(Sain)\mathbb{P}(Positif|Sain)\\ &= 10^{-4} \times 0,95 + (1 - 10^{-4}) \times 0,05\\ &= 5009.10^{-5}\end{aligned}$$

donc la probabilité que M. Legris soit malade sachant que le test est positif est

$$\begin{aligned}\mathbb{P}(Malade|Positif) &= \frac{\mathbb{P}(Malade \cap Positif)}{\mathbb{P}(Positif)}\\ &= \frac{10^{-4} * 0,95}{5009.10^{-5}}\\ &= 9,5/5009 \sim \frac{1}{500}.\end{aligned}$$

3. La probabilité que Mme Souscolline soit malade sachant que le test est négatif est

$$\begin{aligned}\mathbb{P}(Malade|Negatif) &= \frac{P(Negatif|Malade)P(Malade)}{P(Negatif|Malade)P(Malade) + P(Negatif|Sain)P(Sain)}\\ &= \frac{0.05 * 0.0001}{0.05 * 0.0001 + 0.9999 * 0.95} \sim \frac{5}{5 + 95.10^4} \sim \frac{1}{190000}.\end{aligned}$$

Exercice 2 : La famille est sommable si $\sum_{(i,j)\in\mathbb{N}^2} \frac{|\lambda|^{i+j}}{(i+j+1)!} < \infty$. Commençons déjà par calculer cette somme : comme il s'agit d'une somme à termes positifs, on peut regrouper par paquets. On considère les paquets

$$A_k = \{(i, j) \in \mathbb{N}^2 \mid i + j = k - 1\}$$

où $k \in \mathbb{N}^*$. On a l'union disjointe $\mathbb{N}^2 = \cup_{k=1}^{\infty} A_k$, donc

$$\sum_{(i,j)\in\mathbb{N}^2} \frac{|\lambda|^{i+j}}{(i + j + 1)!} = \sum_{k=1}^{\infty} \sum_{(i,j)\in A_k} \frac{|\lambda|^{i+j}}{(i + j + 1)!} .$$

Mais pour tout $(i, j) \in A_k$ on a $\frac{|\lambda|^{i+j}}{(i+j+1)!} = \frac{|\lambda|^{k-1}}{k!}$ et $A_k = \{(i, k - i - 1)|i \in \{0, \cdots , k - 1\}\}$ donc A_k est de cardinal k. D'où

$$\sum_{(i,j)\in\mathbb{N}^2} \frac{|\lambda|^{i+j}}{(i + j + 1)!} = \sum_{k=1}^{\infty} k\frac{\lambda^{k-1}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k - 1)!} = e^{|\lambda|} < +\infty.$$

Ainsi, la famille est sommable. On peut donc appliquer le théorème de sommation par paquet à la famille des $\frac{\lambda^{i+j}}{(i+j+1)!}$, et avec les mêmes paquets A_k que plus haut on obtient

$$\sum_{(i,j)\in\mathbb{N}^2} \frac{\lambda^{i+j}}{(i + j + 1)!} = e^{\lambda}.$$

Exercice 3 :

1. Pour tout t avec $|t| \leq 1$ on a

$$G_{2X}(t) = \mathbb{E}\left(t^{2X}\right) = \mathbb{E}\big(\ (t^2)^X\ \big) = G_X(t^2)$$

et

$$G_{X+1}(t) = \mathbb{E}\left(t^{X+1}\right) = t\ \mathbb{E}\left(t^X\right) = t\ G_X(t).$$

2. On a $\mathbb{P}(X = k) = \frac{1}{d}$ pour tout $k \in \{0, \cdots, d-1\}$. Ainsi

$$\mathbb{E}(X) = \sum_{k=0}^{d-1} \frac{k}{d} = \frac{1}{d}\frac{d(d-1)}{2} = \frac{d-1}{2}\ .$$

De plus, pour tout $t \neq 1$ on a

$$G_X(t) = \sum_{k=0}^{d-1} \frac{t^k}{d} = \frac{t^d - 1}{d(t-1)}\ .$$

et $G_X(1) = 1$ (la fonction G_X est infiniment dérivable sur $\mathbb{R}$ puisque c'est un polynôme).

3. Cette question n'est pas un bonus pour rien. Il faut d'abord calculer $\mathbb{E}(X^2)$, le mieux est de calculer $\mathbb{E}(X(X-1)$. On propose deux méthodes pour ce calcul.

— **Méthode 1 :** On calcule $G''_X(1)$. La méthode la plus rapide pour cela est le développement de Taylor en 1 : quand z tend vers 0 on a

$$G_X(1+z) = G_X(1) + zG'_X(1) + \frac{z^2}{2}G''_X(1) + o(z^2)$$

et par ailleurs

$$(1+z)^d = 1 + \binom{d}{1}z + \binom{d}{2}z^2 + \binom{d}{3}z^3 + o(z^3)$$

donc

$$\begin{aligned} G_X(1+z) &= \frac{1 + dz + \frac{d(d-1)}{2}z^2 + \frac{d(d-1)(d-2)}{6}z^3 - 1 + o(z^3)}{dz} \\ &= 1 + z\frac{d-1}{2} + \frac{z^2}{2}\frac{(d-1)(d-2)}{3} + o(z^2) \end{aligned}$$

et donc $G''_X(1) = \frac{(d-1)(d-2)}{3}$. D'où $\mathbb{E}(X(X-1) = \frac{(d-1)(d-2)}{2}$.

— **Méthode 2 :** On fait une intégration par partie discrète. On a

$$\mathbb{E}(X(X-1)) = \sum_{k=2}^{d-1} \frac{k(k-1)}{d}\ .$$

On remarque que $k = \frac{k(k+1)}{2} - \frac{k(k-1)}{2}$. Posons $x_k = \frac{k(k+1)}{2}$, on a donc que $k = x_k - x_{k-1}$ ("k est la dérivée discrète de x_k"). Ainsi

$$\begin{aligned} \sum_{k=2}^{d-1} k(k-1) &= \sum_{k=2}^{d-1}(k-1)(x_k - x_{k-1}) \\ &= \sum_{k=2}^{d-1}(k-1)x_k - \sum_{k=1}^{d-1}(k-1)x_{k-1} \\ &= \sum_{k=2}^{d-1}(k-1)x_k - \sum_{l=1}^{d-2} lx_l \qquad \text{changement de variables } l = \\ &= \sum_{k=2}^{d-2}\big((k-1)x_k - kx_k\big) + (d-2)x_{d-1} - x_1 \quad \text{on regroupe les deux sommes. C'est l'ipp dis} \\ &= \frac{(d-2)(d-1)d}{2} - \frac{1}{2}\sum_{k=2}^{d-2} k(k+1)\ . \end{aligned}$$

Notons $s=\sum_{k=2}^{d-1}k(k+1)=2\sum_{k=1}^{d-2}x_k$, on a montré que

$$s=\frac{(d-2)(d-1)d}{2}-\frac{1}{2}s$$

donc

$$s+\frac{1}{2}s=\frac{d(d-1)(d-2)}{2}$$

et donc

$$s=\frac{d(d-1)(d-2)}{3}\ .$$

On retrouve que $\mathbb{E}(X)=s/d=\frac{(d-1)(d-2)}{3}$.
La variance de X est donc

$$Var(X)=\frac{(d-1)(d-2)}{3}+\frac{d-1}{2}-\frac{(d-1)^2}{4}=\frac{d^2-1}{12}\ .$$

Méthode 3 :
On montre la formule $\sum_{k=1}^{n}k^2=\frac{(2n+1)(n+1)n}{6}$ Si on ne connaît pas la formule d'avance (où que l'on ne se souvient que de l'équivalent $n^3/3$ venant d'une somme de Riemann, on cherche les coeff. tels que $(2n^3+an^2+bn+c)/6$ donne la bonne formule. Par récurrence, pour $n=1$ les deux membres valent 1=3*2/6. Sinon, au rang n,$\frac{(2n+1)(n+1)n}{6}=\frac{(2n-1)(n-1)n+6n^2}{6}$ (car $(2n+1)(n+1)=2n^2+3n+2=2n^2-3n+2+6n=(2n-1)(n-1)+6n)$ ce qui convient vu $\sum_{k=1}^{n}k^2=n^2+\sum_{k=1}^{n-1}k^2$.
De même, $E(X^2)=\frac{1}{d}\sum_{k=0}^{d-1}k^2=\frac{(2(d-1)+1)(d-1)d}{6d}=\frac{(2d-1))(d-1)}{6}$ Donc $Var(X)=\frac{(2d-1))(d-1)}{6}-\frac{(d-1)^2}{4}=(d-1)\frac{4d-2-3d+3}{12}=(d-1)\frac{d+1}{12}$.

4. Comme les variables $X_0,\cdots,X_n$ sont mutuellement indépendantes, la fonction génératrice de Z est le produit des fonctions génératrices des d^kX_k. De plus, comme dans la question 1,

$$G_{d^kX_k}(t)=\mathbb{E}\left(t^{d^kX^k}\right)=G_{X_k}\left(t^{d^k}\right)$$

et d'après la question 2, pour tout $t\neq 1$ on a

$$G_{d^kX_k}(t)=\frac{\left(t^{d^k}\right)^d-1}{d\left(t^{d^k}-1\right)}=\frac{t^{d^{k+1}}-1}{d\left(t^{d^k}-1\right)}\ .$$

Ainsi, si $t\neq 1$ on a

$$\begin{aligned}
G_Z(t)&=\prod_{k=0}^{n-1}G_{d^kX_k}(t)\\
&=\prod_{k=0}^{n-1}\frac{t^{d^{k+1}}-1}{d\left(t^{d^k}-1\right)}\\
&=\frac{\prod_{k=0}^{n-1}\left(t^{d^{k+1}}-1\right)}{\prod_{k=0}^{n-1}d\left(t^{d^k}-1\right)}\\
&=\frac{\prod_{l=1}^{n}\left(t^{d^l}-1\right)}{\prod_{k=0}^{n-1}d\left(t^{d^k}-1\right)} && \text{en prenant } l=k+1 \text{ dans la somme du numérateur}\\
&=\frac{t^{d^n}-1}{d^n(t-1)} && \text{en éliminant les produits termes à termes}
\end{aligned}$$

On reconnais la fonction génératrice de la loi uniforme sur $\{0,\cdots,d^n-1\}$, qui est donc la loi de Z.

Remarque : On aurait aussi pu trouver directement la loi de Z. En effet, tout entier $k \in \{0, \cdots, d^n - 1$ admet une unique écriture en base d, c'est à dire une unique suite $x_0, \cdots, x_n \in \{0, \cdots, d-1\}$ tels que $\sum_{i=0}^{n-1} x_i d^i = k$. Ainsi, on a

$$\mathbb{P}(Z = k) = \mathbb{P}(\forall i,\ X_i = x_i) = \prod_{i=0}^{n-1} \mathbb{P}(X_i = x_i) = \frac{1}{d^n}$$

et donc c'est bien la loi uniforme.

Exercice 4 :

1. Voir le cour ou les TDs.
2. S'il y a n saumons, le nombre de saumons capturés est le nombre de succès dans une liste de n expériences indépendantes qui réussissent avec probabilité $p = 1/4$. Ainsi, sachant $X = n$ la variable Y est une binomiale de paramètres n et p, et donc
$$\mathbb{P}(Y = k \mid X = n) = \binom{n}{k} p^k (1-p)^{n-k} \ .$$
3. Soit $k \in \mathbb{N}$, calculons la probabilité que $Y = k$. On a
$$\begin{aligned}
\mathbb{P}(Y = k) &= \sum_{n=0}^{\infty} \mathbb{P}(Y = k \ et\ X = n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(Y = k \mid X = n)\mathbb{P}(X = n) \\
&= \sum_{n=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda n} \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda n} \qquad \text{car } \binom{n}{k} = 0 \text{ si } k > n \\
&= \frac{p^k \lambda^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \left(\frac{p}{1-p}\right)^{n-k} \lambda^{n-k} \\
&= \frac{p^k \lambda^k e^{-\lambda}}{k!} e^{\frac{p}{1-p}\lambda} \\
&= \frac{(p\lambda)^k}{k!} e^{-\lambda p} \ .
\end{aligned}$$
Il s'agit donc d'une loi de Poisson de paramètre λp.
4. Les lois Y et $X - Y$ sont indépendantes (bien qu'à n fixé elles ne le soient pas !)
$$\mathbb{P}(Y = k \ et\ X - Y = l) = \mathbb{P}(Y = k \ et\ X = l + k) = \mathbb{P}(Y = k \mid X = l + k)\mathbb{P}(X = l + k) = \frac{\lambda^{k+l} p^k (1-p)^l}{k!l!} e^{-\lambda} \ .$$
Donc en sommant sur k on trouve $\mathbb{P}(X - Y = l) = \frac{\lambda^l (1-p)^l}{l!} e^{-\lambda(1-p)}$ soit $Y \sim \mathcal{P}(\lambda(1-p))$.
Donc pour tous l, k on a
$$\mathbb{P}(Y = k)\mathbb{P}(X - Y = l) = \frac{\lambda^{k+l} p^k (1-p)^l}{k!l!} e^{-\lambda}$$
Cette indépendance est contre-intuitive, mais forme une sorte de réciproque de la question 1.