L3 Informatique - Examen de Probabilités et Statistiques

Jeudi 15 décembre 2016 16h30-18h30

(les supports distribués en cours sont les seuls documents autorisés)

Exercice 1 : Loi discrète (5 points)

Un certain vaccin provoque chez un individu sur 800 environ une réaction dangereuse. Soit X la variable aléatoire qui représente le nombre d'individus présentant une telle réaction.

Nous supposons **dans tout l'exercice** que 3000 personnes, choisies aléatoirement, sont vaccinées.

1. Quelle est le nom de la loi que vous pourriez utiliser pour calculer la probabilité $P(X = k)$ avec $k \in \mathbb{N}$? Quels sont les paramètres qui la définissent ? (1 point)
   loi binomiale avec $p = 1/800$ et $n = 3000$.
2. Quelle est la probabilité qu'il y ait trois réactions dangereuses ? Vous poserez le calcul numérique sans donner la valeur numérique. (1 point)

$$P(X = 3) = \frac{3000!}{3!2997!}\left(\frac{1}{800}\right)^3\left(\frac{799}{800}\right)^{2997}$$

Nous supposons dans la suite que la loi précédente peut être approchée par une loi de Poisson :

$$f(k) = \begin{cases} \dfrac{\lambda^k e^{-\lambda}}{k!} & \text{si k} \in \mathbb{N} \\ 0 & \text{sinon} \end{cases}$$

3. Déterminer, pour 3000 personnes vaccinées :
   a) la valeur de λ, (1 point)
   $\lambda = 3000/800 = 3,75$
   b) la probabilité qu'il y ait trois réactions dangereuses, cette fois calculer la valeur numérique, (1 point)
   $P(k = 3) = 0,2067$
   c) la probabilité qu'il y ait plus de deux réactions dangereuses, donner la valeur numérique. (1 point)
   $P(k > 2) = 1 - P(k = 2) - P(k = 1) - P(k = 0) = 0,7229$

Exercice 2 : Loi réelle (4 points)

Après enquête, nous estimons que le temps de charge d'une batterie de téléphone portable, exprimée en heures, est une variable aléatoire T dont la densité de probabilité est donnée par la fonction f_T définie par :

$$f_T(t) = \begin{cases} \lambda(1 - t^2) \text{ si } t \in [0, 1] \\ 0 \text{ sinon} \end{cases}$$

avec λ réel.

1. Déterminer la constante λ telle que f_T soit une densité de probabilités.
   (0.5pt formule + 1 pt calcul)
   $\int_0^1 \lambda(1-t^2)dt = 1$
   $$\int_0^1 \lambda(1-t^2)dt = \lambda\left[t - \frac{t^3}{3}\right]_0^1 = \frac{2}{3}\lambda = 1$$
   donc $\lambda = \frac{3}{2}$.
2. Déterminer la fonction de répartition F_T. (0.25 pt pour $t < 0$, 0.5 pt pour $t \in [0,1]$ et 0.25 pt pour $t > 1$)
   $$F_T(t) = \begin{cases} 0 \text{ si } t < 0 \\ \frac{3}{2}(t - \frac{t^3}{3}) \text{ si } 0 \leq t \leq 1 \\ 1 \text{ si } t > 1 \end{cases}$$
3. Calculer l'espérance de T. Quel est le temps moyen de charge de la batterie ?
   (1 pt pour calcul de l'espérance et 0.5 pt pour la réponse à la question)
   $$\begin{aligned} E(T) &= \frac{3}{2}\int_0^1 t(1-t^2)dt \\ &= \frac{3}{2}\left[\frac{t^2}{2} - \frac{t^4}{4}\right]_0^1 \\ &= \frac{3}{8} \end{aligned}$$
   Ainsi le temps moyen de passage en caisse est de 3/8 d'heure.

Exercice 3 : Régression et statistiques (6 points)

Nous utilisons les statistiques dans le domaine de la linguistique : nous nous intéressons à la proportion de verbes dans les fragments de texte de certains auteurs. Soit X le nombre de mots, et V le nombre de verbes. Remarque : un texte a réellement été traité pour obtenir ce tableau, dont les valeurs de X ont été ordonnées.

X	6	6	7	9	9	16	34	65	99	127
V	1	3	1	2	2	2	7	13	17	24

1. Construire le tableau représentant la loi jointe (X, V). (1 point loi jointe)
2. Donner les lois marginales empiriques de X et V. (0,25 point par marginale : total = 0,5 point)

X / V	1	2	3	7	13	17	24	
6	0,1	0	0,1	0	0	0	0	0,2
7	0,1	0	0	0	0	0	0	0,1
9	0	0,2	0	0	0	0	0	0,2
16	0	0,1	0	0	0	0	0	0,1
34	0	0	0	0,1	0	0	0	0,1
65	0	0	0	0	0,1	0	0	0,1
99	0	0	0	0	0	0,1	0	0,1
127	0	0	0	0	0	0	0,1	0,1
	0,2	0,3	0,1	0,1	0,1	0,1	0,1	1

3. Calculer les moyennes et écart-types empiriques de X et V. (0,5 point pour 2 formules + 0,25 par calcul : total = 1,5 points)
$\overline{V} = 7,2, \overline{X} = 37,8$
$var(V) = 58,76, var(X) = 1756,16, \sigma(V) = \sqrt{var(V)} = 7,666, \sigma(X) = \sqrt{var(X)} = 41,9066$
4. Calculer la covariance empirique et le coefficient de corrélation $corr(X, V)$. (covariance (0,5 formule + 0,5 calcul) + coeff. de corr. (0.25 + 0.25) : total = 1,5 points)
$cov(X, V) = 319,14, corr(X, V) = \frac{cov(X,V)}{\sigma(X)\sigma(V)} = 0,9938$
5. Sachant que sur une page d'un même texte nous avons 260 mots, combien de verbes doivent être présents (d'après l'équation de la droite de régression) ? Si en realité nous avons 49 verbes, quelle est la deviation Δ ? (eq. de regression (0,5 formule + 0,25 calcul) + prédiction (0,5) + delta (0,25) : total = 1,5 points)
On s'interesse de l'estimation de V par rapport à X, donc on écrit l'équation de regression de V sur X. L'équation est de la forme
$$v = ax + b$$
avec $a = \frac{cov(X,V)}{var(X)} = 0,1817$ et $b = \overline{V} - a\overline{X} = 0,3308$.
Pour obtenir l'estimation on prend $x = 260$ et obtient $v = 0,1817 \cdot 260 + 0,3308 = 47,5795$.
Si la vraie valeur est 49 on trouve Δ de l'équation $49 = 0.1817 \cdot 260 + 0,3308 + \Delta$, alors $\Delta = 1,4205$.

Exercice 4 : Chaîne de Markov (5 points)

Nous plaçons une souris dans la pièce 1 d'un labyrinthe (voir figure 1). Chaque fois qu'elle entend une sonnerie, apeurée, elle change de pièce en choisissant de manière équiprobable une des portes de la pièce dans laquelle elle se trouve. Lorque qu'elle atteint soit la nourriture (fromage en pièce 4), soit sa tanière (pièce 5), elle y reste.

Soit X_n , le numéro de la pièce dans laquelle se situe la souris après n étapes (sonneries).

1. Montrer que X_n est une chaîne de Markov. (0,5 point démo + 1 point résultat : total = 1,5 points)
Il est possible de calculer $P(X_{n+1}/X_n)$ (à montrer ou écrire les cas possibles !) donc l'étape $n + 1$ ne dépend que de l'étape n.

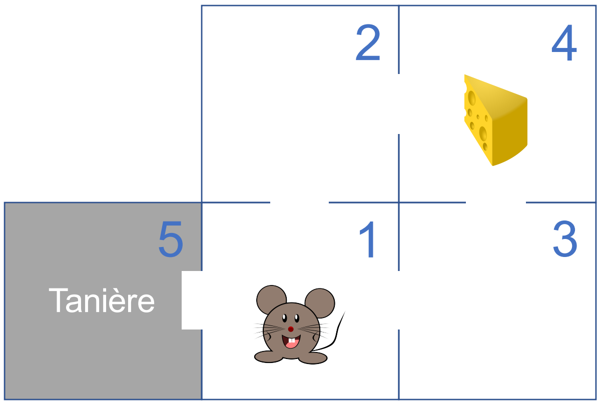
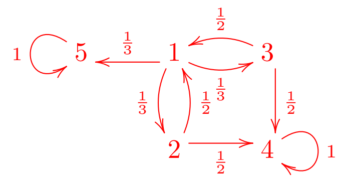


FIGURE 1 – Labyrinthe de la souris.

2. Donner la matrice de transition. (0,5 point)

$$P = \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

3. Tracer le graphe associé. (0,5 point)



4. La chaîne est-elle irréductible ? Justifier... (0,25 point réponse + 0,25 justification)
   Non. Par exemple, à partir de l'état 5, il n'est pas possible d'atteindre un autre état.
5. Donner les états transitoires et persistants. (0,5 point)
   Transitoires={1, 2, 3} et Persistants={4,5}
6. Calculer les probabilités d'absorption (par l'état 4). Partant de la pièce 1, la souris a-t-elle plus de chance de manger le fromage ou de rentrer dans sa tanière ? (0,5 point formule + 0,5 point calcul + 0,5 point réponse à la question)

$$\begin{cases} ab_1 = p_{12}ab_2 + p_{13}ab_3 \\ ab_2 = p_{24} + p_{21}ab_1 \\ ab_3 = p_{34} + p_{31}ab_1 \end{cases}$$

   $ab_1 = \frac{1}{2}$ et $ab_2 = ab_3 = \frac{3}{4}$

   Conclusion : la souris a autant de chance de finir dans la tanière (le ventre vide) que de manger le fromage !