

TP: BOOTSTRAP NON-PARAMÉTRIQUE

Exercice 1 : moyenne et moyenne bootstrap

Supposons que trois souris pèsent respectivement 82, 107 et 93 grammes.

- i) Quel est le poids moyen des souris ?
- ii) Combien d'échantillons bootstrap possibles peut-on composer ? Les lister.
- iii) Calculer la moyenne de chaque échantillon bootstrap.
- iv) Calculer la moyenne des moyennes reconstituées par bootstrap. Comparer avec la moyenne de l'échantillon d'origine. Quelle conclusion tirez-vous ?
- v) Quel est le minimum / maximum des moyennes rééchantillonnées ?

Exercice 2 : intervalle de confiance d'un estimateur de localisation

Téléchargez le logiciel libre R et installez-le sur votre machine. Installez ensuite la librairie R (package) `bootstrap`, ainsi que la librairie `boot`. Chargez ces deux librairies.

Des résidus d'Aflatoxine se trouvent dans le beurre de cacahuète. Douze lots de beurre de cacahuète contiennent ces résidus avec une concentration respective de 4,94 ; 5,06 ; 4,53 ; 5,07 ; 4,99 ; 5,16 ; 4,38 ; 4,43 ; 4,93 ; 4,72 ; 4,92 ; et 4,96.

- i) Combien d'échantillons bootstrap peut-on créer à partir de cet échantillon d'origine ?
- ii) En utilisant R et la fonction `sample()`, ou bien un générateur de nombre aléatoire (voir par exemple `runif` ou `rmultinom`), générer 5 échantillons d'entiers compris entre 1 et 12 (positions des valeurs dans le vecteur des valeurs de concentration). Cela permet de visualiser aisément la variété des échantillons bootstrap créés.
- iii) Pour chacun de ces échantillons, calculer la moyenne de cette concentration.
- iv) Calculer la moyenne de ces moyennes, et la comparer avec la moyenne de l'échantillon d'origine des valeurs observées.
- v) Trouver le minimum et le maximum des 5 moyennes recalculées. C'est un intervalle de confiance bootstrap sur la moyenne. En utilisant 1000 échantillons bootstrap, donner un intervalle de confiance de la moyenne à 95%.

Exercice 3 : estimation du biais d'un estimateur

On se propose dans cet exercice de calculer le biais d'un estimateur. Considérons une variable aléatoire $\mathcal{N}(\mu, \sigma^2)$, dont l'estimateur du maximum de vraisemblance du paramètre σ^2 basé sur un échantillon de taille n est donné par la variance empirique

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

où \bar{X}_n est la moyenne empirique des observations.

On connaît l'expression exacte du biais de cet estimateur, donné par $b = -\frac{\sigma^2}{n}$.

L'estimation bootstrap du biais vaut b .

B^* est défini tel que

$$B^* = \mathbb{E}[\theta^* - \hat{\theta}],$$

où $\hat{\theta}$ représente S_n^2 et $\theta^* = \frac{1}{n} \sum_{i=1}^n (X_i^* - \bar{X}_n^*)^2$. L'approximation Monte Carlo de B^* vaut

$$B_{MC} = \frac{1}{N} \sum_{j=1}^N B_j^*,$$

où B_j^* est l'estimation bootstrap du biais pour le j^e échantillon bootstrap (N est donc le nombre d'échantillons bootstrap), et $B_j^* = \theta_j^* - \hat{\theta}$ avec

$$\theta_j^* = \frac{1}{n} \sum_{i=1}^n (X_{ij}^* - \bar{X}_n^*)^2.$$

Question : à l'aide du logiciel R, estimer numériquement (avec 10 000 échantillons bootstrap) le biais associé à l'estimateur de la variance d'une loi normale $\mathcal{N}(0, 1)$ pour 25 observations générées. Comparer à la valeur théorique du biais.

Exercice 4 : Estimation de la moyenne et de l'écart-type d'un estimateur

D'après l'US National Transportation Safety Board, l'historique du nombre d'accidents d'avion entre 1983 et 2006 est le suivant : 23, 16, 21, 24, 34, 30, 28, 24, 26, 18, 23, 23, 36, 37, 49, 50, 51, 56, 46, 41, 54, 30, 40 et 31.

- i) Pour cet échantillon historique, calculer la moyenne, l'erreur standardisée ($1/\sqrt{n} \times$ l'écart-type) et la médiane.
- ii) En utilisant R, calculer des estimations bootstrap de la moyenne et de la médiane. Donner les erreurs standardisées de ces estimateurs en considérant 1000 échantillons bootstrap.
- iii) Calculer la médiane des estimateurs de médiane.

N.B. : vous pouvez utiliser la fonction `set.seed()` pour que vos résultats soient reproductibles...