



Statistique pour ingénieur

Thème 1 : Exercices de probabilités

D. Pastor, F.-X. Socheleau & C. Garnier, 26 octobre 2016

Exercice 1 : Spams

Dans une entreprise, 40% des courriers électroniques reçus sont des spams. Parmi tous les spams reçus, 50% contiennent le mot «order» et, parmi tous les courriels légitimes, 30% contiennent le mot «order».

1. Quelle est la probabilité qu'un mail arrivant contienne le mot «order» ?
2. Un mail arrivant contient le mot «order», quelle est la probabilité que ce soit un spam ?

Solution :

Introduisons les notations suivantes :

$$S = \{\text{le mail arrivant est un spam}\}$$

$$M = \{\text{le mail arrivant contient le mot «order»}\}$$

On a, d'après les données, $\mathbb{P}(S) = 0,4$, $\mathbb{P}(M|S) = 0,5$ et $\mathbb{P}(M|\bar{S}) = 0,3$. Alors, d'après la formule des probabilités totales avec la partition $\Omega = S \cup \bar{S}$, la probabilité que le mail arrivant contienne le mot «order» est

$$\mathbb{P}(M) = \mathbb{P}(M|S)\mathbb{P}(S) + \mathbb{P}(M|\bar{S})\mathbb{P}(\bar{S}) = 0,5 \times 0,4 + 0,3 \times 0,6 = 0,38.$$

Si le mail contient le mot «order», la formule de Bayes avec la même partition donne la probabilité qu'il s'agisse d'un spam :

$$\mathbb{P}(S|M) = \frac{\mathbb{P}(M|S)\mathbb{P}(S)}{\mathbb{P}(M)} = \frac{0,5 \times 0,4}{0,38} \simeq 0,53.$$

Exercice 2 : Durée de vie

La durée de vie d'un certain type de lampes de vidéoprojecteur suit une loi exponentielle de densité de probabilité

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{sinon.} \end{cases} \quad (1)$$

1. Une variable aléatoire X est dite « sans mémoire », ou « sans vieillissement », si elle vérifie la propriété suivante :

$$\mathbb{P}(X > x + x_0 | X > x_0) = \mathbb{P}(X > x), \quad (2)$$

pour tout x, x_0 positifs.

Montrer que la durée de vie de ces lampes de vidéoprojecteur est sans mémoire.

2. Déterminer la durée de vie moyenne d'une lampe en fonction de λ ?

3. Un fabricant de lampes annonce à son client que la durée de vie moyenne des lampes qu'il produit est d'au moins 10000 heures alors que, dans les faits, lampe sur deux a une durée de vie de moins de 7000 heures. Est-il honnête ? Justifier.

Solution :

1) D'après la formule de Bayes, on a :

$$\mathbb{P}(X > x + x_0 | X > x_0) = \frac{\mathbb{P}(X > x_0 | X > x + x_0) \mathbb{P}(X > x + x_0)}{\mathbb{P}(X > x_0)} = \frac{\mathbb{P}(X > x + x_0)}{\mathbb{P}(X > x_0)}.$$

Pour la loi exponentielle, on obtient donc :

$$\mathbb{P}(X > x + x_0 | X > x_0) = \frac{e^{-\lambda(x+x_0)}}{e^{-\lambda x_0}} = e^{-\lambda x} = \mathbb{P}(X > x).$$

2)

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}^+} x \lambda e^{-\lambda x} dx \\ &= \left[-x e^{-\lambda x} \right]_0^{+\infty} + \int_{\mathbb{R}^+} e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \end{aligned}$$

3) On déduit de l'énoncé que $\mathbb{P}(X < 7000) = \frac{1}{2}$. Avec cette information et à l'aide de la réponse à la question (2), il faut ensuite vérifier si l'assertion "la durée de vie moyenne des lampes est d'au moins 10000 heures" est vraie ou fausse. Par définition $\mathbb{P}(X < 7000) = \mathbb{F}_X(7000) = \frac{1}{2}$, alors

$$1 - e^{-\lambda 7000} = \frac{1}{2} \Rightarrow \lambda = \frac{\ln(2)}{7000}.$$

Or $\lambda = \frac{1}{\mathbb{E}(X)}$, on en déduit que $\mathbb{E}(X) = \frac{7000}{\ln(2)} \approx 10098$. Le constructeur est donc honnête.

Exercice 3 : Consommation énergétique

Pour effectuer un comparatif entre deux modèles de smartphones, une association de consommateurs s'intéresse à leurs consommation électriques respectives. Pour cette étude, elle dispose de données constructeur qui fournissent des informations sur la variabilité de consommation en fonction de l'usage qui est fait du téléphone.

Pour le modèle 1, le fabricant indique que sa consommation est une variable aléatoire X_1 de fonction de répartition \mathbb{F}_{X_1} et de densité f_{X_1} . Pour le modèle 2, la consommation est une variable aléatoire X_2 de fonction de répartition \mathbb{F}_{X_2} et de densité f_{X_2} . On suppose que les consommations des deux modèles sont indépendantes. L'association de consommateur ne dispose que d'un exemplaire de chaque modèle.

1. Pour un même usage, exprimer la probabilité que l'exemplaire du modèle 1 consomme

plus que l'exemplaire du modèle 2 à l'aide de \mathbb{F}_{X_1} et de f_{X_2} .

2. Dans le cas où $\mathbb{F}_{X_1} = \mathbb{F}_{X_2}$, quelle est la valeur de cette probabilité ?

Solution :

1)

$$\begin{aligned}\mathbb{P}(X_1 \geq X_2) &= \iint_{\{x_1 \geq x_2\}} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 \quad (\text{v.a. indépendantes}) \\ &= \int_{-\infty}^{+\infty} \left(\int_{x_2}^{+\infty} f_{X_1}(x_1) dx_1 \right) f_{X_2}(x_2) dx_2 \\ &= 1 - \int_{-\infty}^{+\infty} \mathbb{F}_{X_1}(x_2) f_{X_2}(x_2) dx_2\end{aligned}$$

2) Si $\mathbb{F}_{X_1} = \mathbb{F}_{X_2}$, on a :

$$\mathbb{P}(X_1 \geq X_2) = 1 - \int_{-\infty}^{+\infty} \mathbb{F}_{X_2}(x_2) f_{X_2}(x_2) dx_2$$

On pose $u = \mathbb{F}_{X_2}(x_2) \Rightarrow du = f_{X_2}(x_2) dx_2$, alors

$$\mathbb{P}(X_1 \geq X_2) = 1 - \int_0^1 u du = \frac{1}{2}$$

Exercice 4 : Transformation linéaire d'un vecteur gaussien

Soit $X = (X_1, X_2)^T$, un vecteur aléatoire¹ réel centré, à deux dimensions, de loi gaussienne et de matrice de covariance

$$\Gamma_X = \begin{pmatrix} 3 & \rho\sqrt{3} \\ \rho\sqrt{3} & 1 \end{pmatrix},$$

avec $|\rho| < 1$. On définit un nouveau vecteur aléatoire $(Y_1, Y_2)^T$ avec :

$$\begin{cases} Y_1 &= \frac{X_1}{\sqrt{3}} - X_2, \\ Y_2 &= \frac{X_1}{\sqrt{3}} + X_2. \end{cases}$$

1. Calculer la variance $\mathbb{V}(X_1)$ de X_1 et la covariance $\mathbb{Cov}(X_1, X_2)$ de (X_1, X_2) ?

2. Dans quel cas, les deux variables X_1 et X_2 sont-elles indépendantes ?

3. Calculer la covariance $\mathbb{Cov}(Y_1, Y_2)$ du couple (Y_1, Y_2) . Calculer les variances des variables aléatoires Y_1 et Y_2 .

4. Déterminer les densités de probabilités marginales de Y_1 et Y_2 .

5. Les deux variables Y_1 et Y_2 sont-elles indépendantes ? Justifier clairement votre réponse.

1. Reportez-vous à la vidéo 2 du thème 1 de ce MOOC pour la représentation d'un couple de variables aléatoires sous forme vectorielle.

Solution :

1) Il suffit de lire la matrice de covariance. On a alors $\mathbb{V}(X_1) = 3$ et $\text{Cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))) = \rho\sqrt{3}$.

2) X_1, X_2 indépendantes si et seulement si $\text{Cov}(X_1, X_2) = 0$ car le vecteur X est gaussien. Il y a donc indépendance si et seulement si $\rho = 0$ (coefficient de corrélation).

3) $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = 0$. On calcule maintenant :

$$\begin{aligned}\text{Cov}(Y_1, Y_2) &= \mathbb{E}(Y_1 Y_2) \\ &= \mathbb{E}\left(\left(\frac{X_1}{\sqrt{3}} - X_2\right)\left(\frac{X_1}{\sqrt{3}} + X_2\right)\right) \\ &= \frac{1}{3}\mathbb{E}(X_1^2) + \frac{1}{\sqrt{3}}\mathbb{E}(X_1 X_2) - \frac{1}{\sqrt{3}}\mathbb{E}(X_1 X_2) - \mathbb{E}(X_2^2) \\ &= \frac{1}{3}\mathbb{E}(X_1^2) - \mathbb{E}(X_2^2) = 1 - 1 = 0\end{aligned}$$

On calcule maintenant les variances de Y_1 et de Y_2 . On obtient :

$$\begin{aligned}\mathbb{V}(Y_1) &= \mathbb{E}(Y_1^2) = \mathbb{E}\left(\left(\frac{X_1}{\sqrt{3}} - X_2\right)^2\right) \\ &= \mathbb{E}\left(\frac{X_1^2}{3} + X_2^2 - \frac{2}{\sqrt{3}}X_1 X_2\right) \\ &= 1 + 1 - \frac{2}{\sqrt{3}}\rho\sqrt{3} = 2(1 - \rho)\end{aligned}$$

et

$$\begin{aligned}\mathbb{V}(Y_2) &= \mathbb{E}(Y_2^2) = \mathbb{E}\left(\left(\frac{X_1}{\sqrt{3}} + X_2\right)^2\right) \\ &= \mathbb{E}\left(\frac{X_1^2}{3} + X_2^2 + \frac{2}{\sqrt{3}}X_1 X_2\right) \\ &= 1 + 1 + \frac{2}{\sqrt{3}}\rho\sqrt{3} = 2(1 + \rho)\end{aligned}$$

4) Le vecteur aléatoire Y est obtenu par transformation linéaire du vecteur aléatoire gaussien X . Le vecteur aléatoire Y est donc lui-même gaussien. Par conséquent, ses composantes sont gaussiennes et $Y_1 \sim \mathcal{N}(0, 2(1 - \rho))$ et $Y_2 \sim \mathcal{N}(0, 2(1 + \rho))$.

5) L'indépendance des composantes de Y équivaut à la décorrélation de celles-ci. Comme $\text{Cov}(Y_1, Y_2) = 0$, la matrice de covariance de Y est diagonale. Il s'ensuit que ses composantes Y_1 et Y_2 sont décorréliées et donc indépendantes.

Exercice 5 : Application non probabiliste du théorème central-limite

On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes, qui suivent des lois de Poisson de paramètre unité : pour tout $n = 1, 2, \dots$, $X_n \sim \mathcal{P}(1)$. On pose $S_n = X_1 + X_2 + \dots + X_n$.

Pour l'exercice, on admettra que la somme de deux variables aléatoires de Poisson indépendantes X_1 et X_2 , telles que $X_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 \sim \mathcal{P}(\lambda_2)$ suit une loi de Poisson : $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

1. Quelle est la loi de S_n ? Soit \mathbb{F}_n la fonction de répartition de la variable aléatoire S_n . Montrer que $\mathbb{F}_n(n) = e^{-n} \sum_{k=0}^n \frac{n^k}{k!}$.

2. Montrer que $Z_n = \frac{S_n - n}{\sqrt{n}}$ converge en loi vers la loi normale centrée réduite.

3. Dédurre des questions précédentes que $\lim_{n \rightarrow +\infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = 1/2$.

Solution :

1) Une récurrence immédiate du résultat sur la somme de variables aléatoires de Poisson montre que $S_n \sim \mathcal{P}(n)$. Par définition de la fonction de répartition, on a $\mathbb{F}_n(n) = \mathbb{P}(S_n \leq n)$. Donc, $\mathbb{F}_n(n) = \sum_{k=0}^n \mathbb{P}(S_n = k)$. Puisque $S_n \sim \mathcal{P}(n)$, $\mathbb{P}(S_n = k) = \frac{n^k}{k!} e^{-n}$, d'où le résultat.

2) Les variables aléatoires X_1, \dots, X_n sont identiquement distribuées et indépendantes. D'après le théorème de théorème central-limite, $\frac{S_n - nm}{\sqrt{nm}\sigma}$ converge en loi vers $\mathcal{N}(0,1)$ avec $m = \mathbb{E}(X_i)$ et $\sigma^2 = \mathbb{V}(X_i)$. Comme $X_1 \sim \mathcal{P}(1)$, on a $\mathbb{E}(X_1) = \mathbb{V}(X_1) = 1$. On en déduit donc que $Z_n = \frac{S_n - n}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0,1)$.

3) On a $\{Z_n \geq 0\} = \{S_n \geq n\}$. Donc $\mathbb{P}(Z_n \geq 0) = \mathbb{P}(S_n \geq n)$. Comme Z_n converge en loi vers $\mathcal{N}(0,1)$, $\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \geq 0) = \mathbb{P}(U \geq 0)$ avec $U \sim \mathcal{N}(0,1)$, car 0 est un point de continuité de la fonction de répartition de la loi $\mathcal{N}(0,1)$. Or $\mathbb{P}(U \geq 0) = 1/2$, ce qui permet de conclure.