Université Paris Dauphine — Année 2004-2005
U.F.R. Mathématiques de la Décision
MD3 - Statistique Inférentielle

Projet de statistique

1 Description

Chaque étudiant, ou binôme, reçoit en TD un numéro de projet. Il doit récupérer le fichier correspondant. Ce fichier contient un vecteur R de 1000 réalisations indépendantes d'une variable aléatoire X de loi inconnue appartenant à une liste donnée.

La rédaction par deux d'un rapport écrit est obligatoire. Une soutenance orale peut étre exigée.

Les traitements suivants sont demandés (la liste n'est pas exhaustive).

1. Tracer l'histogramme et la fonction de répartition empirique des données.
2. Estimer la densité par la méthode du noyau.
3. Identifier le type de loi dont sont issues les observations, en utilisant un graphique de type "droite de Henri". (confère annexe, et exercice 1 du partiel).
4. Estimer le ou les paramétre(s) de la loi présumée.
5. Estimer par intervalles et/ou régions de confiance le ou les paramétre(s), par les méthodes classiques et par Bootstrap.
6. Vérifier la qualité de l'ajustement par un test statistique approprié.
7. Effectuer des traitements spécifiques selon la loi trouvée .

La liste des lois possibles est indiquée ci-dessous.

1. La loi continue uniforme sur l'intervalle (a,b), avec $a < b$.
2. La loi exponentielle de paramétres x_0 et b (avec x_0 et $b > 0$) dont une version de la densité f est
$$f(x) = b\exp(-b(x - x_0))1\!\!1_{[x_0,\infty[}(x).$$
3. La loi normale d'espérance m et de variance σ^2.
4. La loi log-normale généralisée de paramétres x_0 ,m et σ^2. X suit cette loi si elle peut s'écrire $X = x_0 + \exp(m + \sigma U)$ où U suit la loi normale $N(0,1)$.
5. La loi de Cauchy avec changement d'origine et d'échelle. X suit cette loi si elle peut s'écrire $X = b + aY$ (avec $a > 0$) où Y suit la loi de Cauchy usuelle.
6. La loi de Weibull de paramétres α et θ dont la fonction de répartition est donnée par
$$F(x) = 1 - exp\left(-\theta x^{\alpha}\right) 1\!\!1_{[0,+\infty[}(x)$$
où les paramétres et sont strictement positifs.
7. La loi Gamma décentrée. X suit une loi Gamma décentrée ssi $X = x_0 + Y$ où Y suit une loi Gamma de paramètres (α,β). On rappelle que la densité d'une loi Gamma s'écrit
$$f_{\alpha,\beta}(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, \quad x \geq 0$$
8. La loi Béta. La densité d'une loi béta de paramètre (a,b) s'écrit:
$$f_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, \quad x \in [0,1].$$

2 Indications

Chaque fois que cela est possible, on utilisera la méthode du maximum de vraisemblance pour déterminer des estimateurs. On considèrera aussi la méthode des moments pour estimer le ou les paramêtres. Cette méthoe est décrite en annexe. Elle correspond aussi à un exercice de la feuille de TD4 (dans un cas particulier). On pourra aussi choisir, dans certains cas de considérer les estimateurs vus dans les différents TDs (autres que le maximum de vraisemblance).
Lorsque les estimateurs obtenus par ces méthodes sont différents, on comparera leurs qualités respectives. Il peut arriver que la méthode des moments ne soit pas applicable, par exemple lorsque l'espérance mathématique n'existe pas. Dans ce cas, on pourra utiliser d'autres indicateurs tels que la médiane (on identifie la médiane empirique avec la médiane théorique), l'écart-interquartile ou encore la probabilité d'appartenir à un intervalle donné. On comparera ensuite les estimateurs obtenus. On fera notamment une étude du biais, de la variance et de la vitesse de convergence

On donne quelques consignes supplémentaires relatives à certains types de lois.

1) Loi de Cauchy :

Supposons que la loi présumée est la loi de Cauchy avec changement d'origine b et changement d'échelle donné par $a > 0$. On est donc en présence d'une v.a. X de la forme $X = aY + b$, où Y est une loi de Cauchy ordinaire. On souhaite déterminer des estimateurs de a et b. La méthode du maximum de vraisemblance est en principe applicable, mais nécessite la résolution numérique d'un systéme de deux équations non linéaires. Quant à la méthode des moments, on ne peut l'utiliser car les moments théoriques n'existent pas. Dans ce cas, il est opportun de se servir des quantiles. L'identification de la médiane empirique et de la médiane théorique, égale à b, fournit immédiatement une estimation de ce paramétre. Pour estimer le paramétre d'échelle a, on peut identifier l'écart interquartile. On peut également obtenir des intervalles de confiance pour a et b, ou encore une région de confiance pour le couple (a,b), gràce au théoréme de Mosteller sur la convergence en loi des quantiles empiriques.

2) Loi de Weibull :

Dans le cas où la loi trouvée est celle de Weibull, dont les paramétres sont θ et α, la méthode des moments est difficile à mettre en oeuvre à cause de l'intervention de la fonction gamma. On utilisera donc la méthode du maximum de vraisemblance. Pour cela, on écrit les équations de la vraisemblance. Il est possible d'éliminer le paramétre θ, ce qui conduit à une équation ne contenant que le paramétre α. On démontrera ensuite que cette équation admet une solution unique dont on déterminera une valeur approchée par une méthode numérique appropriée.