Bootstrap TP 1

Partie 1 : estimateurs bootstrap

Estimation de la variance

Jeu de données "law" : variance de la corrélation

```
law=as.data.frame(read.table("./lsat.dat",header=TRUE))
```

Donner un estimateur bootstrap de la variance de la corrélation pour le jeu de données "law school".

Médiane

- Simuler $n = 10$ réalisations de $Z \sim \mathcal{E}(1)$.
- Calculer la médiane obtenue sur les réalisations puis estimer sa variance par bootstrap. Recommencer en faisant grandir n et comparer à la variance limite.

Estimation de la fonction de répartition

Pour $n = 10$ réalisations de $Z \sim \mathcal{E}(1)$, comparer sur simulation

- la fonction de répartition de la médiane empirique (approchée par Monte Carlo)
- la fonction de répartition de l'approximation asymptotique gaussienne
- la fonction de répartition empirique des estimateurs bootstrapés.

Partie 2 : intervalles de confiance et tests par bootstrap

Soit $X_1, \dots, X_n$ un échantillon i.i.d. de loi gamma $\Gamma(p, \theta)$ et $Y_1, \dots, Y_n$ un échantillon i.i.d. de loi gamma $\Gamma(q, \kappa)$. On veut construire un intervalle de confiance pour le paramètre

$$\tau = \frac{\mathbb{E}(X)}{\mathbb{E}(Y)} = \frac{p\kappa}{q\theta}.$$

On utilise pour cela la statistique

$$T(X_1, \dots, X_n, Y_1, \dots, Y_n) = \frac{\bar{X}_n}{\bar{Y}_n} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i} = \hat{\tau}.$$

On choisit les paramètres

```
n=10
p=0.7
theta=0.007
q=1
kappa=0.02
alpha = 0.05
tau = (p * kappa) / (q *theta)
```

- Créer une fonction qui simule les échantillons $X_1, \dots, X_n$ et $Y_1, \dots, Y_n$.

Intervalle de confiance par approximation normale

On sait, par le Théorème Central Limite, que :

$$\sqrt{n}\left(\bar{X}_n - p/\theta\right) \overset{Loi}{\rightsquigarrow} U \sim \mathcal{N}(0, p/\theta^2) \text{ et } \sqrt{n}\left(\bar{Y}_n - q/\kappa\right) \overset{Loi}{\rightsquigarrow} V \sim \mathcal{N}(0, q/\kappa^2).$$

Par ailleurs, la delta-méthode assure que

$$\sqrt{n}\left(\frac{\bar{X}_n}{\bar{Y}_n} - \frac{p\kappa}{q\theta}\right) = \sqrt{n}\left(\hat{\tau} - \tau\right) \overset{Loi}{\rightsquigarrow} \frac{\kappa}{q} U - \frac{p}{\theta}\left(\frac{\kappa}{q}\right)^2 V.$$

- Quelle est la loi limite de $\sqrt{n}\left(\hat{\tau} - \tau\right)$?
- En déduire que l'intervalle de confiance asymptotique par approximation normale au niveau de confiance $1 - \alpha$ pour τ s'écrit :

$$IC_{\text{AN}}(\alpha) = \left[\hat{\tau} - \frac{\hat{\sigma}}{\sqrt{n}}\Phi^{-1}(1-\alpha/2); \hat{\tau} + \frac{\hat{\sigma}}{\sqrt{n}}\Phi^{-1}(1-\alpha/2)\right]$$

  où

$$\hat{\sigma} = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\bar{Y}_n^2} + \frac{\bar{X}_n^2}{\bar{Y}_n^4}\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}.$$

Intervalles de confiance par bootstrap

Les IC par bootstrap

- Ecrire une fonction qui calcule les intervalles de confiance du bootstrap basique et du percentile pour τ au niveau de confiance $1 - \alpha$ (les fonctions utiles sont `sample` et `sort(v)`.
- De même pour l'intervalle de confiance du t-bootstrap pour τ au niveau de confiance $1 - \alpha$.

Comparaison des intervalles de confiance

- Écrire qui calcule les bornes inférieures et supérieures des différents intervalles de confiance avec $B = 1000$ tirages pour le bootstrap.
- Evaluer les performances des différents intervalles de confiance en terme de couverture et de longueur par Monte Carlo ($M = 1000$).

Tests par bootstrap

On veut tester $\mathcal{H}_0 : \tau = 2$.

- Calculer les niveaux des tests par intervalles de confiance
- Estimer la puissance de ces tests sous $\mathcal{H}_1 : \tau = 3$ on prendra :

```
p=0.7
theta=0.007
q=1
kappa=0.03
alpha = 0.05
```