

TP no 2 - Méthodes de Monte Carlo

Master parcours SSD - UE Statistique Computationnelle

1 Méthodes MC pour l'estimation

Exercice 1 (Comparaison d'estimateurs)

Dans cet exercice nous allons évaluer la robustesse des estimateurs décrits dans le cours pour estimer la moyenne d'une loi normale quand on s'éloigne des hypothèses de normalité. Pour cela, nous allons générer des observations selon une loi normale $\mathcal{N}(0, \sigma_1)$ "contaminée" par seconde loi normale de variance différente $\mathcal{N}(0, \sigma_2)$, où $\sigma_2 > \sigma_1$, selon le modèle suivant :

- avec probabilité $p \in [0, 1]$, l'observation courante est tirée selon la loi "nominale" $\mathcal{N}(0, \sigma_1)$,
- avec probabilité $(1 - p)$, elle est tirée selon la loi "contaminante" $\mathcal{N}(0, \sigma_2)$.

Le paramètre $p \in [0, 1]$ contrôle l'ampleur de cette contamination.

1. Implémenter une procédure pour simuler des observations selon ce modèle de mélange.
2. En se basant sur la procédure décrite dans le cours, représenter l'évolution de la précision (quantifiée en terme d'erreur quadratique moyenne) des différents estimateurs de la moyenne proposés (moyenne empirique, moyenne empirique "trimmée" et médiane) pour $\sigma_1 = 1$, $\sigma_2 = 10$, et $p \in \{1; 0.95; 0.9; 0.8; 0.7\}$. Comment interpréter ces résultats ?
 - Question subsidiaire : que se passe-t-il si on fait varier p de 0 à 1 ?

On basera les estimations sur des n -échantillons de taille 20 et un schéma MC de $m = 1000$ répétitions.

Exercice 2 (Estimation d'un niveau de confiance)

Dans cet exercice, nous allons évaluer empiriquement la couverture de l'intervalle de confiance associé à l'estimation de la variance d'une loi normale, quand on s'éloigne des hypothèses de normalité. Pour cela nous allons considérer le modèle de loi contaminée introduit dans l'exercice précédent.

Reprendre la procédure décrite dans le cours pour estimer un intervalle à 95% de la variance d'intérêt σ_1^2 quand $\sigma_2 \in \{2; 5; 10\}$ et $p \in \{1; 0.95; 0.9; 0.8; 0.7\}$.

On basera les estimations sur des n -échantillons de taille 20 et un schéma MC de $m = 1000$ répétitions.

2 Méthodes MC et tests d'hypothèses

Exercice 3 (risque de première espèce empirique)

On suppose que X suit une loi normale $\mathcal{N}(\mu, \sigma)$ de moyenne et variance inconnues. On veut tester l'hypothèse $H_0 : \mu = \mu_0 = 500$ contre l'hypothèse alternative $H_1 : \mu = \mu_1 > 500$ à un niveau de significativité $\alpha = 0.05$.

1. Quelle statistique de test convient-il d'utiliser ?
2. Vérifier empiriquement que le risque de première espèce est bien celui attendu, en considérant des échantillons de taille $n = 20$ et $m = 10000$ itérations de la procédure de MC décrite dans le cours. Pour les simulations, nous prendrons (arbitrairement) $\sigma = 100$.

Exercice 4 (puissance empirique)

Reprendre la configuration de l'exercice 3 et estimer la puissance de détecter l'hypothèse alternative quand μ_1 varie de 500 à 700 (par pas de 10).

Reproduire cette analyse pour des valeurs croissantes de n (50, 100 et 200).