TP Statistiques 2

Juhyun Park, Phuong Thuy Vo, Atef Lechiheb

1 mars 2024

Simulation et convergence

Simulation dans R

Ex1. Si $X \sim \mathcal{N}(\mu = 20, \sigma^2 = 60)$, quelle est la probabilité qu'on observe une valeur supérieure à 35? Hint: vous pouvez utiliser les fonctions internes de densité et de répartition dans R pour évaluer les quantités théoriques.

```
?rnorm
# list of available distributions
?Distributions
```

Ex2. Simuler un échantillon de taille $n = 20$ d'un loi de X. Obtenez la probabilité empirique que la valeur dépasse 35. Créez un histogramme de votre échantillon et commentez la forme de votre histogramme. Superposer la vrai densité.

```
n =
x = rnorm(n, mean=, sd=)
phat =
hist(x, probability = TRUE, main=paste("Emp.prob=", phat))
## evaluate normal density at grid xx
xx = seq()
yy = dnorm(xx, mean=, sd=)
## add density plot
lines(xx, yy, col="red")
abline(v=35, col="gray", lty=5)
```

Ex3. Répétez cette opération 10 fois et stockez les données dans la matrice ou data frame (disons, `datmat`) Calculez la probabilité empirique pour chaque ensemble de données. Sont-elles proches de la valeur théorique ? Commentez les différences entre les histogrammes que vous obtenez à chaque fois. Utilisez la même limite sur les axes pour faciliter la comparaison.

```
## plots in multi panel
ncol(datmat) #
par(mfrow=c(2,2)) # 2 x 2 panel
for (j in 1:4)
{
  hist(datmat[,j])
}
```

Ex4. Augmentez la taille de votre échantillon à 50 et répétez votre expérience. Que remarquez-vous?

Moyenne et phénomène de concentration.

Ex5. Pour quantifier théoriquement l'erreur attendue dans l'expérience précédente, définir $\theta = P(X > 3.5)$. Montre que l'estimateur $\hat{\theta}$ de la probabilité empirique peut être expriméé comme la moyenne des variables aléatoires Y_i, en définissant Y_i et en identifiant sa loi.

Ex6. Utilisant l'inégalité de Bienaymé Chebychev, donner une borne de quantité

$$P(|\hat{\theta} - \theta| \geq \delta)$$

De combien d'échantillons auriez-vous besoin pour que cette quantité pour $\delta = 1.5\sigma$ soit inférieure à 0.1 ?

Théorème Central Limite et Estimation Monte Carlo (à soumettre)

Pour $X \sim \text{Pareto}(a, \alpha)$, la densité est $f(x; a, \alpha) = \alpha \frac{a^\alpha}{x^{\alpha+1}} 1_{[a,+\infty[}$.

Ex7. Vérifier que l'espérance théorique d'une loi de Pareto est $E[X] = \frac{\alpha a}{\alpha - 1}$ si $\alpha > 1$ et ∞ si $\alpha \leq 1$ (avec la formule $\int_0^\infty P(X > t)dt$). On rappelle que la variance d'une Pareto est $V(X) = \left(\frac{\alpha a}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2}$ si $\alpha > 2$ et ∞ si $\alpha \leq 2$.

Ex8. Simuler $B = 500$ échantillons i.i.d de loi commune Pareto(a, α) (avec votre choix de paramètres) de taille $n = 20, 100, 200$ et calculer les moyennes et variances empiriques $\bar{X}_{n,i}$ et $S_{n,i}, i = 1, \ldots, B$.

Ex9. Tracer l'histogramme des moyennes empiriques.

Ex10. A l'aide d'une renormalisation adéquate (a_n, b_n), montrer que $U_{n,i} = \frac{\bar{X}_{n,i} - a_n}{b_n}$ a une loi que vous pouvez approchez. Comparez histogramme de les moyennes empiriques normalisées, $U_{n,i}$, et *distribution théorique approchée*. Quelle est l'influence de la taille de l'échantillon n sur la qualité de cette approximation?

Quand le théorème de central limite ne s'applique pas (à soumettre)

La densité de la loi de Cauchy $\mathcal{C}(\theta)$ est $f(x, \theta) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$ pour tout $x \in \mathbb{R}$.

Ex11. Simuler un échantillon de taille $n = 50$ d'une loi de $\mathcal{C}(2)$ et calculer la moyenne empirique $\bar{X}_n$.

Ex12. Faites varier la taille de l'échantillon $n = 100, 1000$ et 10000. Qu'en déduire ?

Ex13. Expliquer ce comportement. On se rappellera notamment que la fonction caractéristique s'écrit $\phi_\theta(t) = \exp(i\theta t - |t|)$.

Ex14. Quelle est la médiane d'une loi de Cauchy $\mathcal{C}(\theta)$?

Ex15. En déduire un estimateur de θ et evaluer la performance de ce estimateur de médian pour $n = 20, 100, 1000$.