UNIVERSITÉ DE NANTES

Master Ingénierie mathématique, Univ. Nantes
Option Mathématiques et applications, ECN

Statistique Inférentielle.

Anne Philippe
Université de Nantes, LMJL

Fiche 3. Régions et intervalles de confiance

Exercice 1. Situation gaussienne

Soit $X_1, \ldots, X_n$ un n-échantillon distribué suivant la loi normale de paramètres (μ, σ^2). On veut estimer μ.

On note
- q_α le quantile d'ordre α de la loi normale standard
- $t_{n,\alpha}$ le quantile d'ordre α de la loi de Student à n degrés de liberté.

si σ^2 est connue,

$$I_n = [\bar{X}_n - \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}; \bar{X}_n - \frac{q_{\alpha/2}\sigma}{\sqrt{n}}]$$

est un intervalle de confiance de niveau $1-\alpha$ pour le paramètre μ c'est à dire

$$P_\mu(\mu \in I_n) = 1 - \alpha$$

si σ^2 est inconnue, on pose

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

et

$$I'_n = [\bar{X}_n - \frac{t_{n-1,1-\alpha/2}\hat{\sigma}_n}{\sqrt{n}}; \bar{X}_n - \frac{t_{n-1,\alpha/2}\hat{\sigma}_n}{\sqrt{n}}]$$

est un intervalle de confiance de niveau $1-\alpha$.

1) Simuler $N = 1000$ échantillons de taille $n = 10$ iid suivant la loi normale de paramètres $(1, 1)$
2) Calculer pour chaque échantillon les régions de confiance I_n et I'_n au niveau 95%
3) Représenter les régions de confiance (on peut se limiter aux 100 premières) et utiliser des couleurs pour distinguer les intervalles qui contiennent $\mu = 1$.

4) Evaluer la fréquence de l'évènement μ *appartient à l'intervalle de confiance*. Commenter le résultat.

5) Reprendre les questions 1) -> 4) avec $n = 50$

Exercice 2. Situation asymptotique

On suppose que $X_1, \dots, X_n$ sont iid, L^2
On construit un intervalle de confiance sur la moyenne à partir de la loi limite de

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1))$$

On obtient un intervalle asymptotiquement de niveau $1 - \alpha$ en prenant

$$I_n(X_1, \dots, X_n) = \left[\bar{X}_n - q_{1-\alpha/2} \frac{S_n}{\sqrt{n}} \ ; \ \bar{X}_n - q_{\alpha/2} \frac{S_n}{\sqrt{n}} \right]$$

où

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

On considère $(X_1, ..., X_n)$ iid suivant la loi de student à 3 ddl. On veut estimer le niveau exact de l'intervalle de confiance $\beta_n = P_\theta(\theta \in I_n)$ en fonction de n par la simulation. On utilise la méthode de Monte Carlo suivante :

étape 1 On simule N échantillons indépendants de taille n suivant la loi de Student à 3 ddl

étape 2 Pour chacun des échantillons, on calcule l'intervalle de confiance : $I_n(i), \quad i = 1, \dots, N$

étape 3 On approche la probabilité β_n par

$$\hat{\beta}_n(N) := \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_{I_n(i)}(\theta)$$

avec N grand.

1) Justifier la méthode

2) Pour n = 15, simuler $N = 10000$ échantillons et représenter $\hat{\beta}_n(N)$ en fonction de N. En déduire une estimation de β_n

3) Reprendre la question précédente pour $n = 25, 50, 100, 500$. Superposer les différentes courbes.

4) Faire le lien entre les résultats numériques et le niveau asymptotique des intervalles de confiance.

Exercice 3. Choix de l'estimateur de la variance

Soit $X_1, \ldots, X_n$ iid suivant la loi exponentielle de paramètre θ^{-1}

— La densité s'écrit $f_\theta(x) = \theta^{-1} e^{-x\theta^{-1}} \mathbb{I}_{\mathbb{R}^+}(x)$

— On a $E_\theta(X) = \theta$ et $\mathrm{Var}_\theta(X) = \sigma^2 = \theta^2$

— Convergence en loi

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \theta) \xrightarrow{loi} N(0,1)$$

On peut estimer la variance σ^2 par les estimateurs consistants suivant :

— la variance empirique S_n^2

— la méthode des moments : on estime $\sigma^2 = h(\theta)$ par $h(\bar{X}_n) = \bar{X}_n^2$

On peut donc construire les intervalles de confiance de niveau asymptotique $1-\alpha$ suivant

$$I_n(X_1, \ldots, X_n) = \left[\bar{X}_n - q_{1-\alpha/2}\frac{S_n}{\sqrt{n}} \; ; \; \bar{X}_n - q_{\alpha/2}\frac{S_n}{\sqrt{n}}\right]$$

ou

$$I'_n(X_1, \ldots, X_n) = \left[\bar{X}_n - q_{1-\alpha/2}\frac{\bar{X}_n}{\sqrt{n}} \; ; \; \bar{X}_n - q_{\alpha/2}\frac{\bar{X}_n}{\sqrt{n}}\right]$$

On note

$$\beta_n = P_\theta\left(\theta \in I_n\right)$$
$$\beta'_n = P_\theta\left(\theta \in I'_n\right)$$

1) Pour $\theta = 1$ et $n = 25, 50, 100, 200$, simuler $N = 10000$ échantillons et superposer les courbes $N \mapsto \hat{\beta}_n(N)$ et $N \mapsto \hat{\beta}'_n(N)$. En déduire des valeurs approchées de β_n et β'_n.

2) Même question pour $\theta = 2$

3) Commenter les résultats obtenus