

TP1 : Bootstrap

Marie-Anne Poursat

8 septembre 2025

Préparer l'environnement de travail en créant un fichier qui contiendra votre code. Il est conseillé d'inclure les commandes suivantes pour assurer un environnement vierge et bien initialisé.

```
rm(list=objects())      # supprime les objets existant en session
graphics.off()          # supprime les graphiques existant en session
setwd("MonRepertoire")  # définit le répertoire en cours
```

Partie 1

Les maxima des hauteurs de vague sur une digue ont été relevés plusieurs années consécutives. On suppose que les mesures X_1, X_2, \dots, X_n sont générées par une loi de Rayleigh de densité:

$$f(x) = \frac{x}{a^2} \exp\left(-\frac{x^2}{2a^2}\right), \quad x > 0, \quad a > 0.$$

Les observations sont :

```
x=c(3.1,2.4,2.6,2.2,1.9,2.8,1.1,0.7,2.3,4.3)
n=length(x)
```

Cette densité est codée dans R dans la fonction `drayleigh` du package `VGAM`.

1. Calculez l'estimateur du maximum de vraisemblance \hat{a} du paramètre a .
Représentez graphiquement l'ajustement de la densité de Rayleigh aux données.
2. Générez un échantillon bootstrap des données de départ (à l'aide de la fonction `sample`). Tracez son histogramme. Calculez la valeur de l'estimateur bootstrap sur cet échantillon.
Visualisez plusieurs échantillons bootstrap obtenus par *bootstrap non-paramétrique*.
3. Utilisez la fonction `replicate` pour générer $B = 1000$ échantillons bootstrap et représentez la distribution d'échantillonnage bootstrap de l'estimateur de a .
4. Calculez l'estimation bootstrap du MSE de \hat{a} .
5. Calculez un intervalle de confiance bootstrap de a .
6. Comparez avec l'intervalle de confiance classique (fondé sur l'approximation gaussienne de la loi de \hat{a})

A faire avant lundi 22 septembre 8h

Vous déposerez sur eCampus un script R nommé **Nom.R** comportant un code

- opérationnel, c'est à dire que lorsque je le lancerai il marchera sans intervention de ma part et reproduira vos résultats,
- commenté avec les **explications méthodologiques** de chaque étape,
- avec vos **résultats affichés** en commentaire.
- **La première ligne mentionnera votre nom en commentaire.**

Travail demandé:

1. Lire le chapitre sur le bootstrap du livre de Larry Wasserman *All of statistics*, disponible sur le cours eCampus (ou en ligne <https://upsaclay.focus.universite-paris-saclay.fr/>, avec vos identifiants).
2. Reproduire avec R l'exemple 8.7: calcul de $\hat{\theta}$ et de son écart-type estimé par bootstrap, ainsi que la figure de la loi d'échantillonnage bootstrap de $\hat{\theta}$.

Attention: comme le bootstrap est fondé sur un rééchantillonnage aléatoire, vous ne retrouverez pas des résultats identiques mais similaires, veillez à ce que vos résultats soient reproductibles.

3. Commentez soigneusement votre code en expliquant la démarche et l'objectif de chaque étape.

Expliquez la construction de l'intervalle de confiance bootstrap du paramètre d'intérêt (paramètre de *bioéquivalence*). Affichez les résultats obtenus explicitement en commentaire du code et interprétez-les..