Université de Strasbourg
UFR de Mathématique
et d'Informatique

Licences 3
Statistique : Étude de Cas
Octobre 2019

Contrôle Continu N^0 1

Responsable : H LI

Les notes de cours et de travaux dirigés sont autorisées,
le courriel (e-mail) et le téléphone ne sont pas autorisés.
Durée de l'épreuve : 60 minutes
La rédaction et les commandes doivent être reportées sur la copie
avec les résultats numériques éventuels.
Le sujet est à rendre en même temps que la copie.
Responsable : H LI

NOM : Prénom :

NOM : Prénom :

Exercice 1 : Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0;1[$:

$$\mathbb{P}(X = k) = (1-p)^{k-1}p,\ k \geq 1$$

1. Soit $(1, 1, 3, 1, 2)$ une réalisation d'un échantillon $(X_1, ..., X_5)$ de loi $\mathcal{G}(p)$. Simplifiez l'expression suivante :

$$L(p) = \mathbb{P}\left(X_1 =, 1 X_2 = 1, X_3 = 3, X_4 = 1, X_5 = 2\right)$$

2. Calculez les valeurs de $L(p)$ pour $p = 0.5$, 0.3 et 0.1.
3. On définit :
   (a) $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$,
   (b) $S_c^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$.

   Ces deux variables sont elles des statistiques ? Sont-elles indépendantes ? Pourquoi ?
4. On simule $N = 10000$ réalisations d'un échantillons de taille $n = 5$ d'une loi géométrique de paramètre $p = 3/4$ à l'aide de la fonction `rgeom(n,p)+1`. Dans le but d'estimer $\theta = \frac{1}{p^2}$ $(= 1.777778)$, en vous basant sur ces 10000 réalisations, reportez les résultats de l'évaluation des performances des estimateurs suivants (sans le code) :
   (a) $S_c^2 + \overline{X}$,
   (b) $\dfrac{\left(\displaystyle\sum_{i=1}^{n} X_i - 1\right)^2}{n(n-1)}$,
   (c) $\dfrac{n\overline{X}^2 + \overline{X}}{n+1}$.

Exercice 2 :

1. Soit U une variable aléatoire suivant la loi uniforme continue sur l'intervalle $[0;1]$. On définit une seconde variable aléatoire $X = \dfrac{\lambda}{(1-U)^{1/3}}$ où $\lambda > 0$ est un paramètre. ($x^{1/3}$ désigne la racine cubique de x, c'est -à-dire l'unique racine positive telle que $\left(x^{1/3}\right)^3 = x$).
   (a) Calculez la fonction de répartition de X, c'est à dire :
   $$F_X(x) = \mathbb{P}(X \leq x).$$
   (b) Montrez que sa dérivée $F'(x)$ est égale à la la densité de probabilité d'une loi de Paréto de paramètre λ :
   $$f_\lambda(x) = \frac{3\lambda^3}{x^4}\mathbf{1}_{x>\lambda}$$
2. (a) En exploitant les résultats précédents, donner le code pour simuler $N = 1000$ réalisations d'un échantillon $(X_1, ..., X_{50})$ $(n = 50)$ d'une loi de Paréto de paramètre $\lambda = 2$ via la commande `runif(...)`.
   Admis : Si X suit une loi de Paréto de paramètre $\lambda > 0$, alors $\mathbb{E}(X) = \frac{3\lambda}{2}$ et $Var(X) = \frac{3\lambda^2}{4}$.
   (b) Étuider les des variables aléatoires suivantes :
      i. $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$. Quelle est la moyenne de $N = 1000$ réalisations de $\overline{X}$? Si On augmente la taille N à 10^9, 10^{90}, ..., quelle sera la valeur limite ? En vertu de quel théorème ?
      ii. $S_c^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(\overline{X} - X_i\right)^2$. Quelle est la moyenne de $N = 1000$ réalisations de S_c^2 ? Si On augmente la taille N à 10^9, 10^{90}, ..., quelle sera la valeur limite ? En vertu de quel théorème ?
      iii. $T = (n-1)\dfrac{S_c^2}{\sigma^2}$. La variable aléatoire T est-elle une statistique ? Connaissez vous la loi de T ? Si oui, laquelle ?
   (c) Sauvez les 1000 réalisations associées des variables aléatoires suivantes :
      i. $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
      ii. $Z = \dfrac{\sqrt{n}\left(\overline{X} - \mu\right)}{\sigma}$,
3. (a) À l'aide de la fonction `hist(...,probability=TRUE)`, donner le code pour afficher l'histogramme de $Z/1000$.
   (b) Superposez-lui le graphe de la densité d'une loi normale aux paramètres bien choisis à l'aide des fonctions `curve` et `dnorm`. Justifiez vos choix de paramètres.
   (c) Quel théorème venez-vous d'illustrer ? Énoncez-le.
4. À l'aide des fonctions `mean` et `var`, estimez l'espérance et la variance de $\overline{X}$ en vous basant sur ces 1000 réalisations. Ces résultats sont-ils cohérents avec les points 2.(b)i. et 2.(b)ii. ?