EXAMEN DE STATISTIQUES

Les notes de cours et de TP sont autorisées. L'examen comporte quatre exercices indépendants. Il est conseillé de lire tout le sujet et de tenir compte du barème.

1 Exercice 1 (4 points)

On étudie les revenus mensuels (en euros) d'un ensemble de familles d'un quartier de Toulouse :

Revenus	$[700; 900[$	$[900; 1100[$	$[1100; 1300[$	$[1300; 1400[$	$[1400; 1500[$	$[1400; 1600]$
Effectifs	13	219	20	46	50	82

1. Quelle est la moyenne des revenus ? Préciser la formule utilisée.
2. Quelle est l'écart-type des revenus ? Préciser la formule utilisée.
3. Dans quel intervalle se trouve la médiane ? La calculer en faisant une interpolation linéaire.
4. Sur un graphique soigné, représenter l'histogramme correspondant à cette distribution et placer la médiane en abscisse. Que remarquez-vous ?

2 Exercice 2 (6 points)

Un hypermarché dispose de 20 caisses. Le tableau ci-dessous donne le temps moyen d'attente à une caisse en fonction du nombre de caisses ouvertes :

Nombre de caisses ouvertes (caractère x)	3	4	5	6	8	10	12
Temps moyen d'attente (caractère y) en minutes	16	12	9,6	7,9	6	4,7	4

1. Construire le nuage de points $P_i = (u_i, v_i)$ correspondant à cette série statistique. Unités à utiliser pour le graphique : $1cm$ pour une caisse ouverte en abscisse ; $1cm$ pour une minute d'attente en ordonnée.
2. Calculer les coordonnées du *centre de gravité* $G = (\overline{x}, \overline{y})$ du nuage et le placer sur le graphique.
3. Calculer le coefficient de corrélation linéaire r du nuage.
4. Déterminer l'équation de la droite de régression D du nuage. Tracer D sur le graphique (marquer les points utilisés pour tracer D).
5. Estimer à l'aide de la droite de régression : le nombre de caisses à ouvrir pour que le temps moyen d'attente à une caisse soit de 5 minutes ; le temps moyen d'attente à la caisse lorsque 15 caisses sont ouvertes. Ces prédictions semblent-elles fiables (justifier la réponse) ?
6. On considère la fonction f définie sur $]0; +\infty[$ par $f(x) = \dfrac{\lambda}{x}$. Déterminer λ de façon à avoir $f(3) = 16$. Tracer le graphe de f sur le graphique précédent. Estimer en utilisant la fonction f : le nombre de caisses à ouvrir pour que le temps moyen d'attente à une caisse soit de 5 minutes ; le temps moyen d'attente à la caisse lorsque 15 caisses sont ouvertes.

3 Exercice 3 (7 points)

Une usine produit des pièces qui sont contrôlées en fin de fabrication : en prélevant et contôlant des pièces au hasard, on note le nombre de pièces correctement fabriquées jusqu'à l'obtention d'une première pièce défectueuse. Ce nombre est une variable aléatoire X de loi géométrique caractérisée par un paramètre p :

$$P[X = x] = p(1-p)^x \qquad x \in \mathbb{N} \tag{1}$$

Pour une telle loi géométrique, on a $E[X] = \frac{1-p}{p}$ et $Var[X] = \frac{1-p}{p^2}$. On désire estimer le paramètre p à partir de n observations de la variable X notées $(x_1, x_2, \ldots, x_n)$. On suppose que ces observations sont issues d'un échantillon iid $(X_1, X_2, \ldots, X_n)$ dont chaque composante est de même loi que X.

1. Proposer un estimateur de p selon la méthode des moments.
2. Donner l'expression de la vraisemblance de l'échantillon $L(x_1, x_2, \ldots, x_n, p)$ pour les réalisations $(x_1, x_2, \ldots, x_n)$.
3. Si on note $\theta = \frac{1-p}{p}$, alors en exprimant p en fonction de θ, on peut exprimer la vraisemblance précédente en fonction de θ. On la note $L(x_1, x_2, \ldots, x_n, \theta)$. Donner cette nouvelle expression.
4. Montrer que l'estimateur du maximum de vraisemblance de θ est égal à $\widehat{\theta_{MV}} = \frac{1}{n}\sum_{i=1}^n X_i$. On peut admettre le résultat pour répondre aux questions suivantes. On veillera, en cas d'échec dans les calculs liés à cette question, à toutefois bien préciser la méthode suivie.
5. L'estimateur $\widehat{\theta_{MV}}$ est-il sans biais ? convergent ? Il faut observer que θ a une relation simple avec les moments de la loi de X pour répondre à cette question simple.

4 Exercice 4 (3 points)

Soit un échantillon $(X_1, X_2, \ldots, X_n)$ de n variables aléatoires iid. Chaque X_i étant une v.a.r. continue distribuée selon une loi uniforme sur l'intervalle réel $[0, 2]$.

1. Rappeler la fonction densité de probabilité associée à chaque X_i.
2. Donner l'expression $L(x_1, x_2, \ldots, x_n)$ de la vraisemblance de l'échantillon pour les réalisations $(x_1, x_2, \ldots, x_n)$.
3. Pour $n = 4$, que vaut $L(x_1, x_2, x_3, x_4)$ avec $(x_1, x_2, x_3, x_4) = (1.5, 0.5, 1.8, 0.01)$ et avec $(x_1, x_2, x_3, x_4) = (1.1, 0.75, 3.3, 0.9)$? Donner une interprétation de vos réponses.