RPM4 2024: TD de statistique

Clément Gauchy

Novembre 2024

Exercice 1: Estimation Bayésienne des paramètres d'une loi de Poisson

On souhaite estimer l'intensité d'une source émettrice de particules. On effectue K comptages $(N_i)_{1 \leqslant i \leqslant k}$ de particules de façon indépendantes sur un détecteur. On suppose que chaque N_i suit une loi de Poisson $\mathcal{P}(\lambda)$ i.e. $\mathbb{P}(N_i = n) = e^{-\lambda}\frac{\lambda^n}{n!}$.

1) Quel est le modèle statistique ? Selon vous que devons nous estimer dans ce problème ?

2) Calculez la log-vraisemblance de ce modèle et l'estimateur du maximum de vraisemblance.

3) On suppose pour λ une loi *a priori* de type Gamma $\Gamma(\alpha, \beta)$ dont la densité de probabilité est $f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$. Calculez la loi *a posteriori* à l'aide de la règle de Bayes. Que remarquez vous ?

4) Calculez la moyenne *a posteriori* et l'estimateur MAP.

Exercice 2: Estimation du temps de détection entre deux particules.

On souhaite estimer le temps d'attente entre la détection de deux photons sur un détecteur. On suppose que le temps d'attente T est une variable aléatoire suivant une loi exponentielle de paramètre λ i.e. sa densité de probabilité est $f_T(t) = \lambda e^{-\lambda t}$. Nous avons en notre possession un jeu de données $(T_1, \ldots, T_n)$ i.i.d. de même loi que T.

On définit les lois Gamma $\mathcal{G}(\alpha, \beta)$ par la densité de probabilité $f_{\alpha,\beta}(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$. Sa moyenne est α/β et son mode $(\alpha - 1)/\beta$ pour $\alpha \geqslant 1$.

1) Quel est le modèle statistique ? Selon vous que devons nous estimer dans ce problème ?

2) On sait que la loi exponentielle de paramètre λ admet $1/\lambda$ comme moyenne. Proposez un estimateur de λ par méthode des moments.

3) Écrivez la vraisemblance de ce modèle. Déterminez l'estimateur du maximum de vraisemblance $\widehat{\lambda}_{\mathrm{MV}}$ pour λ (utilisez pour cela la log-vraisemblance, on admettra que la log-vraisemblance est concave). Que remarquez vous ?

4) On admet que $\sum_{i=1}^{n} T_i \sim \mathcal{G}(n, \lambda)$. Calculez $\mathbb{E}\left[\frac{1}{\sum_{i=1}^{n} T_i}\right]$ avec un calcul d'intégrale (on utilisera la propriété $\Gamma(n) = (n-1)\Gamma(n-1)$).

5) À l'aide de la question 4), calculez le biais de $\widehat{\lambda}_{\text{MV}}$.

On va désormais utiliser des méthodes bayésiennes. On utilisera comme loi *à priori* pour λ une loi Gamma de paramètre α et β.

6) À l'aide de la règle de Bayes, calculez la loi *a posteriori* de λ sachant $(T_1, \ldots, T_n)$.

7) Déterminez l'estimateur de la moyenne et du mode *a posteriori* de λ.

Exercice 3: Estimateur Monte-Carlo pour une étude de radio-protection

On considère un matériau homogène de section efficace macroscopique d'absorption μ pour une particule non spécifié. On ne considèrera que l'interaction par absorption dans cet exercice. On souhaite protéger un opérateur d'un terme source émetteur de particules énergétiques à l'aide de ce matériau. On suppose que l'opérateur est situé derrière une épaisseur L de ce matériau.

Soit X la variable aléatoire désignant la longueur parcourue dans le matériau par la particule.

1) Rappelez sans démonstration la loi de X en utilisant les informations du paragraphe ci dessus

2) Pour la radioprotection de l'opérateur, une quantité très importante est $p = \mathbb{P}(X > L)$. Que représente cette quantité ? Pourquoi est elle importante ?

3) On suppose que l'on sait générer $(X_i)_{1\leqslant i\leqslant N}$ i.i.d. suivant la loi de X, écrivez $\widehat{p}_{\text{MC}}$ l'estimateur Monte-Carlo de p.

4) Écrivez la variance de $\widehat{p}_{\text{MC}}$. En supposant que $p = 10^{-6}$, combien doit valoir N pour que le coefficient de variation de $\widehat{p}_{\text{MC}}$ soit de 1% ? (Pour rappel, le coefficient de variation est le ratio écart-type sur moyenne).

On va construire un estimateur par échantillonage d'importance (*importance sampling*) pour réduire la variance de $\widehat{p}_{\text{MC}}$. Pour $0 < \theta < \mu$, on définit la densité d'échantillonage $h_\theta(x) = (\mu - \theta)e^{-(\mu-\theta)x}$.

5) De quelle loi h_θ est elle la densité de probabilité ? Quelle interprétation physique peut on faire si on diminue la section efficace macroscopique μ par θ ?

6) On génère $(Y_i)_{1\leqslant i\leqslant N}$ selon la densité de probabilité h_θ. Ecrire l'estimateur par échantillonage d'importance $\widehat{p}_{IS}$.

7) Écrivez la variance de $\widehat{p}_{IS}$. **Bonus:** Proposez une borne supérieur de cette variance sachant que $e^{-\theta x} < e^{-\theta L}$ pour $x > L$. Trouvez les conditions sur θ, μ et L pour que $\text{Var}(\widehat{p}_{IS}) < \text{Var}(\widehat{p}_{MC})$