ÉTUDE DE CAS

Corrigé TD 4
Février 2011

Exercice 1. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ définie comme l'instant de premier succès dans un schéma de Bernoulli de paramètre $q \in]0, 1[$.

1.1. Vérifier que la loi de X est une loi géométrique dont on précisera le paramètre.

Soit $x \in \mathbb{N}^*$, $\mathbb{P}_q(X = x) = (1 - q)^{x-1}q$. La loi de X est la loi géométrique de paramètre q.

1.2. Vérifier qu'il s'agit d'un modèle exponentiel. Donner une statistique exhaustive.

La loi est $p(x, q) = (1 - q)^{x-1}q$, on a :

$$\ln p(x, q) = \ln q - \ln(1 - q) + x \ln(1 - q)$$

Il s'agit d'un modèle exponentiel. On obtient une statistique exhaustive suivante (D'après Théorème de Darmois) $T(X) = X$ qui est la statistique canonique.

1.3. Déterminer I(q), l'information de Fisher sur q d'un échantillon de taille 1.

On a $\frac{\partial}{\partial q} \ln p(x, q) = \frac{1}{q} - \frac{x-1}{1-q}$ et $\frac{\partial^2}{\partial q^2} \ln p(x, q) = -\frac{1}{q^2} - \frac{x-1}{(1-q)^2}$. On en déduit

$$I(q) = \frac{1}{q^2} + \frac{E_q[X] - 1}{(1 - q)^2} = \frac{1}{q^2(1 - q)}$$

1.4. Soit $X_1, ..., X_n$ un échantillon indépendant de taille n de même loi que X. Déterminer $\hat{q}_n$, l'estimateur du maximum de vraisemblance de q.

La vraisemblance du n échantillon est

$$p_n(x_1, .., x_n, q) = q^n(1 - q)^{\sum_{i=1}^n x_i - n}.$$

On regarde les zéros de la dérivée de la log-vraisemblance $l_n = \ln p_n$, et on vérifie que $\hat{q}_n = \frac{n}{\sum\limits_{i=1}^{n} X_i} = \frac{1}{\bar{X}_n}$ est l'EMV de q.

1.5. Montrer que l'estimateur du maximum de vraisemblance est asymptotiquement normal.

On peut appliquer le T.C.L. Comme $\mathrm{Var}_q(X) = \frac{1-q}{q^2}$, il vient

$$\sqrt{n}(\bar{X}_n - \frac{1}{q}) \mapsto^{loi} \mathcal{N}(0, \frac{1 - q}{q^2}).$$

Comme la fonction $h(u) = \frac{1}{u}$ est de classe $\mathcal{C}^1$ pour $u > 0$, on a la convergence en loi suivante

$$\sqrt{n}(\frac{1}{\bar{X}_n} - q) \mapsto^{loi} \mathcal{N}(0, \frac{1 - q}{q^2}q^4).$$

On obtient ainsi les propriétés asymptotiques de l'EMV. Ainsi,

$$\sqrt{n}(\hat{q}_n - q) \mapsto^{loi} \mathcal{N}(0, (1-q)q^2).$$

Exercice 2. On considère le modèle d'échantillonnage $X_1 ..., X_n$ de taille n associé à la famille de lois exponentielles $P = \{\mathcal{E}(\lambda), \lambda > 0\}$. On veut estimer λ

Pour les 4 premières questions, voir TD 2 Ex 2.

2.1. A partir de la méthode des moments, construire un estimateur convergent $\hat{\lambda}_n$ de λ.
2.2. Vérifier qu'il s'agit de l'estimateur du maximum de vraisemblance.
2.3. Déterminer la loi de $\Sigma_{i=1}^n X_i$. Calculer $E_\lambda[\hat{\lambda}_n]$. L'estimateur est-il sans biais ?
2.4. Déterminer un estimateur $\hat{\lambda}_n^*$ sans biais et un estimateur $\hat{\lambda}_n^o$ qui minimise le risque quadratique parmi les estimateurs

$$\hat{\lambda}_n^{(c)} = \frac{c}{\Sigma_{i=1}^n X_i}, \text{ où } c > 0$$

2.5. Calculer l'information de Fisher et la borne Cramer-Rao.

Par définition, on a

$$\begin{aligned} \ln L_1(x_1, \lambda) &= \ln \lambda - \lambda x_i + \ln(\mathbf{1}_{x_1 > 0}) \\ \frac{\partial}{\partial \lambda} L_1(x_1, \lambda) &= \frac{1}{\lambda} - x_1 \text{ et } \frac{\partial^2}{\partial \lambda^2} L_1(x_1, \lambda) = -\frac{1}{\lambda^2} \end{aligned}$$

On obtient donc,
$I_1(\lambda) = -E_\lambda \left[\frac{\partial}{\partial \lambda} L_1(x_1, \lambda)\right] = \frac{1}{\lambda^2}$ et $\text{BCR}(\lambda) = \frac{1}{nI_1(\lambda)} = \frac{\lambda^2}{n}$.

2.6. Les estimateurs étudiés font intervenir la statistique $S_n = \sum_{i=1}^n X_i$. Est-elle exhaustive ?

On vérifie que le modèle est exponentiel en regardant

$$\ln f(x, \lambda) = \ln \lambda - \lambda x + \ln(\mathbf{1}_{x>0})$$

ainsi $f(x, \lambda)$ peut bien s'écrire sous forme exponentielle. D'après Th. de Darmois, on obtient une statistique exhaustive : $S_n = S_n(X_1, ..., X_n) = \sum_{i=1}^n X_i$.

2.7. Résumé : quelles propriétés $\hat{\lambda}_n^*$ a-t-il :
(a) Sans biais : $\hat{\lambda}_n^*$ a été choisi sans biais.
(b) Efficace : L'estimateur sans biais $\hat{\lambda}_n^*$ a comme risque quadratique

$$\text{Var}(\hat{\lambda}_n^*) = R(\hat{\lambda}_n^*, \lambda) = \frac{\lambda^2}{(n-1)(n-2)}[(n-1)^2 - (n-1)(n-2)] = \frac{\lambda^2}{n-2}.$$

Il n'atteint pas la borne BCR. Il n'est donc pas efficace.
(c) Asymtotiquement normal. Pour répondre à cette question, nous avons besoin du Théorème de Slutsky qui n'a pas été traité en cours !