Master Ingénierie mathématique, Univ. Nantes

Option Mathématiques et applications, ECN

Statistique Inférentielle.

Anne Philippe
Université de Nantes, LMJL

Fiche 2. Modèle logistique et efficacité asymptotique

On considère un n-échantillon $(X_1, \cdots, X_n)$ d'une loi logisitique de paramètre $\theta \in \mathbb{R}$, de densité

$$f(x) = \frac{e^{-(x-\theta)}}{\left(1 + e^{-(x-\theta)}\right)^2}, \quad x \in \mathbb{R}$$

Exercice 1. Partie théorique :

1) Montrer que f est symétrique autour de θ. En déduire l'espérance de X_1.

2) En déduire un estimateur de θ.

3) En admettant que $Var(X_1) = \pi^2/3$, donner une approximation de la loi asymptotique de votre estimateur.

4) On admet que l'information de Fisher apportée par une observation vaut $I(\theta) = 1/3$. L'estimateur proposé précédemment est-il asymptotiquement efficace ?

5) Montrer que le maximum de vraisemblance existe. Peut-on en donner une formule explicite ? Quelle est approximativement sa loi asymptotique ?

6) Proposer un estimateur asymptotiquement efficace qui ne nécessite pas le calcul du maximum de vraisemblance.

7) Montrer que la médiane $m(X_1)$ de X_1 est égale à θ ,

$$m(X_1) = F_\theta^{-1}(1/2) = 1/2.$$

Exercice 2. Avec R

1) La médiane empirique $\hat{m}_n$ (cf. cours sur les quantiles empiriques) est un estimateur de la médiane qui vérifie la propriété suivante

$$\sqrt{n}(m_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{4}\frac{1}{f(\theta)^2})$$

Illustrer ce résultat par la simulation.
Indications :

— Simuler $N = 500$ échantillons de taille n ($= 50$ augmenter n pour visualiser la convergence) suivant la loi la loi logistique de paramètre $\theta = 1$ (utiliser la fonction `rlogis`).
— Calculer sur chaque échantillon $T_{n,i} = \sqrt{n}(m_n - \theta)$, pour $i = 1, ..., N$.
— comparer la densité de la loi limite avec une estimation de la densité calculée sur l'échantillon $T_{n,1}, ..., T_{n,N}$

Objectif :

On veut comparer les cinq estimateurs de θ suivants :
— la moyenne empirique $\bar{X}_n$,
— la médiane empirique m_n,
— le maximum de vraisemblance $\hat{\theta}_n$
(indication : pour trouver le zéro d'une fonction, on pourra utiliser la fonction `uniroot` sous R , qui approche les zéros sur un intervalle que l'on fixe assez grand [-10,10]),
— l'estimateur asymptotiquement optimal associé à la moyenne empirique

$$\widetilde{\theta}_n^1 = \bar{X}_n - \frac{h_n(\bar{X}_n)}{h'_n(\bar{X}_n)}$$

— l'estimateur asymptotiquement optimal associé à la médiane

$$\widetilde{\theta}_n^2 = m_n - \frac{h_n(m_n)}{h'_n(m_n)}$$

où

$$h_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log(f_\theta(X_i))$$

2) Pour chacun des 5 estimateurs, construire une fonction dont le paramètre d'entrée est X un vecteur de longueur n et le paramètre de sortie est le vecteur constitué des estimations de θ calculées sur $(X_1, ..., X_k)$ pour k variant de 10 à n par pas de $d = 10$. Indications : pour éviter des boucles !

```
  est.1 = function(k,x)
{
 calcul de l estimateur sur x[1:k]
}
```

```
est.V1 =function(x,d)
{
sapply(seq(d,n,d) , est.1 ,x)
}
```

3) Simuler un échantillon de taille $n = 300$, calculer et représenter l'évolution des estimateurs en fonction du nombre d'observations pour k variant de d à n par pas de $d = 10$.

   On veut maintenant comparer les variances des 5 estimateurs.

4) Reprendre la question précédente sur $N = 200$ échantillons de taille $n = 100$ (il faudra éventuellement augmenter $n = 200$ et $N = 500$ pour bien estimer les variances des lois limites. Si le temps de calcul est trop long augmenter le pas d entre deux valeurs calculées) .

   Pour chaque estimateur, stocker les résultats dans une matrice à N colonnes. Soit η_k un des estimateurs calculé sur les k premières observations. Soit $\eta_k(1), ...\eta_k(N)$ les N estimations obtenues sur les N échantillons. On note $\widehat{\mathrm{Var}}(\eta_k)$ la variance empirique de $\eta_k(1), ...\eta_k(N)$.
   Remarque : c'est aussi la variance empirique de $\eta_k(1) - \theta, ...\eta_k(N) - \theta$

5) Pour chacun des estimateurs, calculer $\widehat{\mathrm{Var}}(\eta_k)$ pour k variant de 1 à n par pas de d.

6) Superposer les courbes $(k, \widehat{\mathrm{Var}}(\eta_k))$ des 5 estimateurs

7) Superposer les courbes $(log(k), \log(\widehat{\mathrm{Var}}(\eta_k))$ des 5 estimateurs

8) Superposer les courbes $(k, k * \widehat{\mathrm{Var}}(\eta_k))$ des 5 estimateurs

9) Les résultats des simulations sont-ils en accord avec la théorie ?