Partiel n°2

Il sera tenu compte du soin apporté à la rédaction.

Attention toute réponse devra être précisément justifiée. Une réponse oui ou non ne sera pas suffisante afin d'avoir les points.

$\boxed{1}$

Un vol Paris New York est assuré par un Airbus A380 d'une capacité de 538 places. La probabilité qu'une personne confirme sa réservation et retire son billet est notée p. On suppose que les comportements des voyageurs sont indépendants les uns des autres. On note X_n la variable aléatoire désignant le nombre de personnes se présentant à l'embarquement. La compagnie fait du surbooking (elle vend plus de billets qu'il n'y a de places, espérant ainsi remplir son appareil, et comptant sur le fait que des voyageurs ne se présentent pas) et vend n billets. Toutefois, si une personne se présente à l'embarquement et ne peut monter à bord car toutes les places sont prises, celle ci sera dédommagée.

Pour $1 \leq i \leq n$, on note Y_i la variable aléatoire qui vaut 1 si la $i$ième personne s'est présenté à l'embarquement et 0 sinon.

1. Déterminer la loi de X_n .

2. On suppose que $p \in]0,1[$. Donner un intervalle de confiance asymptotique de p au niveau 95 % basé sur l'observation des Y_i.

3. On suppose que $p = 0,9$, combien doit valoir au plus n de manière à ce qu'il y ait une probabilité d'au moins 95% que la compagnie aérienne n'ait pas fait de surbooking ?

$\boxed{2}$

En 1897, l'économiste suisse Vilfredo Pareto (1848-1923), professeur d'économie politique à l'université Lausanne, eut l'idée de modéliser la loi des revenus en postulant que le nombre relatif de personnes dont le revenu dépasse une valeur x est inversement proportionnel à une puissance de x. La définition suivante fut adoptée : une variable aléatoire X, absolument continue, suit une loi de Pareto de paramètres α et θ si sa densité est donnée par

$$f_{\alpha,\theta}(x) = (\alpha - 1)\theta^{\alpha-1}x^{-\alpha}\mathbb{1}_{[\theta,\infty[}(x).$$

où $\alpha > 3$ et $\theta > 0$ On considère un n-échantillon de loi de Pareto de paramètres α et θ. On suppose $\alpha > 0$ connu et on va vouloir estimer θ.

[a] Écrire le modèle sous la forme d'un triplet, espace d'observation, tribu, famille de probabilités.

[b] Ce modèle est-il dominé ? Identifiable ?

[c] Déterminer l'estimateur du maximum de vraisemblance de θ, noté $\hat{\theta}_0$.

[d] Soit $\hat{\theta}_1 = \min_{1\leq i\leq n} X_i$. Quelle est la loi de $\hat{\theta}_1$?

[e] Calculer le biais de $\hat{\theta}_1$. Que pouvez vous en déduire ?

[f] Calculer l'espérance de X_1 et en déduire un estimateur des moments de θ, noté $\hat{\theta}_2$.

[g] Quelle est la vitesse de convergence de l'estimateur $\hat{\theta}_1$? On pourra penser à étudier $Z_n = n(\hat{\theta}_1 - \theta)$.

[h] Quelle est la vitesse de convergence de l'estimateur $\hat{\theta}_2$?

[i] Question de cours (aucun calcul n'est demandé) : sur quel(s) critère(s) pourriez vous comparer $\hat{\theta}_1$ et $\hat{\theta}_2$ afin de savoir lequel est le plus pertinent ?

$\boxed{3}$ **Modèle à trois variables explicatives** On considère le modèle de régression suivant :

$$\forall i \in \{1, ..., n\} \quad y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \epsilon_i.$$

Les $x_{i,j}$ sont des variables exogènes du modèle, les ϵ_i sont des variables aléatoires indépendantes, de loi normale centrée admettant la même variance σ^2. En posant :

$$X = \begin{pmatrix} 1 & x_{1,2} & x_{1,3} & x_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,2} & x_{n,3} & x_{n,4} \end{pmatrix} \quad , \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \text{et} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix},$$

on a observé

$$X'X = \begin{pmatrix} 50 & 0 & 0 & 0 \\ 0 & 20 & 15 & 4 \\ 0 & 15 & 30 & 10 \\ 0 & 4 & 10 & 40 \end{pmatrix}, \qquad X'Y = \begin{pmatrix} 100 \\ 50 \\ 40 \\ 80 \end{pmatrix}, \qquad Y'Y = 640.$$

[a] Déterminer la valeur de n, calculer $\hat{\beta}$, estimateur des moindres carrés de β, la somme des carrés des résidus $\sum_{i=1}^{n} \hat{\epsilon}_i^2$ et donner un estimateur de σ^2.

[b] Donner un intervalle de confiance pour β_2 au niveau 95%. Faire de même pour σ^2 (on donner $c_1 = 29$ et $c_2 = 66$ pour les quantiles d'ordre $2,5\%$ et $97,5\%$ d'un chi-deux à 46 degré de liberté).