Statistique de Base - M1 Ingé. Math.
Partiel du 7 novembre 2011 - 2 heures

Formulaire sur copie double autorisé
Calculatrice autorisée
Le barème n'est que prévisionnel.

chi.tran@math.univ-lille1.fr

Exercice 1 (Loi de Poisson, 7 points)
On considère un échantillon de variables aléatoires $X_1, \ldots, X_n$, i.i.d. de loi de Poisson $\mathcal{P}(\theta)$ avec $\theta > 0$. On rappelle que :

$$\mathbb{P}(X_i = k) = e^{-\theta}\frac{\theta^k}{k!}$$

On cherche à estimer $\mathbb{P}(X_i = 0)$.

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

2. On s'intéresse à la statistique $S_n = \sum_{i=1}^n X_i$. Cette statistique est-elle exhaustive ? totale ?

3. Calculer $\mathbb{P}(X_i = 0)$. Montrer que $\mathbf{1}_{X_1=0}$ est un estimateur sans biais de cette probabilité.

4. Rappeler (calcul non nécessaire si vous vous en souvenez) quelle est la loi de la somme de deux variables aléatoires de Poisson indépendantes et de paramètres respectifs λ et μ.

5. Calculer $\mathbb{P}(X_1 = k \mid S_n = s)$ et en déduire que la loi de X_1 conditionnellement à S_n est une loi binomiale de paramètre $(S_n, 1/n)$.

6. Calculer $\xi_n = \mathbb{E}(\mathbf{1}_{X_1=0} \mid S_n)$. En déduire que ξ_n est le SBVM de $\mathbb{P}(X_i = 0)$.

7. La statistique ξ_n est-elle efficace ?

Exercice 2 (Modèle de machine industrielle, 15 pts)
On observe n machines indépendantes. Chacune des machines, par exemple la machine $i \in \{1, \ldots, n\}$, est composée d'un générateur dont la durée de vie est modélisée par une variable aléatoire (v.a.) C_i de loi exponentielle de paramètre $\mu > 0$ et d'un module dont la durée de vie est modélisée par une v.a. Z_i de loi exponentielle de paramètre $\lambda > 0$. Les v.a. Z_i et C_i sont indépendantes.

Partie A : observation parfaite

Dans un premier temps, on suppose qu'on observe les durées de vie des générateurs et des modules séparément. On dispose d'observations $(z_i, c_i)_{i\in\{1,\ldots,n\}}$.

1. Ecrire le modèle statistique correspondant. S'agit-il d'un modèle exponentiel ?

2. Exhiber une statistique exhaustive. Cette statistique est-elle complète ?

3. Calculer l'estimateur du maximum de vraisemblance $(\widehat{\lambda}_n, \widehat{\mu}_n)$ de (λ, μ).

4. Ces estimateurs sont-ils convergents ? sans biais ? Calculer le biais de $\widehat{\lambda}_n$. On pourra utiliser que la densité de la loi $\Gamma(n, \lambda)$ est :

$$f(x; n, \lambda) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \mathbb{1}_{x>0}$$

5. Donner un résultat de normalité asymptotique pour $(\widehat{\lambda}_n, \widehat{\mu}_n)$ lorsque n tend vers l'infini.

6. Proposer des estimateurs SBVM de λ et μ.

Partie B : observation imparfaite

On suppose maintenant que seules les variables aléatoires $X_i = \min(Z_i, C_i)$ sont observées. En effet, lorsque le générateur ou le module tombent en panne, la machine s'arrête de fonctionner.

7. Calculer la fonction de survie de X_i.

8. Ecrire le nouveau modèle statistique.

9. Ecrire la vraisemblance. Quelles sont les fonctions identifiables de λ et μ ?

10. On pose $\theta = \lambda + \mu$. Quels sont les estimateurs du maximum de vraisemblance de θ fondés :
- sur les observations $X_1, \ldots, X_n$?
- sur les observations $(Z_1, C_1), \ldots, (Z_n, C_n)$ (dans le cas d'une observation parfaite) ?

Est-il naturel que ces estimateurs soient différents ?

11. Ecrire un résultat de normalité asymptotique pour l'estimateur $\widehat{\theta}_n$ fondé sur les observations $X_1, \ldots, X_n$. Comparer en utilisant la question 6.

Partie C : conclusion

12. Déduire des parties A et B l'expression de l'espérance conditionnelle

$$\mathbb{E}\Big(\frac{1}{\sum_{i=1}^n \min(Z_i, C_i)} \mid \sum_{i=1}^n Z_i, \sum_{i=1}^n C_i\Big).$$