TP R sur l'estimation

Emmanuel Rachelson and Matthieu Vignes

18 septembre 2013, SupAero - ISAE

1 Estimateurs pour la loi exponentielle

1.1 Simulations: l'EMV

Soit $X_1 \dots X_n$ un échantillon provenant d'une loi de densité exponentielle de paramètre θ notée $\mathcal{E}(\theta)$. Cette loi est utilisée pour modéliser des effets sans mémoire comme le temps entre deux pannes d'un appareil.

1. Regarder la densité en la représentant graphiquement pour différentes valeurs de θ. (aide: `dexp()`) Quelles sont les valeurs prises par une variable exponentielle ?

2. Générer une réalisation d'un échantillon de taille $n = 20$ de loi $\mathcal{E}(2)$. (aide: `rexp()`)

3. Créer une fonction `LV` de la variable θ qui calcule la log-vraisemblance en θ de l'échantillon précédent.

4. Calculer `LV`(θ) pour $\theta = 1.5, 2, 2.5$. Lequel choisir ?

5. Rechercher l'EMV θ_{20} de θ.

6. On refait tout ça avec $n = 10,000$.

7. Visualiser la loi limite de l'EMV. Pour cela, on pourra générer 200 échantillons de taille $n = 20$.

1.2 La médiane

On s'intéresse à la loi $\mathcal{E}(\lambda)$.

1. Générer un échantillon de taille $n = 100$ et paramètre $\lambda = 0.5$.

2. Calculer la médiane de l'échantillon. (aide: `median()`) La comparer à la médiane théorique m. (aide: `qexp()`)

3. refaire $B = 150$ fois cette expérience et stocker les valeurs de $Z_n^b := \sqrt{n}(M_n^b - m)$ dans un vecteur. Estimer la variance V_n de la médiane empirique. Tracer l'histogramme des Z_n^b et superposer à la densité de $\mathcal{N}(0;1)$. Qu'en dire ?

4. Si on suppose ne pas connaître la valeur de λ (donc de m). Estimer la moyenne et l'écart-type de la médiane empirique et tracer l'histogramme des valeurs centrées réduites T_n^b. Qu'obtient-on ? Commentaires ?

2 Données réelles

1. Charger la librarie `MASS` et les données `quine`. (aide: `library()` et `data()`). Dans `help.start()`, aller voir *MASS/quine* pour voir ce que sont ces données. On travaillera avec le nombre de jours d'absence. Les ranger dans un vecteur `obs`.

2. Faire un histogramme et trouver une modélisation *ad hoc* (quelle densité voulez vous considérer ?).

3. En créant une fonction de vraisemblance avec la loi choisie (qui dépend d'un paramètre θ), chercher l'EMV du paramètre. On l'appelera *EQuasiMV*.

4. Retracer l'histogramme du point 2. en superposant la loi que vous avez utilisé.

3 Estimateurs pour la loi uniforme $\mathcal{U}[0;a]$

En adaptant les codes vus ci-avant, étudier comparativement les EMV et l'estimateur de la méthode des moments ($2\bar{X}$ puisque la moyenne est $a/2$) pour le paramètre a.

4 Intervalles de confiance

4.1 Intervalles de confiance exacts (données gaussiennes)

On travaille avec la loi $\mathcal{N}(m = 2, s^2 = 1^2)$, $n = 100$ et on refait $2,000$ fois l'expérience.

1. Stocker les réalisations de la loi gaussienne dans une matrice de taille $B \times n$. (aide: utiliser `rnorm()`)

2. Calculer les moyennes empiriques $\bar{X}_n^b$ des $2,000$ expériences. Que devrait-on approximativement obtenir ?

3. Calculer les variances empiriques V_n^b pour les $2,000$ expériences. Que devrait-on approximativement obtenir ?

4. Soit α la proportion d'indices b pour lequels $m = 2$ tombe dans l'intervalle $I^b = [\bar{X}_n^b + / - \frac{1.64s}{\sqrt{n}}]$. Calculer le quantile d'ordre 0.95 de la loi $\mathcal{N}(0,1)$ et le comparer Ã la valeur observée pour α.

5. Même question avec les intervalles $\tilde{I}^b = [\bar{X}_n^b + / - \frac{1.64\sqrt{V_n^b}}{\sqrt{n}}]$

4.2 Intervalles de confiance approchés (données non gaussiennes)

Refaire le même exercice avec un échantillon de loi $\mathcal{E}(5)$.

Donner un IC pour $1/\theta$ puis pour la variance. (aide: si $X \sim \mathcal{E}(\theta)$, alors $E[X] = 1/\theta$ et $Var(X) = 1/\theta^2$; penser aussi au TCL)

4.3 Influence de n

1. Charger les données `quine`. On travaille avec `quine$Days`. Représenter graphiquement les données. Superposer la droite représentant la moyenne empirique.

2. Chercher un IC (niveau $\alpha = 0.9$) pour la moyenne construit sur les $n = 10$ premières données. Superposer les bandes de confiance en rouge.

3. Recommencer en prenant la moitié des données et mettre les bandes de confiance en vert. Puis en utilisant toutes les données avec les bandes en jaune. Que remarque-t-on ?