Université Joseph Fourier **Frédérique Leblanc**
L2/STA401 **leblanc@imag.fr**

TP6 : Estimation paramétrique

Objectifs : *Etudier numériquement les propriétés d'estimateurs d'un ou plusieurs paramètres de la loi d'une variable continue.*

1 Modèle gaussien

Dans cette partie on considère un échantillon i.i.d. de X pour X de loi normale $\mathcal{N}(\mu, \sigma^2)$ et de taille n où μ et σ sont les paramètres inconnus du modèle.

Exercice 1 : Estimer l'espérance dans le Modèle Normal

Le but ici est l'étude et la comparaison de deux estimateurs du paramètre μ : le premier T_1 est l'estimateur usuel $\bar{X}_n$ et le second que l'on notera T_2 est defini par $T_2 = (\min(X_i) + \max(X_i))/2$.

1. Définir $\mu = 3$, $\sigma = 1$, $N = 1000$ et $n = 10$. Générer N tirages d'échantillons de taille n et les affecter dans une matrice nommée `dataG` à N lignes et n colonnes.

2. Calculer les N réalisations de T_1 et de T_2 obtenues pour les N échantillons et les affecter à `est1` et `est2`.

3. Calculer la moyenne empirique de `est1` (resp. de `est2`) et l'affecter à `moyest1` (resp. de `moyest2`). Calculer les écart-type empiriques corrigés (avec `sd()`) de chacune des deux séries `est1` et `est2` et les affecter à `etest1` et `etest2`. Comparer les résultats et commenter.

4. Au vu des résultats obtenus peut-on dire que T_1 (resp. T_2) est sans biais ? Lequel de ces deux estimateurs a la plus faible variabilité ? En déduire le meilleur de ces deux estimateurs pour μ (meilleur au sens de "celui qui est sans biais et de variance minimale").

5. Représenter pour finir dans une même fenêtre et l'un au dessus de l'autre, les histogrammes des vecteurs `est1` et `est2` et y ajouter en rouge la verticale passant par le points d'abscisse μ et en vert la verticale qui passe par la moyenne empirique de l'échantillon. Les limites des axes des abscisses et des ordonnées devront être les mêmes pour les deux histogrammes, afin qu'ils soient visuellement facilement comparables. Commenter.

6. Refaire ces deux graphiques avec $n = 50$. La différence entre les deux estimateurs est-elle plus nette que lorsque $n = 10$?

7. Revenir à $n = 10$ et calculer les N bornes inférieures et bornes supérieures, de l'intervalle de confiance de niveau de confiance 95% pour μ et basé sur l'estimateur T_1 et en supposant $\sigma = 1$ connu. Affecter les résultats à `Binf` et `Bsup`. Que fait la commande `mean((Binf<=mu)&(mu<=Bsup))` ?

8. Représenter sur un même graphique le nuage des points `(k,Binf[k])` et `(k,Bsup[k])` et relier les deux points obtenu pour chaque valeur de k par un segment vertical (c'est l'intervalle de confiance obtenu pour le kième échantillon de taille n). Ajouter l'horizontale qui passe par l'ordonnée μ. Commenter.

Exercice 2 : Estimer la variance dans le Modèle Normal

On s'intéresse ici aux deux estimateurs usuels de σ^2 que l'on notera comme précédemment $T_1 = S_n^2 = (\sum_{i=1}^n X_i^2)/n - \bar{X}_n^2$ (variance empirique de l'échantillon aléatoire) et $T_2 = S_n'^2 = T_1 n/(n-1)$ (variance empirique corrigée).

1. On garde $\mu = 3$, $\sigma = 1$, $N = 1000$ et $n = 10$. Générer N tirages d'échantillons de taille n et les affecter dans une matrice nommée `dataG` à N lignes et n colonnes.
2. Calculer les N réalisations de T_1 et de T_2 obtenues pour les N échantillons et les affecter à `est1` et `est2` (rappel : utiliser la fonction `var()` pour calculer T_2 et en déduire ensuite le calcul de T_1).
3. Représenter sur une même page les histogrammes des vecteurs `est1` et `est2` et y ajouter en rouge la verticale passant par le points d'abscisse σ^2 et en pointillés vert la verticale qui passe par la moyenne empirique des N réalisations de l'estimateur étudié. Pour le graphique décrivant T_1 on ajoutera aussi en trait plein vert la verticale passant par le point d'abscisse $(n-1)\sigma/n$. Quelles propriétés des estimateurs ces graphiques mettent-ils en évidence (biais, variance et lois des estimateurs) ? Superposer la densité g de la loi théorique de T_1 à l'histogramme précédent qui porte sur T_1 et la densité h sur l'histogramme de T_2. Rappelons que la densité de la loi théorique de T_1 est la fonction $g(t) = n/\sigma^2 f(tn/\sigma_2)$ (ex. à faire en TD) où f désigne ici la densité d'un chi-deux à $n-1$ degrés de liberté puisque nT_1/σ^2 suit une loi $\mathcal{X}_{n-1}$ (la commande R pour calculer sa densité en x est `dchisq(x,df=n-1)`). Celle de T_2 est $h(t) = (n-1)/\sigma^2 f(t(n-1)/\sigma_2)$
4. (facultatif) Afficher dans une fenêtre coupées en 6 morceaux (2 lignes et 3 colonnes) les histogrammes de T_1 obtenus pour les valeurs $10, 20$ et 100 de n en première ligne et ceux obtenus pour T_2 en deuxième ligne. Comparer et commenter.

2 Autres modèles

Exercice 3 : Modèle de Bernoulli

Dans cet exercice on considère un échantillon i.i.d. de X pour X de loi de Bernoulli $\mathcal{B}(p)$ avec $p \in]0, 1[$ et de taille n.

1. Définir $p = 0.5$, $N = 1000$ et $n = 4$. Générer N tirages d'échantillons de taille n et les affecter dans une matrice nommée `dataB` à N lignes et n colonnes.
2. Calculer les N réalisations de l'estimateur usuel de p : $F_n = \bar{X}_n$ obtenues pour les N échantillons de taille n tirés. Affecter le résultat à `est`.
3. Représenter l'histogramme de `est` et y superposer la densité d'une $\mathcal{N}(p, p(1-p)/n)$.
4. Essayer plusieurs valeurs de n (entre 4 et 100) et apprécier graphiquement à partir de quelle valeur de n on peut considérer que la loi de l'estimateur F_n est normale. L'estimateur F_n est-il sans biais quelque soit n ?

Exercice 4 : Modèle Uniforme

Dans cet exercice on considère un échantillon i.i.d. de X pour X de loi de Uniforme $\mathcal{U}[0, a]$ avec $a \in]0, \infty[$ et de taille n.

Avec R la loi de uniforme est décrite avec la fonctions `dunif`.

1. Définir $a = 2$, $N = 1000$ et $n = 10$. Générer N tirages d'échantillons de taille n et les affecter dans une matrice nommée `dataU` à N lignes et n colonnes.

2. Calculer les N réalisations de l'estimateur usuel de a : $T_1 = 2\bar{X}_n$ et $T_2 = \max(X_i)$.

3. Faire une étude analogue à celle proposées dans les exercices précédents permettant de valider numériquement les propriétés mathématiques de ces deux estimateurs. Etude du biais de la variance et de la loi des deux estimateurs proposés.

4. Proposer un troisième estimateur de a : T_3 proportionnel à T_2 et sans biais. Evaluer sa variance. Lequel des trois estimateurs étudiés est-il le plus performant (et en biais et en variance) ?

Exercice 5 : Modèle Exponentiel

Ce modèle est très utile pour modéliser des durées (date de la première panne d'un système, temps nécessaire pour qu'un automate effectue une tâche, durée de vie d'une tumeur, durée sans sinistre déclaré ....) lorsque l'on ne peut pas faire l'approximation par une loi normale. Dans cet exercice on considère un échantillon i.i.d. de X pour X de loi de exponentielle $\mathcal{E}(\lambda)$ avec $\lambda \in]0, \infty[$ et de taille n.

Avec R la loi de uniforme est décrite avec la fonction `dexp`.

1. Représenter sur un même graphique les densité des lois exponentielles de paramètres $\lambda \in 0.1, 0.5, 1, 2, 10$ avec une couleur différente pour chaque λ.

2. Définir $\lambda = 2$, $N = 1000$ et $n = 15$. Générer N tirages d'échantillons de taille n et les affecter dans une matrice nommée `dataE` à N lignes et n colonnes.

3. Calculer les N réalisations de l'estimateur de λ : $T_1 = \bar{X}_n^{-1}$ Quel méthode d'estimation permet-elle d'aboutir à cet estimateur (moments, max de vraissemblance, moindres carrés,...) ?

4. Faire une étude analogue à celle proposées dans les exercices précédents permettant d'étudier numériquement les propriétés mathématiques de T_1 (étude du biais de la variance).