Corrigé rapide de l'examen de juin 2007 - Proba 1 - L1 MMIA

Exercice 1

Cet exercice a été traité en TD.

1. Y suit une loi de Bernoulli de paramètre $p = 0,75$. On a $\mathbb{E}(Y) = p$ et $\mathrm{Var}(Y) = p(1-p)$.

2. Lorsque le médicament est inefficace, X suit une loi de Poisson de paramètre 5, d'espérance et de variance 5. Lorsque le médicament fait effet, X suit une loi de Poisson de paramètre 2.

3. L'événement "Stéphanie attrape 2 rhumes" est $\{X = 2\}$. L'énoncé se traduit par :

$$\mathbb{P}\left(X = 2 \quad \middle| \quad Y = 1\right) = \frac{2^2}{2!}e^{-2} = 2e^{-2}, \qquad \text{et} \qquad \mathbb{P}\left(X = 2 \quad \middle| \quad Y = 0\right) = \frac{5^2}{2!}e^{-5} = \frac{25}{2}e^{-5}.$$

Nous cherchons $\mathbb{P}(Y = 1 \,|\, X = 2)$. Par la formule de Bayes :

$$\mathbb{P}(Y = 1 \,|\, X = 2) = \frac{\mathbb{P}(X = 2 \,|\, Y = 1)\,\mathbb{P}(Y = 1)}{\mathbb{P}(X = 2 \,|\, Y = 1)\,\mathbb{P}(Y = 1) + \mathbb{P}(X = 2 \,|\, Y = 0)\,\mathbb{P}(Y = 0)} = \frac{1,5e^{-2}}{1,5e^{-2} + 3,125e^{-5}}.$$

Exercice 2

1. La probabilité qu'Esterinna tire une boule verte est $v/(r+v+b)$. La probabilité pour qu'elle tire une boule rouge ou verte est $(r+v)/(r+v+b)$.

2. X est à valeurs dans $\mathbb{N}^*$ (Esterinna peut tirer tout de suite une boule rouge ou verte, auquel cas $X = 1$, ou tirer indéfiniment des boules blanches). X suit une loi géométrique de paramètre $(r+v)/(r+v+b)$, la probabilité de tirer une boule soit verte soit rouge (le jeu s'arrête dès que cet événement se réalise).

3. La probabilité que Rosanna gagne au $i^{\text{ième}}$ tirage est la probabilité que $X = i$ et que la boule tirée par Esterinna soit verte. La probabilité de cet événement est :

$$\begin{aligned}\mathbb{P}(\{X = i\} \cap \{\text{la ième boule est verte}\}) &= \mathbb{P}\left(\text{la ième boule est verte} \quad \middle| \quad X = i\right)\mathbb{P}(X = i) \\ &= \frac{v}{r+v} \times \left(1 - \frac{r+v}{r+v+b}\right)^{i-1} \frac{r+v}{r+v+b} = \left(\frac{b}{r+v+b}\right)^{i-1} \frac{v}{r+v+b}.\end{aligned}$$

On peut aussi directement écrire que c'est la probabilité de tirer $i-1$ boules blanches et une boule verte au tirage i.

4. La fonction génératrice G_X de X est :

$$G_X(s) = \mathbb{E}\left(e^{sX}\right) = \sum_{k=1}^{+\infty} e^{sk}\mathbb{P}(X = k) == \frac{r+v}{b}\sum_{k=1}^{+\infty}\left(e^s \frac{b}{r+v+b}\right)^k$$

en utilisant le résultat de la question précédente. La somme dans cette expression est finie si et seulement si $e^s b/(r+v+b) < 1$ c'est à dire, si et seulement si $s < \ln(1 + (r+v)/b)$. Dans ce cas :

$$G_X(s) = \frac{r+v}{b}e^s \frac{b}{r+v+b}\frac{1}{1 - \frac{e^s b}{r+v+b}} = \frac{(r+v)e^s}{r+v+b-e^s b}.$$

5. Nous rappelons que $\mathbb{E}(X) = G'_X(0)$ et $\mathbb{E}(X^2) = G''_X(0)$.

$$\mathbb{E}(X) = \frac{(r+v)(r+v+b)}{(r+v+b-b)^2} = 1 + \frac{b}{r+v}, \quad \mathbb{E}(X^2) = \left(\frac{r+v+b}{r+v}\right)^3$$

$$\mathrm{Var}(X) = \left(1 + \frac{b}{r+v}\right)^3 - \left(1 + \frac{b}{r+v}\right)^2 = \frac{b}{r+v}\left(1 + \frac{b}{r+v}\right)^2.$$

Exercice 3

1. La variable X est à valeurs dans $[\![0,n]\!]$. Pour $k \in [\![1,n]\!]$ et $i \in [\![0,n]\!]$, $\mathbb{P}\left((X,Z)=(i,k)\right)=0$ si $i>k$ et sinon :

$$\mathbb{P}\left((X,Z)=(i,k)\right)=\mathbb{P}\left(\{X=i\}\cap\{Z=k\}\right)=\mathbb{P}\left(X=i\,|\,Z=k\right)\mathbb{P}\left(Z=k\right)=\frac{\mathbb{P}\left(Z=k\right)}{k+1}.$$

On en déduit la loi marginale de X. Pour $i \in [\![0,n]\!]$:

$$\mathbb{P}\left(X=i\right)=\sum_{k=1}^{n}\mathbb{P}\left((X,Z)=(i,k)\right)=0+\sum_{k=i}^{n}\frac{\mathbb{P}\left(Z=k\right)}{k+1}.$$

2. On déduit de la question précédente :

$$\begin{aligned}\mathbb{E}\left(X\right)&=\sum_{i=0}^{n}i\mathbb{P}\left(X=i\right)=\sum_{i=0}^{n}i\sum_{k=i}^{n}\frac{\mathbb{P}\left(Z=k\right)}{k+1}=\sum_{k=1}^{n}\frac{\mathbb{P}\left(Z=k\right)}{k+1}\sum_{i=0}^{k}i\\&=\sum_{k=1}^{n}\frac{\mathbb{P}\left(Z=k\right)}{k+1}\frac{k(k+1)}{2}=\frac{1}{2}\sum_{k=1}^{n}k\mathbb{P}\left(Z=k\right)=\frac{\mathbb{E}\left(Z\right)}{2}.\end{aligned}$$

3. La variable $Z-X$ est à valeurs dans $[\![0,n]\!]$. Soit $j \in [\![0,n]\!]$.

$$\mathbb{P}\left(Z-X=j\right)=\mathbb{P}\left(\bigcup_{k=1}^{n}\{Z-X=j\}\cap\{Z=k\}\right)=\sum_{k=j}^{n}\mathbb{P}\left(\{Z-X=j\}\cap\{Z=k\}\right),$$

car les événements $\{Z-X=j\}\cap\{Z=k\}$ pour $k\in[\![1,n]\!]$ sont disjoints, et car $Z=X+j\geq j$. En conditionnant par rapport à Z pour utiliser le fait que la loi de X conditionnellement à Z est une loi uniforme :

$$\mathbb{P}\left(Z-X=j\right)=\sum_{k=j}^{n}\mathbb{P}\left(Z-X=j\,|\,Z=k\right)\mathbb{P}\left(Z=k\right)=\sum_{k=j}^{n}\frac{\mathbb{P}\left(Z=k\right)}{k+1}=\mathbb{P}\left(X=j\right),$$

par la question **1**.

Exercice 4

1. Pour estimer p, on peut considérer la moyenne empirique $\bar{X}_n=\sum_{i=1}^{n}X_i/n$.
2. $\mathbb{E}(\bar{X}_n)=p$, $\mathrm{Var}(\bar{X}_n)=\sum_{i=1}^{n}\mathrm{Var}(X_i)/n^2=p(1-p)/n$ par indépendance des $(X_i)_{i\in[\![1,n]\!]}$.
3. Par la loi des grands nombres, $\lim_{n\to+\infty}\bar{X}_n=\mathbb{E}\left(X_1\right)=p$.
4. Pour $a>0$, par l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}\left(\left|\bar{X}_n-p\right|\geq a\right)\leq\frac{\mathbb{E}\left(\left|\bar{X}_n-p\right|^2\right)}{a^2}=\frac{\mathrm{Var}(\bar{X}_n)}{a^2}=\frac{p(1-p)}{a^2n}$$

en utilisant les résultats de la question **2**.
5. $f'(p)=1-2p$. $f'(p)=0 \quad\Leftrightarrow\quad p=1/2$. f est croissante sur $[0,1/2]$ et décroissante sur $[1/2,1]$. Elle atteint son maximum en $p=1/2$ et $f(1/2)=1/4$.
6. On déduit des questions **4** et **5** :

$$\mathbb{P}\left(\left|\bar{X}_n-p\right|\geq a\right)\leq\frac{1}{4a^2n}.\tag{1}$$

Pour que $\mathbb{P}\left(\left|\bar{X}_n-p\right|\geq 0,01\right)<0,05$, il est donc suffisant que :

$$\frac{1}{4\times 0,01^2n}<0,05\quad\Leftrightarrow\quad n>50\,000.$$

Il faut interroger au moins 50 000 personnes pour avoir, dans 95% des cas, une erreur d'approximation inférieure à 1%.
7. L'inégalité (1) se retraduit par :

$$\mathbb{P}\left(p\notin\left]\bar{X}_n-a,\bar{X}_n+a\right[\right)\leq\frac{1}{4a^2n}.$$

Pour $n=1000$, l'intervalle $\left]\bar{X}_n-a,\bar{X}_n+a\right[$ contient p dans plus de 95% des cas dès que :

$$\frac{1}{4a^2 1000}<0,05\quad\Leftrightarrow\quad a^2>5.10^{-3}.$$

Numériquement, $\sqrt{5.10^{-3}}\simeq 7\%$ et pour $\bar{X}_n=51\%$, on obtient que la proportion des électeurs ayant l'intention de voter pour Nicolas est $p\in]44\%,58\%[$. Cet intervalle n'exclut pas que $p<50\%$! *Cet exercice est inspiré d'un sondage réalisé par la TNS-Sofres le 24 avril 2007. Le cadre de cet exercice est bien sûr simplificateur, par les hypothèses faites sur les répondants et par les techniques d'estimation de p choisies (nécessité de redressement des fausses ou non-réponses, méthodes de sondage par "quota" pour garantir la représentativité de l'échantillon de personnes interrogées...)*