LMAFY1101 - Exercices - Série 6

Variables aléatoires: Quelques lois usuelles
PARTI II

Distribution d'échantillonnage et TCL

Exercice 1

Une agence de voyage loue un bateau de plaisance pour 50 clients. Le poids des clients est supposé avoir une moyenne de 80kg et un écart-type de 18kg. La capacité du bateau est de 4.2 tonnes. Répondez aux deux questions suivantes en supposant la normalité des poids et sans cette hypothèse.

1. Quelle est la probabilité que le poids total excède la capacité du bateau ?
2. L'agence s'apprête à louer un deuxième bateau, toujours pour 50 clients. Quelle doit être sa capacité maximale si l'agence souhaite que le dépassement de la capacité maximale n'arrive qu'avec une probabilité de 10% ?

Exercice 2

Dans un service automatisé, le temps pour préparer une commande est en moyenne de 70 secondes, avec un écart-type de 15 secondes. Sur un échantillon de 100 commandes, quelle est la probabilité que la moyenne de l'échantillon dépasse 71 secondes ?

Exercice 3

Une compagnie aérienne remarque que 5% des personnes ayant effectué une réservation ne se présentent pas le jour du vol. Cette compagnie vend 160 billets alors qu'il n'y a que 155 places dans l'avion. Le jour du vol, quelle est la probabilité que tous les individus présents puissent s'asseoir dans l'avion? Répondez à cette question par *deux approches différentes.*

Excercice 4

Soit X_1, X_2 et X_3 trois variables aléatoires indépendantes et identiquement distribuées de loi $N(\mu, 12)$, où μ est la moyenne inconnue de cette distribution. Soit $\overline{X} = \frac{1}{3}(X_1+X_2+X_3)$. Calculer $P(|\overline{X} - \mu| < 3)$.

Excercice 5

Une équipe de zoologues prépare une expédition pour mesurer la taille des manchots empereurs. Des études précédentes montrent que dans la population des manchots, la taille a un écart-type σ de 5cm. Cependant, la moyenne (de la population) μ est inconnue. Combien de manchots faudrait-il échantillonner pour que $|\overline{X} - \mu|$, càd. la distance entre la moyenne de l'échantillon $\overline{X}$ et μ, soit inférieure à 1cm avec une probabilité de 95% ?

Excercice 6

Un institut de sondage veut effectuer une enquête pour déterminer la proportion d'individus soutenant la loi climat. Quelle taille d'échantillon (n) doit-il viser pour que, *dans le pire des cas*, la distance entre la proportion d'individus soutenant la loi climat dans la population (p) et dans l'échantillon soit inférieure à 0.01 avec une probabilité de 95%. [Tuyau]: Vous devez d'abord exprimer n en fonction de p et puis maximiser cette fonction en prenant la "pire" valeur possible de p.

Excercice 7

Soit $X_1, \ldots, X_n$ des variables indépendantes de distribution $Pois(2)$. Soit

$$T_n = X_1 + X_2 + \ldots + X_n.$$

1. On sait montrer, par des calculs théoriques, que quel que soit n,

   $$T_n \sim Pois(2n)$$

   On vous propose de vérifier ce résultat pour $n = 3$, à l'aide de simulations en R. Pour cela,

   a. Générez trois observations d'une $Pois(2)$ et prenez leur somme.
   b. Répétez cette opération 10^6 fois.
   c. Calculez les proportions des valeurs ainsi obtenues.
   d. Faites un graphique pour comparer ces proportions aux probabilités théoriques, càd celles d'une $Pois(6)$.

D'après le TCL, puisque $E(X_i) = Var(X_i) = 2$, on sait que, pour n suffisamment grand,

$$T_n \sim_a N(2n, 2n)$$

Par la suite, nous souhaitons vérifier cette approximation.

2. On commence par $n = 3$.

   a. Générez trois observations d'une $Pois(2)$ et prenez leur somme.
   b. Répétez cette opération 10^6 fois.
   c. Calculez la moyenne et la variance des 10^6 valeurs ainsi obtenues. Que pouvez-vous déjà constater?
   d. Tracez l'histogramme des 10^6 sommes obtenus en (b). Superposez à l'histogramme obtenu la courbe de la densité (théorique) d'une $N(2n, 2n)$. Que pouvez-vous constater?

3. Refaire les questions 2 pour $n = 10$ puis pour $n = 50$. Que constatez-vous?

Conformité à la loi normale

Excercice 8

Cet exercice porte sur le vecteur `Revenus` contenu dans ce fichier `Revenus.Rdata` disponible, pour téléchargement, ici. Ce vecteur donne le salaire mensuel d'un échantillon de 1000 individus. Commencez donc par charger ce fichier dans votre session `R`.

1. En inspectant uniquement la moyenne et la médiane, que pouvez-vous dire sur l'asymétrie de la distribution ?
2. Affinez vos conclusions en examinant un boxplot et un histogramme. Une distribution normale pourrait-elle être adéquate pour modéliser le revenu ?
3. Affinez davantage vos conclusions à l'aide d'un QQlot.
4. Refaites la questions 1-3 mais cette fois ci pour (i) $W = \sqrt{\text{Revenus}}$ et (ii) $Z = \log(\text{Revenus})$. Tirez une conclusion générale

Excercice 9

1. Générez 10000 échantillons, chacun de taille $n = 1$, d'une $Unif(0, 1)$. Calculer les 10000 moyennes puis comparer leur distribution avec la distribution normale à l'aide d'un QQplot.
2. Refaire cette analyse pour différentes valeurs de n, par exemple $n = 5, 10, 20, 100$. Que constatez-vous?
3. Quelle taille d'échantillon est nécessaire pour obtenir une distribution proche d'une distribution normale ? Pour cette dernière question, il n'y a pas de réponse unique, mais il faudra justifier votre démarche.