Licence Informatique S4
Probabilités-Statistiques TP 2

30 mars 2020

1 Répartition observée et répartition théorique

1. Utiliser `help` de R pour comprendre les commandes `dbinom, rbinom`. Expliquer les résultats obtenus quand on exécute `dbinom(5, 10,0.5)`, `dbinom(5, 4,0.8)`, `dbinom(5, 5,1)`, `rbinom(10, 5, 0.8)`. **Indication.** Dans `http://pages.stat.wisc.edu/~larget/R/prob-R.pdf` vous trouverez quelques indications utiles au sujet de cet exercice.
2. Simuler $N = 10$ lancer d'une pièce équilibrée en utilisant R. Cela signifie qu'il faut produire, en utilisant les commandes de R, une liste de 1 (=pile) et de 0 (=face), qui est aléatoire. Varier la valeur de N. Vérifier que le nombre de "piles" vaut à peu près $(1/2)N$ lorsqu'on prend N de plus en plus grand. Si la probabilité d'obtenir "pile" est $2/3$ plutôt que $1/2$, comment faut-il changer vos commandes ? Et comment peut-on toujours vérifier que cette fois-ci le nombre de "piles" vaut à peu près $(2/3)N$?
3. Utiliser `help` de R pour comprendre les commandes `dunif, runif`. Expliquer les résultats obtenus quand on exécute `dunif(5, 0,10)`, `dunif(5, 0,4)`, `dunif(5, 0,20)`, `runif(10, -2,3)`.
4. Simuler $N = 100$ réalisations de la loi uniforme sur $[-2,3]$. Tracer l'histogramme correspondant. Augmenter N pour voir comment l'histogramme évolue.
5. Simuler $N = 10$ jets de dé en utilisant la commande `runif()` de R. Tracer l'histogramme correspondant, en choisissant bien les classes de façon à avoir une présentation intuitivement claire. Augmenter N et observer l'évolution de l'histogramme.
6. Tracer l'histogramme de ces réalisations et sur le même graphique, la densité de la loi uniforme sur $[1,7]$.

7. Reprendre les questions qui précèdent avec cette fois-ci, respectivement, la loi binomiale, la loi exponentielle de paramètre α (que vous ferez varier) et la loi normale $\mathscr{N}(-1,4)$ (4 étant la variance).

2 Loi forte des grands nombres

Lorsque $X_1, \ldots, X_N$ sont des v.a. indépendantes de même loi admettant une espérance $\mu = \mathbb{E}(X_1)$ alors on a avec probabilité 1 (on dit aussi presque sûrement)

$$\frac{S_N}{N} = \frac{X_1 + \cdots + X_N}{N} \xrightarrow[N \to +\infty]{} \mu.$$

Illustration pour la loi uniforme sur [0,1].

1. Commencer par simuler un vecteur de M variables aléatoires de loi uniforme sur [0,1] ($M = 2000$) $(X_1, ..., X_M)$.
2. Construire le vecteur des sommes successives, c'est à dire le vecteur $(S_1, ..., S_M)$, où $S_N = X_1 + ... + X_N$.
   Astuce R : regarder la fonction "cumsum"
3. Afficher maintenant $\frac{S_N}{N}$ en fonction de N.
   Astuce R : regarder la fonction "plot"
4. Rajouter sur le graphique une ligne horizontale d'ordonnée $E[X_1]$.
   Astuce R : vous pouvez utiliser "abline"
5. Pourquoi y a-t-il de fortes oscillations lorsque N est petit ?
6. Pourquoi le graphe change à chaque fois que l'on compile le code ?

Illustrations pour d'autres lois.

7. Reprendre les questions 1 à 4 pour une loi uniforme sur $[a, b]$. Faire varier a et b et observer comment le graphe change.
   Astuce R : on peut créer une fonction qui prend a, b et N en arguments
8. Reprendre les questions 1 à 4 pour une loi exponentielle de paramètre $\alpha > 0$. Faire varier α et observer comment le graphe change.
   Astuce R : on peut créer une fonction qui prend alpha et N en arguments
9. La loi de Cauchy, qui a pour densité $\rho(x) = \frac{1}{\pi(1+x^2)}$, n'admet pas d'espérance. Reprendre les questions 1 à 4 pour la loi de Cauchy, que constate-t-on graphiquement ? (exécutez plusieurs fois votre code pour voir différentes simulations).
   Astuce R : loi uniforme avec "runif", loi exponentielle avec "rexp", loi de Cauchy avec ...

3 Théorème de la limite centrale (TLC)

Dans l'exercice sur la loi des grands nombres nous avons illustré le fait que pour $X_1, ..., X_N$ des variables indépendantes de même loi et d'espérance $\mu = E[X_1]$ nous avons $\frac{S_N}{N} \to \mu$ lorsque $N \to +\infty$, où $S_N = X_1 + ... + X_N$. Autrement dit, lorsque N est grand, $\frac{S_N}{N} \approx \mu$. Mais vous avez dû voir sur vos simulations que $\frac{S_N}{N}$ est aléatoire autour de μ. Le Théorème Centrale Limite (TCL) donne une description de la façon dont $\frac{S_N}{N}$ fluctue autour de μ. Grosso modo, le TCL nous dit que lorsque N est grand

$$\frac{S_N}{N} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right),$$

où σ est l'écart-type de la variable aléatoire X_1. C'est à dire que $\frac{S_N}{N}$ suit à peu près une loi normale de valeur moyenne μ et d'écart type $\frac{\sigma}{\sqrt{N}}$.

Le but de cet exercice est d'illustrer ce comportement. Pour cela, nous allons fixer N assez grand ($N = 100$ par exemple), simuler un grand nombre de fois S_N/N (M fois, avec $M = 10000$ par exemple) et en faire un histogramme. Nous comparerons ensuite cet histogramme avec la densité de la loi normale $\mathcal{N}(\mu, \sigma^2/N)$.

Illustration pour la loi uniforme sur [0,1].

1. Faire une matrice avec N lignes et M colonnes dont les entrées sont des variables aléatoires de loi uniforme sur [0,1].
2. Construire, à partir de la matrice, un vecteur dont la nième coordonnée est la moyenne des valeurs de la nième colonne de la matrice (c'est donc la moyenne de N variables aléatoires uniformes).
   Astuce R : Pas besoin de faire des boucles, regarder les fonctions disponibles dans "help(rowSums)" ?
3. Faire un histogramme avec les valeurs obtenues.
4. Ajouter sur l'histogramme la courbe de la loi normale de valeur moyenne μ et d'écart type $\sigma/\sqrt{N}$.
   Astuce R : la formule de la loi normale peut être longue à taper, on peut facilement ajouter la courbe avec "curve(dnorm(x,mean,sd), add=TRUE)" ?

Illustrations pour d'autres lois

5. Reprendre les questions 1 à 4 avec des variables de loi uniforme sur un intervalle $[a,b]$.
   Astuce R : on peut créer une fonction qui prend N, M, a et b en arguments
6. Reprendre les questions 1 à 4 avec des variables de loi exponentielle de paramètre $\alpha > 0$.
   Astuce R : on peut créer une fonction qui prend N, M et $alpha$ en arguments

7. Reprendre les question 1 à 3 pour la loi de Cauchy. Faire varier N, que remarque-t-on ? Essayer d'identifier la loi de la moyenne de variables aléatoires de Cauchy.

4 Étudier la formule du transfert

1. Lire l'entrée correspondante dans le Dictionnaire de ProbaStat. Exécuter le code qui y est donné. Varier les paramètres N et a. Que se passe-t-il quand on remplace la loi uniforme sur $[-a,a]$ par une loi uniforme sur $[0,a]$. Pourquoi ?
2. On suppose à nouveau que X suit une loi uniforme sur $[-a,a]$. Remplacer dans le code $Y = X^2$ par $Y = X^3$. Observez l'histogramme pour Y. Pouvez-vous intuitivement expliquer son allure ?
3. On souhaite trouver la densité de probabilité $\rho_Y(y)$ de Y dans ce cas-ci. On supposera qu'elle est de la forme $\rho_Y(y) = \frac{A}{a}\frac{1}{|y|^B}$, avec A et B des nombres positifs. Par tatonnement, trouver B et A. Ils ne dépendent pas de a.
4. On suppose maintenant que X suit une loi exponentielle de paramètre α. Simuler la loi de Y^2 pour différentes valeurs de α. On peut montrer que, dans ce cas, on a
$$\rho_Y(y) = \frac{\alpha}{\sqrt{y}}\exp(-\alpha\sqrt{y}).$$
Tester cette affirmation numériquement. **Indication :** Pour y voir clair, il va falloir ajuster les classes de l'histogramme et l'affichage des courbes à la valeur de α. Sinon vous ne verrez pas grand chose.

   Complément. Les formules pour $\rho_Y(y)$ peuvent être expliquées de la façon suivante. Supposons que $Y = f(X)$ et, pour faire le plus simple possible, que f est une fonction strictement croissante de X. Comme par exemple $f(X) = X^3$ ou encore $f(X) = X^2$, lorsque $X > 0$. Il en résulte que, pour tout x,
$$P(Y \le f(x)) = P(X \le x).$$
Par ailleurs
$$P(Y \le f(x)) = \int_{-\infty}^{f(x)} \rho_Y(y)\mathrm{d}y.$$
Changeons de variable dans cette dernière intégrale en posant $y = f(t)$. Alors $\mathrm{d}y = f'(t)\mathrm{d}t$ et donc
$$P(Y \le f(x)) = \int_{-\infty}^{x} \rho_Y(f(t))f'(t)\mathrm{d}t.$$
On conclut de la première et de la dernière de ces équations que, toujours pour tout $x \in \mathbb{R}$,
$$P(X \le x) = \int_{-\infty}^{x} \rho_Y(f(t))f'(t)\mathrm{d}t.$$

Mais, par ailleurs,

$$P(X \leq x) = \int_{-\infty}^{x} \rho_X(t)\mathrm{d}t.$$

Donc, pour tout $x \in \mathbb{R}$,

$$\int_{-\infty}^{x} \rho_X(t)\mathrm{d}t = \int_{-\infty}^{x} \rho_Y(f(t))f'(t)\mathrm{d}t.$$

En prenant la dérivée des deux cotés de cette équation par rapport à x on trouve

$$\rho_X(x) = \rho_Y(f(x))f'(x), \quad \text{ou} \quad \rho_Y(f(x)) = \frac{\rho(x)}{f'(x)}.$$

Sous cette forme, la formule paraîencore un peu mystérieus. En posant $y = f(x)$, on trouve

$$\rho_Y(y) = \frac{\rho(x)}{f'(x)}, \quad \text{où} \quad y = f(x).$$

5 Un autre exemple de convergence

Considérons $X_1,...,X_N$ des variables indépendantes de loi uniforme sur $[a,b]$. On note $M_N = \max(X_1,...,X_N)$ la plus grande des variables et $m_N = \min(X_1,...,X_N)$ la plus petite.

1. Selon vous, est-ce que M_N et m_N vont converger lorsque N va tendre vers l'infini ? Si oui vers quelles valeurs ?
2. En vous inspirant de l'exercice sur la loi forte des grands nombres, illustrer les comportements de M_N et m_N lorsque N tend vers l'infini.
   Astuce `R` : quel pourrait être l'équivalent de "cumsum" pour le minimum et le maximum ?

6 Un autre exemple de loi limite

On considère $X_1,...,X_N$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$, on introduit leur maximum : $M_N = \max(X_1,...,X_N)$. Vous avez normalement illustré dans la partie 3 le fait qu'avec probabilité 1 on a

$$M_N \xrightarrow[N\to+\infty]{} 1.$$

Néanmoins on a toujours des fluctuations aléatoires de M_N en dessous de 1. On peut montrer que si N est grand, $1 - M_N$ suit approximativement une loi exponentielle.

En vous inspirant de l'exercice précédent, illustrer ce phénomène, en particulier identifier le paramètre de la loi exponentielle, comment dépend-t-il de N ?

Complément théorique [Si vous vous ennuyez]
Pour $x \geq 0$, calculer $P(N(1-M_N) \leq x)$ et regarder la limite lorsque $N \to +\infty$. En déduire que $1-M_N$ suit approximativement une loi exponentielle lorsque N est grand, et trouver son paramètre.