

Illustrations : Cas général de valeurs extrêmes

Les jeux de données pertinents pour cette session d'illustrations sont :

- Des données de température à Perth, Australie, voir `Perth_min_temp.csv` et `Perth_max_temp.csv`,
- Des données anémométriques (relatives au vent) pour douze stations en République d'Irlande, accessibles par le package R `gstat`.

Vous pouvez également utiliser les données illustrant la première session.

Dans chaque cas, il est intéressant de coder soi-même les méthodes et de se comparer aux procédures classiques des packages `evir`, `evd`, `evt0` et `extRemes`. Vous pourrez d'abord :

1. Implémenter les estimateurs de maximum de vraisemblance pour une loi GP. Le choix de la méthode d'optimisation (descente de gradient, Newton-Raphson, Nelder-Mead, brute force...) est libre, mais on vérifiera qu'on obtient bien un maximum de la vraisemblance. On pourra démarrer l'algorithme par des estimateurs de moments pondérés.
2. Implémenter une méthode de diagnostic (QQ-plot...)
3. Télécharger quelques packages, par exemple `evir`, `evd` and `extRemes`, et les comparer avec vos implémentations.

1. Données de température

Les variables d'intérêt sont les minima et maxima mensuels de température à Perth, Australie, entre 1944 et 2019. Ces données sont accompagnées du mois et de l'année d'enregistrement de la donnée.

1. Représenter les données en ignorant la structure temporelle puis en tenant compte de cette structure. Y a-t-il de la saisonnalité, de l'autocorrélation, de la non-stationnarité ?
2. Peut-on proposer un modèle statistique raisonnable pour la totalité de la distribution des données ?

3. Estimer l'indice des valeurs extrêmes des données en justifiant le choix du/des paramètres utilisés.
4. Calculer un estimateur de quantile extrême de la variable d'intérêt aux niveaux 0.99, 0.995 et 0.999. Qu'en pensez-vous ?
5. Peut-on calculer un intervalle de confiance pour l'indice des valeurs extrêmes et les quantiles extrêmes ? Si oui, pourquoi et comment ? Sinon, peut-on suggérer une méthode qui le permettrait ?

2. Données d'anémométrie

La variable d'intérêt est la vitesse moyenne journalière du vent à certaines stations météo en Irlande. On choisira une station pour l'analyse, mais il est intéressant de comparer les résultats pour plusieurs stations. La date d'enregistrement de la donnée est également incluse. On obtiendra ces données en chargeant le package `gstat` et en exécutant `data(wind)`.

1. Représenter les données en ignorant la structure temporelle puis en tenant compte de cette structure. Y a-t-il de la saisonnalité, de l'autocorrélation, de la non-stationnarité ?
2. Peut-on proposer un modèle statistique raisonnable pour la totalité de la distribution des données ?
3. Estimer l'indice des valeurs extrêmes des données en justifiant le choix du/des paramètres utilisés.
4. Calculer un estimateur de quantile extrême de la variable d'intérêt aux niveaux 0.995, 0.999 et $1 - 1/n$. Qu'en pensez-vous ?
5. Peut-on calculer un intervalle de confiance pour l'indice des valeurs extrêmes et les quantiles extrêmes ? Si oui, pourquoi et comment ? Sinon, peut-on suggérer une méthode qui le permettrait ?