Fiche 3 - Statistique - M1 Ingé Math

Tran Viet Chi, `chi.tran@math.univ-lille1.fr`, bureau 316 (bâtiment M3).

Exercice 1 (Calcul d'information de Fisher)
Calculer les informations de Fisher pour les modèles suivants :

1. Loi de Poisson $\mathcal{P}(\lambda)$. Commenter le comportement en fonction de λ.
2. Loi de Pareto de paramètres α et θ, de densité

$$f_{\alpha,\theta}(x) = \frac{\alpha - 1}{\theta} \exp\big(\alpha \ln(\theta/x)\big)\mathbf{1}_{x>\theta}.$$

   Expliquer pourquoi il faut fixer θ.

3. Loi multinomiale $\mathcal{M}(n; p_1, \ldots, p_K)$ avec $K \in \mathbb{N}^*$ et $p_1 + \cdots + p_K = 1$.

Exercice 2 (Estimation de la variance d'une loi normale)
Soit $X_1, \ldots X_n$ un échantillon i.i.d. de loi $\mathcal{N}(0, \theta)$, $\theta > 0$.

1. Montrer que $S_n = X_1^2 + \cdots + X_n^2$ est une statistique exhaustive complète de θ.

2. Calculer $\mathbb{E}(S_n^2)$.

3. Trouver l'estimateur SBVM de θ^2.

Exercice 3 (Estimation de la survie d'une loi exponentielle)
Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de loi exponentielle $\mathcal{E}(1/\theta)$ dont la densité est :

$$f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}} \mathbf{1}_{t>0}.$$

1. Ecrire le modèle statistique et calculer l'information de Fisher du modèle.

2. Calculer l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$ de θ. Cet estimateur $\widehat{\theta}_n$ est-il convergent ? sans biais ?

3. Calculer la fonction de survie $S(t) = \mathbb{P}(X_1 > t)$. Connaissez-vous un estimateur de cette quantité ?

4. On décide d'estimer $S(t)$ par :

$$\widehat{S}_n(t) = e^{-\frac{t}{\widehat{\theta}_n}}.$$

Cet estimateur est-il convergent ? sans biais ?
5. Ecrire un théorème de normalité asymptotique pour $\widehat{\theta}_n$, c'est-à-dire, montrer que :

$$\sqrt{n}\big(\widehat{\theta}_n - \theta\big) \to \mathcal{N}(0, \sigma^2) \tag{1}$$

la convergence étant en loi et σ^2 étant une constante à déterminer.

6. On rappelle la règle suivante, appelée δ**-méthode** :

Si on a (1) *et si g est une fonction de classe $\mathcal{C}^1$ et de dérivée non nulle au voisinage de θ, alors :*

$$\sqrt{n}\big(g(\widehat{\theta}_n) - g(\theta)\big) \to \mathcal{N}(0, g'(\theta)^2\sigma^2),$$

la convergence étant en loi. □

En utilisant la δ-méthode, rappelée ci-dessus, établir un théorème de normalité asymptotique pour $\widehat{S}_n(t)$. Que remarquez-vous quant à la variance asymptotique ?

7. Construire un intervalle de confiance asymptotique à 95% pour $S(t)$.

8. On définit $T_i = \mathbf{1}_{Y_i > t}$ et $\Sigma = X_1 + \cdots + X_n$. Donner la densité de la loi jointe de (X_1, Σ). En déduire la densité de X_1 conditionnellement à Σ.

9. On pose $T^* = \mathbb{E}(T|\Sigma)$. Calculer T^* en fonction de Σ et t.

10. Montrer que T^* est l'estimateur SBVM de $S(t)$.

11. T^* est-il efficace à distance finie (c'est-à-dire pour n fixé ne tendant pas vers l'infini) ?

Exercice 4 (Estimation par capture-recapture)
On veut compter le nombre de poissons θ dans un lac fermé. Initialement, aucun poisson n'est marqué. On réitère l'expérience suivante n fois : on tire au hasard un poisson et on le marque s'il n'est pas déjà marqué. Chacun des tirages i peut être représenté par une variable aléatoire Y_i qui vaut 1 si le poisson est déjà marqué, 0 sinon. Par construction, $Y_1 = 0$. Le but est de trouver une statistique exhaustive de θ.

1. La difficulté réside ici dans le fait que les Y_i ne sont pas i.i.d. Ecrire le modèle statistique.

2. Montrer que :

$$\begin{aligned}\mathbb{P}\big(Y_n = y_n, \ldots, Y_1 = y_1\big) =& \mathbb{P}\big(Y_n = y_n \,|\, , Y_{n-1} = y_{n-1}, \ldots, Y_1 = y_1\big)\\ &\times \mathbb{P}\big(Y_{n-1} = y_{n-1} \,|\, Y_{n-2} = y_{n-2}, \ldots, Y_1 = y_1\big)\\ &\times \cdots \times \mathbb{P}(Y_1 = y_1).\end{aligned}$$

3. Montrer que la loi conditionnelle de Y_n sachant $Y_{n-1}, \ldots, Y_1$ est une Bernoulli de paramètre

$$p_n = \frac{n - 1 - \sum_{i=1}^{n-1} Y_i}{\theta}.$$

4. En déduire que la vraisemblance est proportionnelle à :

$$\prod_{i=2}^{n} \frac{(\theta - i + 1 + \sum_{j=1}^{i-1} Y_j)^{1-Y_i}}{\theta}. \tag{2}$$

5. On pose $S_n = \sum_{i=1}^{n} Y_i$. Montrer que (2) peut se réécrire sous la forme :

$$\frac{1}{\theta^n} \prod_{k=0}^{n-1-S_n} (\theta - k) = \frac{1}{\theta^{n-1}} \frac{(\theta - 1)!}{(\theta - n + S_n)!}$$

6. En déduire que S_n est une statistique exhaustive pour θ.