Méthodes Statistiques pour l'Informatique
Corrigé

Université Bordeaux I — Licence Informatique

Questions

1. La somme des v.a. normales indépendantes est une v.a. normale dont l'espérance est la somme des espérances et la variance la somme des variances. Elle est donc de type $\mathcal{N}(2-5+4, \sqrt{9+4+9}) = \mathcal{N}(1, \sqrt{22})$.
2. Par une énumération directe, on obtient : $Pr(1) = 1/36, Pr(2) = 2/36, Pr(3) = 2/36,$ $Pr(4) = 3/36, Pr(5) = 2/36, Pr(6) = 4/36, Pr(8) = 2/38, Pr(9) = 1/36,$ $Pr(10) = 2/36, Pr(12) = 4/36, Pr(15) = 2/36, Pr(16) = 1/36, Pr(18) = 2/36,$ $Pr(20) = 2/36, Pr(24) = 2/36, Pr(25) = 1/36, Pr(30) = 2/36, Pr(36) = 1/36.$
   D'où $F(x)$ vaut :
$$\begin{gathered}0 \text{ pour } x < 1, 1/36 \text{ pour } 1 \leq x < 2, 3/36 \text{ pour } 2 \leq x < 3;\\ 5/36 \text{ pour } 3 \leq x < 4, 8/36 \text{ pour } 4 \leq x < 5, 10/36 \text{ pour } 5 \leq x < 6,\\ 14/36 \text{ pour } 6 \leq x < 8, 16/36 \text{ pour } 8 \leq x < 9, 17/36 \text{ pour } 9 \leq x < 10,\\ 19/36 \text{ pour } 10 \leq x < 12, 23/36 \text{ pour } 12 \leq x < 15, 25/36 \text{ pour } 15 \leq x < 16,\\ 26/36 \text{ pour } 16 \leq x < 18, 28/36 \text{ pour } 18 \leq x < 20, F(x) = 30/36 \text{ pour } 20 \leq x < 24;\\ 32/36 \text{ pour } 24 \leq x < 25, 33/36 \text{ pour } 25 \leq x < 30, 35/36 \text{ pour } 30 \leq x < 36,\\ 1 \text{ pour } x \geq 36.\end{gathered}$$
3. Soit X un v.a. de type $\mathcal{N}(\mu, \sigma)$. On a $Pr(|X-\mu| \geq 4\sigma) = Pr(|\frac{X-\mu}{\sigma}| \geq 4)$. Mais $\frac{X-\mu}{\sigma}$ est une v.a. normale centrée-réduite et la probabilité recherchée vaut approximativement 0.00006.
4. D'après les propriétés sur la somme des v.a. indépendantes, on peut confirmer que $4X - 3Y$ est une v.a. dont l'espérance vaut $4\mathbb{E}X - 3\mathbb{E}Y$ et la variance $16\mathbb{V}X + 9\mathbb{V}Y$.

Choisissez 3 exercices parmi les 4 exercices proposés

Exercice 1

1. On $Pr(X_1 = k) = (\frac{5}{6})^{k-1}\frac{1}{6}$, pour $k = 1, 2, 3, ...$ C'est une distribution géométrique.
2. On a :

$$\begin{aligned} Pr(X_1 \leq k) &= \sum_{i=1}^{k} (\frac{5}{6})^{i-1}\frac{1}{6} \\ &= 1 - (\frac{5}{6})^k, \ k = 1, 2, 3, ... \end{aligned}$$

3. Il est facile de voir qu'on répète le lancer de chaque dé jusqu'à ce que la face obtenue soit un as. Dans ce cas, le nombre de jets (simultanés) est le nombre maximum de lancers des dés pour obtenir un as pour chaque dé. On en déduit que $X_n \leq k$, ssi, pour chaque dé le nombre de lancers nécessaires pour l'obtention un as (i.e. X_1 pour chacun des n dés) est inférieur ou égal à k. En supposant que chaque dé se comporte de façon indépendante, on aura :

$$Pr(X_n \leq k) = [1 - (\frac{5}{6})^k]^n, \ k = 1, 2, 3, ...$$

4. Nous avons alors :

$$\begin{aligned} Pr(X_n = k) &= Pr(X_n \leq k) - Pr(X_n \leq k-1) \\ &= [1 - (\frac{5}{6})^k]^n - [1 - (\frac{5}{6})^{k-1}]^n, \ k = 1, 2, 3, ... \end{aligned}$$

Exercice 2

1. Soit X le nombre de pièces défectueuses dans le lot choisi. On a :

$$\begin{aligned} Pr(X > 5) &= Pr(A).Pr(X > 5 \ / \ A) + Pr(B).Pr(X > 5 \ / \ B) \\ &= \frac{1}{2}\sum_{n=6}^{\infty} e^{-5}\frac{5^n}{n!} + \frac{1}{2}\sum_{n=6}^{\infty} e^{-8}\frac{8^n}{n!}. \end{aligned}$$

2. Le nombre de pièces défectueuses admet l'espérance :

$$\begin{aligned} \mathbb{E}X &= \frac{1}{2}5 + \frac{1}{2}8 \\ &= 6.5 \end{aligned}$$

et la variance (égale à son espérance) :

$$\begin{aligned} \mathbb{V}X &= \frac{1}{2}5 + \frac{1}{2}8 \\ &= 6.5. \end{aligned}$$

3. En vertu du théorème de Bayes, nous avons :

$$\begin{aligned}
Pr(A \ / \ X=5) &= \frac{Pr(A).Pr(X=5 \ / \ A)}{Pr(A).Pr(X=5 \ / \ A)+Pr(B).Pr(X=5 \ / \ B)} \\
&= \frac{\frac{1}{2}\frac{e^{-5}5^5}{5!}}{\frac{1}{2}\frac{e^{-5}5^5}{5!}+\frac{1}{2}\frac{e^{-8}8^5}{5!}} \\
&= \frac{e^3 5^5}{e^3 5^5+8^5}.
\end{aligned}$$

Exercice 3

1. Le nombre de cases occupées est n et, donc, la probabilité pour qu'un essai trouve une case libre vaut $\frac{m-n}{m} = 1-\alpha$. La probabilité pour que le k-ème essai aboutisse (et donc les précédents échouent) vaut alors $p_k = \alpha^{k-1}(1-\alpha), \ k = 1, 2, 3....$ Il s'agit d'une distribution géométrique.
2. La fonction génératrice du nombre d'essais est donnée par :

$$G(z) = \sum_{k=1}^{\infty} p_k z^k = \sum_{k=1}^{\infty}(1-\alpha)\alpha^{k-1}z^k = \frac{(1-\alpha)z}{1-\alpha z}.$$

   Et son espérance vaut $\frac{1}{1-\alpha}$.
3. Dans une recherche négative, on fait le même cheminement que lors d'une insertion pour arriver à la conclusion que l'élément recherché ne figure pas dans la table. On en déduit que nombre moyen d'essais, dans une recherche négative est égal à l'espérance du nombre d'essais, ce qui vaut $\frac{1}{1-\alpha}$.

Exercice 4
La fonction de vraisemblance est donnée par $L_n(\lambda) = \lambda^n exp(-\lambda \sum_{i=1}^n x_i)$. Pour maximiser cette fonction, il suffit de maximiser son logarithme qui vaut $n\ln(\lambda) - \lambda\sum_{i=1}^n x_i$. On dérive alors ce logarithme (par rapport à λ), on aura :

$$\frac{d}{d\lambda}\ln L_n(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i.$$

Elle s'annule pour $\lambda = \widehat{\lambda}_n(x_1, ..., x_n) = \frac{n}{\sum_{i=1}^n x_i}$. Par ailleurs, elle est positive si $\lambda < \widehat{\lambda}$ et négative si $\lambda > \widehat{\lambda}$. On en déduit que l'estimateur qui maximise la fonction de vraisemblance est $\widehat{\lambda}_n(x_1, ..., x_n) = \frac{n}{\sum_{i=1}^n x_i}$, c'est-à-dire l'inverse de la moyenne arithmétique.

1. D'après la première partie, l'estimateur du maximum de vraisemblance est l'inverse de la durée moyenne, ce qui vaut ici $100/88912 \approx 0.0011247076$.
2. La f.r. de la durée de vie vaut $F(t) = Pr(T \leq t) = 1 - e^{-\frac{11t}{10000}}$. On en déduit $Pr(500 \leq T \leq 1000) = e^{-0.55} - e^{-1.1}$.