

TP Statistiques 3

Charantonis Anastase & Brunel Nicolas & Julien Floquet

23 Mars 2018

Maximum de vraisemblance

Ce TP a pour objectif de vous familiariser avec la méthode des estimateurs de maximum de vraisemblance. Cette méthode est utilisée pour estimer, pour une loi donnée et un échantillon, les paramètres qui maximisent la probabilité que cette loi ait généré l'échantillon.

1) Ajuster une loi Bernoulli

Soit une loi Bernoulli (`rbinom`) avec $p = 0.7$. Simuler un échantillon i.i.d de taille $N = 10$

- Quelle est une façon simple d'estimer p ?
- Générez une fonction de vraisemblance, nommée `L_bern`, qui donne la vraisemblance d'un échantillon de Bernoulli pour une valeur donnée de p .
- Pour votre échantillon, estimez la vraisemblance de l'échantillon pour 100 lois Bernoulli de paramètres p allant de 0 à 1. Tracez la courbe des valeurs calculées. Que remarquez-vous?
- En utilisant la fonction `optimize` de R, trouvez la valeur de p la plus probable d'avoir généré cet échantillon. (Attention à utiliser l'option `maximum = TRUE` et de borner les valeurs de p en utilisant l'option `interval`).
- Testez avec des échantillons de taille allant de $N = 10$ à $N = 2000$ et comparez l'écart entre la valeur théorique attendue et la valeur obtenue. Que remarquez-vous? Comment combattre l'instabilité numérique due aux multiplications de probabilités?
- Chargez le fichier "distribution_inconnue_2_100_realisations.csv" et trouvez l'estimateur de maximum de vraisemblance si on considère que c'est une loi Bernoulli. Pourquoi cette hypothèse est possible?

2) Ajuster d'une loi normale d'écart type connu

Soit une loi Normale (`rnorm`) avec $\mu = 2$ et $\sigma = 1$. Simuler un échantillon i.i.d de taille $N = 30$

- Générez une fonction de vraisemblance, nommée `L_norm`, qui donne la vraisemblance d'un échantillon pour une valeur donnée de μ . (Voir la fonction `dnorm`).
- Pour votre échantillon, estimez la vraisemblance de l'échantillon pour 100 lois normales de paramètres μ allant de 0 à 4. Tracez la courbe des valeurs calculées. Que remarquez-vous?
- En utilisant la fonction `optimize` de R, trouvez la valeur de p la plus probable d'avoir généré cet échantillon. (Attention à utiliser l'option `maximum = TRUE` et de borner les valeurs de p en utilisant l'option `interval`).
- Testez avec des échantillons de tailles allant de $N = 10$ à $N = 2000$ et comparez l'écart entre la valeur théorique attendue et la valeur obtenue. Utilisez une technique qui permet d'éviter ces instabilités en calculant des sommes.

3) Ajuster d'une loi à plusieurs paramètres

Pour des lois à plusieurs paramètres, résoudre un problème numérique nécessite de trouver un maximum sur une surface. Ceci pourrait être résolu en utilisant des solveurs plus compliqués que “optimize”. Un solveur typique est le solveur mle (maximum likelihood estimator) du package stats4 (install.packages(stats4), puis library(stats4)).

Générez des échantillons issus:

1. de la loi exponentielle translatée avec paramètres $\lambda = 2$ et $L = 4$. (Pdf: $f(x) = \lambda * e^{-\lambda(x-L)}$ avec $x \geq L$)
2. de la loi de Cauchy de paramètres $x_0 = -2$ and $\alpha = 0.4$. (Pdf: $f(x; x_0, \alpha) = \frac{1}{\pi} \frac{\alpha}{(x - x_0)^2 + \alpha^2}$, avec $\alpha > 0$)

Pour ces deux lois, calculez la log-vraisemblance des échantillons que vous avez générés, en faisant varier un paramètre à la fois. Présentez graphiquement la surface de log-vraisemblance que vous calculez, en utilisant la librairie ggplot2. Commentez.

Comparez la valeur obtenue de cette façon à la valeur que vous obtenez en appliquant le solveur mle de la librairie stats4. Commentez.

Effectuez une étude de sensibilité des vos résultats en fonction de la taille de votre échantillon.