UNIVERSITÉ KASDI MERBAH, OUARGLA
DÉPARTEMENT D'INFORMATIQUE ET TI
MODULE : STATISTIQUES POUR SCIENCES DE DONNÉES.

CORRIGÉ TYPE DE L'EXAMEN SEMESTRIEL

Exercice 1. (8 pt). Une usine produit et commercialise 40 téléviseurs par mois. Le coût de fabrication d'un téléviseur est de 5000 DA. L'usine fait réaliser un test de conformité sur chacun de ses téléviseurs. Le test est positif dans 95% des cas et un téléviseur reconnu conforme peut être vendu k DA. Si le test est en revanche négatif, le téléviseur est bradé au prix de 2500 DA.
Soit X la v.a. qui indique le nombre de téléviseurs conformes parmi les 40 téléviseurs produits par l'usine.

1. Quelle est la loi de probabilité de X ? donner son expression et calculer son espérance.
2. Quelle est la probabilité que tous les téléviseurs soient conformes ?
3. Quelle est la probabilité qu'au maximum 38 téléviseurs soient conformes ?
4. On note Y la variable aléatoire qui désigne le bénéfice mensuel de l'usine.
   (a) Donner l'expression de Y en fonction de X et k.
   (b) Calculer l'espérance de Y (en fonction de k).
   (c) Quelle doit être la valeur minimale de k, pour que l'usine ne fasse pas faillite ?

Solution :

1. X est par définition le nombre de téléviseurs conformes parmi 40. Donc X est une variable aléatoire binomiale de paramètres $n = 40$ et $p = 0.95$. La loi de X est donnée par

$$p(X = x) = \begin{cases} C_{40}^{x}(0.95)^{x}(0.05)^{(40-x)} & \text{si } x \in \{0, 1, 2, \ldots, 40\} \\ 0 & \text{sinon} \end{cases} \quad \textbf{(1pt)}$$

   $E(X) = np = 38$ **(0.5pt)**.
2. $P(X = 40) = C_{40}^{40}(0.95)^{40}(0.05)^{(40-40)} = 0.1285$ **(1pt)**.
3. $P(X \leq 38) = 1 - P(X \geq 39) = 1 - (P(X = 40) + P(X = 39)) = 1 - (0.1285 + 0.27) \approx 0.6015$ **(1pt)**.
4. Soit Y la variable aléatoire qui désigne le bénéfice mensuel de l'usine.
   (a) $Y = k \times X + 2500(40 - X) - 5000 \times 40 = (k - 2500)X - 100000$ **(1.5pt)**.
   (b) $E(Y) = E((k - 2500)X - 100000) = (k - 2500)E(X) - 100000 = 38k - 195000$ **(1.5pt)**.
   (c) Si on veut que l'usine ne tombe pas faillite, il faut que $E(Y) \geq 0$. Donc $38k - 195000 \geq 0$ et par conséquent, $k \geq 5131,58$ **(1.5pt)**.

Exercice 2. **(5 pt).** Dans un bureau de poste, il y a deux guichets. Chacune des personnes arrivant à la poste choisit le premier guichet avec une probabilité p, ou le deuxième guichet avec une probabilité $q = 1 - p$.

Les personnes effectuent leur choix de façon indépendante. En une heure, le nombre X de personnes arrivés à la poste suit une loi de Poisson $\mathcal{P}(\lambda)$

On désigne par Y le nombre de personnes ayant choisi le premier guichet.

1. Exprimer la probabilité conditionnelle de $Y = k$ sachant que $X = n$.
2. En déduire la loi conjointe du couple (X, Y).
3. Déterminer la loi de Y.

Solution :

1. Si on connaît le nombre de personnes qui arrivent au bureau de poste et est n, alors Y, qui désigne le nombre de personnes qui choisissent le guichet 1, est une variable binomiale de paramètres n et p. Donc

$$p(Y = k \mid X = n) = \begin{cases} \mathrm{C}_n^k (p)^k (1-p)^{(n-k)} & \text{si } k \leq n \\ 0 & \text{sinon} \end{cases} \textbf{(1pt)}$$

2. La loi conjointe $p(X = n, Y = k) = p(X = n) \times p(Y = k \mid X = n)$. Donc on a :

$$p(X = n, Y = k) = \begin{cases} \mathrm{e}^{-\lambda} \dfrac{\lambda^n}{n!} \mathrm{C}_n^k (p)^k (1-p)^{(n-k)} & \text{si } k \leq n \\ 0 & \text{sinon} \end{cases} \textbf{(1.5pt)}$$

3. La loi de Y est alors obtenue par

$$\begin{aligned}
\sum_{n \in \mathbb{N}} p(X = n, Y = k) &= \sum_{n \leq k} p(X = n, Y = k) + \sum_{n \geq k} p(X = n, Y = k) \\
&= 0 + \sum_{n \geq k} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!} \mathrm{C}_n^k (p)^k (1-p)^{(n-k)} \\
&= \mathrm{e}^{-\lambda} (p)^k \lambda^k \sum_{n \geq k} \frac{\lambda^{n-k}}{n!} \frac{n!}{(n-k)!k!} (1-p)^{(n-k)} \\
&= \frac{\mathrm{e}^{-\lambda} (p)^k \lambda^k}{k!} \sum_{n \geq k} \frac{\lambda^{n-k} (1-p)^{(n-k)}}{(n-k)!} \\
&= \frac{\mathrm{e}^{-\lambda} (p)^k \lambda^k}{k!} \sum_{l \geq 0} \frac{(\lambda(1-p))^l}{l!} \\
&= \frac{\mathrm{e}^{-\lambda} (p)^k \lambda^k}{k!} \mathrm{e}^{\lambda(1-p)} \\
&= \mathrm{e}^{-\lambda p} \frac{(\lambda p)^k}{k!} \textbf{(2.5pt)}
\end{aligned}$$

On reconnaît la loi de Poisson de paramètre λp

Exercice 3. **(7 pt).** On suppose que le poids d'un nouveau né est une variable aléatoire normale de moyenne m et d'écart type σ.

Dans un premier temps, un échantillon de 10 nouveaux nés est choisi. Le poids est alors donné comme suit : 3.5 ; 3.8 ; 4.2 ; 4.1 ; 3.6 ; 2.8 ;3.9 ;4 ; 3.7 ; 2.9

1. Donner une estimation ponctuelle de m et σ.
2. Déterminer un intervalle de confiance à 95% pour la moyenne.
3. Déterminer un intervalle de confiance à 90% pour l'écart type.

   Considérons maintenant un échantillon de 100 nouveaux nés où le poids moyen est égal 3.8 où l'estimation ponctuelle de l'écart type est égale à 2.5.
4. Déterminer le taux de risque que nous devons courrir si on veut que la longueur (amplitude) de l'intervalle de confiance pour la moyenne soit égale à 1.

Solution : On a $A = \sum_{i=1}^{10} x_i = 36.5$ et $B = \sum_{i=1}^{10} x_i^2 = 135.25$.

1. On calcule $\bar{X} = \dfrac{A}{10} = 3.65$ qui est un estimateur de la moyenne des poids **(1pt)**.

   $\sigma_{ech}^2 = \dfrac{B}{10} - \bar{X}^2 = 0.2025$, donc $\sigma_{ech} = 0.45$ et $s = \sqrt{\dfrac{10}{9}}\sigma_{ech} \approx 0.474$, l'estimateur ponctuel de σ **(1pt)**.
2. L'intervalle de confiance pour m est donné par $[\bar{X} - t_{\alpha/2}\dfrac{\sigma_{ech}}{\sqrt{n-1}}; \bar{X} + t_{\alpha/2}\dfrac{\sigma_{ech}}{\sqrt{n-1}}]$, avec $t_{\alpha/2}$ est tel que $P(|T| \leq t_{\alpha/2}) = 0.95$ où T suit une loi de Student à 9 degrés de liberté. Donc $t_{\alpha/2} = 2.262$ et par conséquent, L'IC est donné par $[3.3107; 3.9893]$ **(1.5pt)**.
3. On cherche L'IC pour la variance, qui est de la forme $[\dfrac{n\sigma_{ech}^2}{\chi_u^2}; \dfrac{n\sigma_{ech}^2}{\chi_l^2}]$ où χ_u^2 et χ_l^2 sont tels que $P(Y \leq \chi_l^2) = P(Y \geq \chi_u^2) = 0.05$ avec Y qui suit la loi de χ^2 à 9 degrés de liberté. On obtient $\chi_u^2 = 16.92$ et $\chi_l^2 = 3.33$. Par conséquent, l'IC pour la variance est $[0.12; 0.608]$ et celle pour l'écart type [0.346 ;0.78] **(1.5pt)**.
4. On a $n = 36 > 30$. Donc L'IC pour la moyenne s'écrit $[\bar{X} - t_{\alpha/2}\dfrac{s}{\sqrt{n}}; \bar{X} + t_{\alpha/2}\dfrac{s}{\sqrt{n}}]$. Donc si on veut que longueur de l'intervalle égale à 1, alors on aura $\bar{X} + t_{\alpha/2}\dfrac{s}{\sqrt{n}} - \bar{X} + t_{\alpha/2}\dfrac{s}{\sqrt{n}} = 1$, où $t_{\alpha/2}$ est tel que $P(|T| \geq t_{\alpha/2}) = \alpha$ avec T une variable normale centrée réduite. Par conséquent, $t_{\alpha/2} = \dfrac{\sqrt{n}}{2.s} = 1.2$ et d'après la table de la loi normale, $\alpha = 0.23$, donc le taux de risque d'erreur est égal à 23% **(2pts)**.