Statistique computationnelle

Série d'exercices #1

Si nécessaire, se familiariser avec les éléments de R. Au besoin, on pourra consulter:

1. Site officiel de R: `http://www.r-project.org/` où l'on trouve logiciels à télécharger, documentation complète, contributions des utilisateurs, etc.
2. Au site `http://statwww.epfl.ch/splus/` où l'on trouve également une documentation abondante.

Voici une liste des notions qu'il est important de maîtriser pour les applications. Toutes ces notions sont couvertes dans la documentation qui précède.

- Propriétés et manipulation des objets de base: vecteurs, matrices, tableaux, listes.
- Opérateurs logiques et opérateurs de comparaison.
- Extraction des composantes d'un objet, que ce soit un vecteur, une matrice, un tableau ou une liste.
- Saisie et lecture de données. Noter que dans R il est préférable de lire un fichier externe avec la fonction `read.table` (plutôt que `scan`).
- Fonctions fondamentales statistiques et mathématiques.
- Fonctions de simulation.
- Création et modification de fonctions.
- Boucles.

–1–

Soit l'échantillon de 5 observations: $-8, 0, 2, 10, 100$. Répondre aux questions suivantes *sans utiliser l'ordinateur*.

1. Enumérer tous les échantillons jackknife.
2. Calculer toutes les réplications jackknife de $\overline{X}$ et de la médiane échantillonnale.
3. Pour chacun des estimateurs du point 2, calculer explicitement l'estimateur du biais de Quenouille et l'estimateur de la variance de Tukey.

–2–

Calculer les estimateurs jackknife de l'écart-type et du biais pour le coefficient de corrélation de Pearson du jeu de données `law` du package `bootstrap`.

N.B. La fonction R pour la corrélation de Pearson est `cor`.

–3–

Engendrer 100 échantillons de taille 20 de la loi $N(\theta, 1)$, où $\theta = 1$.

a) Pour chaque échantillon, calculer l'estimateur jackknife de la variance pour $\widehat{\theta} = \overline{X}$. (Noter que la variance est ici connue.) Calculer ensuite la moyenne et l'écart type de ces 100 estimations.

b) Refaire a) avec $\widehat{\theta} = \overline{X}^2$.

–4–

Supposons qu'un échantillon de taille 20 d'une certaine population produise les observations 3.56, 0.69, 0.10, 1.84, 3.93, 1.25, 0.18, 1.13, 0.27, 0.50, 0.67, 0.01, 0.61, 0.82, 1.70, 0.39, 0.11, 1.20, 1.21, 0.72. On s'intéresse à estimer l'écart type σ de la population, utilisant pour cela $\widehat{\sigma} = \sqrt{\sum_i (X_i - \overline{X})^2/20}$.

N.B. Les données ont été engendrées en simulant 20 valeurs de la variable exponentielle standard. On peut donc calculer σ explicitement à titre d'information.

À l'aide du jackknife, estimer

a) le biais de $\widehat{\sigma}$;

b) l'écart type de $\widehat{\sigma}$;

c) le biais de $\widehat{\sigma}^2$ comme estimateur de σ^2;

d) calculer l'estimateur corrigé pour le biais de $\widehat{\sigma}$ et de $\widehat{\sigma}^2$.

–5–

Soit $\widehat{\theta} = \widehat{\sigma}^2 = \sum_1^n (X_i - \overline{X})^2/n$. Vérifier en détail que

$$\widehat{\theta}_{\text{Jack}} = \sum_1^n (X_i - \overline{X})^2/(n-1).$$

–6–

Les suites grand ordre d'une puissance négative de n ont une arithmétique spéciale. Vérifier que l'on a les identités suivantes:

1. $nO(1/n^3) = O(1/n^2)$;
2. $O(1/n^2) + O(1/n^2) = O(1/n^2)$;
3. $O(1/n^2) + O(1/n^3) = O(1/n^2)$;
4. $nO(1/(n-1)^3) = O(1/n^2)$;
5. Identité de la page 3 des notes de cours:
$$(n-1)\left[O(1/(n-1)^3) - O(1/n^3)\right] = O(1/n^2)$$
6. Identité de la page 4 des notes de cours:
$$-\frac{b}{n(n-1)} + O\left(\frac{1}{n^2}\right) = O\left(\frac{1}{n^2}\right).$$

Jean-Claude Massé
Professeur