Université de Bretagne-Sud

Statistique Bayésienne

Travaux dirigés 5

Exercice 1 – On considère la loi de Dirichlet de paramètres $(\alpha_1, \alpha_2; \alpha_3)$.

1. Vérifier que la somme de la densité sur le simplexe est 1.
2. Calculer les lois marginales de X_1 et X_2.

Exercice 2 – Montrer le résultat suivant :

Soient $X_1, X_2, \cdots, X_k$, $k+1$ variables aléatoires indépendantes de loi Gamma de paramètres respectifs $(\alpha_i, 1)$, $i = 1, \cdots, k+1$. On considère les variable :

$$Y_i = \frac{X_i}{X_1 + \cdots + X_{k+1}} \ , \quad i = 1, \cdots, k.$$

Alors $(Y_1, \cdots, Y_k)$ suit une loi de Dirichlet de paramètre $(\alpha_1, \alpha_2, \cdots, \alpha_k; \alpha_{k+1})$.

Exercice 3 – Soient $Y_1, \cdots, Y_n$, n variables aléatoires indépendantes de loi bêta de paramètres respectifs (α_i, β_i), $i = 1, \cdots, n$. On considère les v.a. :

$$X_i = Y_i \prod_{j=1}^{i-1} (1 - Y_j).$$

Montrer que $(X_1, \cdots, X_n)$ suit une loi de Dirichlet généralisée de paramètres $((\alpha_1, \xi_1), \cdots, (\alpha_n, \xi_n))$ où $\xi_i = \beta_i - \alpha_i - \beta_i + 1$ pour $i = 1, \cdots, n-1$ et $\xi_n = \beta_n + 1$.

Exercice 4 – Dans de nombreuses études statistiques, l'observation se résume à un effectif enregistré dans un intervalle donnée. On parle de *données groupées*. On dispose alors d'un échantillon $(k_1, \cdots, k_{n+1})$ où k_i est le nombre de valeurs de la v.a. X dans l'intervalle $[t_{i-1}, t_i)$. $t_0 = 0$ et $t_{n+1} = +\infty$. On note $n = \sum_{i=1}^{n+1} k_i$.
On considère $f = (f_1, \cdots, f_{n+1})$ une discrétisation de la densité de la loi de probabilité de X. C'est le paramètre que l'on souhaite étudier.

1. Ecrire la vraisemblance
2. On veut faire une estimation bayésienne de f. Proposer une loi a priori.
3. Calculer la loi a posteriori et donner un estimateur de f.
4. En déduire un estimateur de F.

Exercice 5 – Soit un échantillon $(X_1, \cdots, X_n)$ de durées i.i.d. de loi $f(x)$ définie par :

$$f(x) = f_i \quad \text{pour} \quad x \in [t_{i-1}, t_i[, \quad i = 1, \cdots, m \quad \text{avec} \quad t_0 = 0 \quad \text{et} \quad t_m = +\infty.$$

Soit la fonction $\lambda(x)$ définie par :

$$\lambda(x) = \lim_{dx \to 0} \frac{P(X \leq x + dx | X > x)}{dx}$$

On note, pour $x \in [t_i - 1, t_i[$, $\lambda(x) = \lambda_i$ et $R(t_i) = P(X > t_i) = R_i$, $i = 1, \cdots, m$. On note $(k_1, \cdots, k_m)$ le vecteur du nombre d'observations dans chaque intervalle i.e. k_i est le nombre de durées dans dans l'intervalle $[t_i - 1, t_i[$. $\sum_{i=1}^m k_i = n$.

1. Montrer que $\lambda(x) = f(x)/(1 - F(x))$ où $F(x)$ est la fonction de répartition de X i.e. $\int_0^x f(t)dt$.
2. En considérant l'observation d'un vecteur de pannes, donner l'expression de la fonction de vraisemblance pour les paramètres $(f_1, \cdots, f_m)$.
3. Donner cette expression pour les paramètres $(\lambda_1, \cdots, \lambda_m)$.
4. En considérant une loi conjuguée pour le modèle obtenu précédemment, proposer un estimateur de Bayes du taux de survie. Commenter.
5. Construire une loi non informative (On pourra appliquer la règle de Jeffreys) et donner alors une estimation bayésienne du vecteur λ.

On souhaite maintenant introduire dans l'analyse l'idée a priori que le taux de survie est croissant. On pose $u_i = 1 - \lambda_i$. n.b. $u_i = P(X > t_i / X > t_{i-1})$. On se donne une loi a priori de type Dirichlet sur le vecteur $((y_1, \cdots, y_k)$ où $y_i = u_{i-1} - u_i$, $u_0 = 1$ et $y_{m+1} = 1 - \sum_{j=1}^m y_j = u_m$. Les paramètres $(\beta\alpha_1, \cdots, \beta\alpha_m)$ de cette loi sont tels que $\beta > 0$, $\alpha_i > 0$ pour tout i et $\sum_{j=1}^{m+1} \alpha_j = 1$.

1. Montrer que l'espérance mathématique des v.a. u_i est $\sum_{j=i+1}^{m+1} \alpha_j$.

2. En utilisant la formule du binôme de Newton, montrer qu'une estimation Bayésienne des λ_i pour la loi a priori définie ci-dessus et sous l'hypothèse d'un coût quadratique a la forme suivante :

$$\tilde{\lambda}_i = \frac{C_{i,1}}{C_0} \ , \quad i = 1, \cdots, m$$

où $C_{i,1}$ est égal à C_0 où k_i est devenu $k_i + 1$, C_0 étant :

$$C_0 = \sum_{l_m=0}^{k_m} \cdots \sum_{l_1=0}^{k_1} \begin{matrix} k_i \\ l_i \end{matrix} \ (-1)^{l_i} B(\beta\alpha_i, \xi_i)$$

où B est le coefficient bêta.