8 TD8

8.1 Laplace, mélange et bayésien hiérarchique

On considère l'expérience botanique suivante : une plante située à l'origine émet une graine suivant un axe fixé, disons l'axe des abscisses. La graine est soumise à un mouvement aléatoire dû à l'effet d'un vent de force τ et aterrit à l'ascisse $X \mid \tau \sim \mathcal{N}(0, 2\tau)$. La force du vent est inconnue et modélisée par une loi exponentielle $\tau \sim \mathcal{E}(\lambda)$, $\lambda > 0$.

1. Ecrire la densité (marginale) $f_X(x)$ de X comme une intégrale que l'on ne cherchera pas à calculer.
2. Calculer la fonction génératrice $M_X(t) = E[e^{tX}]$ pour tout t tel que $|t| < \sqrt{\lambda}$.
3. Montrer que si Y est une loi de Laplace de paramètre μ, alors $M_Y(t) = \mu^2/(\mu^2 - t^2)$, pour tout $|t| < \mu$.
4. Identifier la loi marginale de X et interpréter le résultat dans un contexte bayésien hiérarchique : quelle loi a priori correspond à la hiérarchie $\theta \mid \tau \sim \mathcal{N}(0, 2\tau)$ et $\tau \sim \mathcal{E}(\alpha)$ pour α une constante positive fixée ?

8.2 Test bayésien (1)

Soient $X = (X_1, \ldots, X_n) \mid \theta \sim \mathcal{N}(\theta, \sigma^2)^{\otimes n}$ et $\theta \sim \Pi = \mathcal{N}(\mu, \tau^2)$, où σ^2, τ^2 sont fixés.

1. Déterminer la loi a posteriori.
2. On veut tester $H_0 = \{\theta \geq 1\}$ contre $H_1 = \{\theta < 1\}$ du point de vue bayésien. Pour un test $\varphi = \varphi(X)$ et $\theta \in \mathbb{R}$, on considère une fonction de perte équilibrée
$$\ell(\theta, \varphi) = \mathbb{1}_{\theta \in \Theta_0} \mathbb{1}_{\varphi = 1} + \mathbb{1}_{\theta \in \Theta_1} \mathbb{1}_{\varphi = 0}.$$
Construire le test bayésien correspondant pour la loi Π ci-dessous.
3. Que devient le test si on remplace H_0 par H_1 et vice-versa ?

8.3 Test bayésien (2)

Soit $X = X_1 \mid \theta \sim \mathcal{N}(\theta, 1)$. Soient les deux problèmes de test

$$H_0^1 : \theta = 0 \quad \text{vs.} \quad H_1^1 : \ \theta \neq 0$$
$$H_0^2 : |\theta| \leq \varepsilon \quad \text{vs.} \quad H_1^2 : \ |\theta| > \varepsilon$$

1. Proposer une loi a priori avec une partie gaussienne $\mathcal{N}(0, \sigma^2)$ pour chaque situation.
2. Comparer les tests bayésiens correspondants lorsque ε et σ varient, dans le cas d'une fonction de perte équilibrée.

8.4 Estimation de densité marginale via importance sampling

Soit (X, Y) un couple de variables de loi $P_{X,Y}$ de densité $f_{X,Y}$ sur $\mathbb{R}^2$. Soient $(X_1, Y_1), \ldots, (X_N, Y_N)$ i.i.d. de loi $P_{X,Y}$ et soit w une densité quelconque sur $\mathbb{R}$.

1. Montrer que
$$\hat{f}_X(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{f_{X,Y}(x, Y_i) w(X_i)}{f_{X,Y}(X_i, Y_i)}$$
est un estimateur consistant de (i.e. converge en probabilité vers) $f_X(x)$.

2. Donner une expression de la variance de cet estimateur.
3. Dans le cas où $Y \sim \mathcal{N}(0,1)$ et $X \mid Y = y \sim \mathcal{N}(y, 1+y^2)$, proposer une mise en oeuvre de la méthode ci-dessus pour estimer $f_X(x)$.

8.5 Autour de l'échantillonage de Gibbs

Dans la suite, (X,Y) est un couple de variables réelles à densité non nulle sur tout $\mathbb{R}^2$. Les densités des lois de (X,Y), $Y \mid X$ et $X \mid Y$ sont notées respectivement $f(x,y) = f_{X,Y}(x,y)$, $f_{Y \mid X}(y \mid x)$ et $f_{X \mid Y}(x \mid y)$.

L'algorithme de Gibbs permet de simuler suivant une approximation de la loi jointe de (X,Y) si l'on sait simuler suivant chacune des lois conditionnelles $\mathcal{L}(Y \mid X)$ et $\mathcal{L}(X \mid Y)$.

1. Pour vérifier, déjà, que cela est concevable, montrons que connaissant les densités $f_{X \mid Y}$ et $f_{Y \mid X}$ il est possible de reconstruire $f_{X,Y}$. Pour cela
   (a) En passant par $f_{X,Y}$, écrire une identité reliant $f_{X \mid Y}, f_{Y \mid X}$ et les lois marginales.
   (b) En déduire une façon d'écrire f_X (puis f_Y) sous la forme d'une intégrale et des densités conditionnelles et conclure.
   (c) Pourrait-on "prendre" $f_{X \mid Y}$, $f_{Y \mid X}$ densités d'une loi exponentielle ?
2. *Algorithme de Gibbs à deux variables*

   i) Générer X_0, Y_0 de façon arbitraire sur $\mathbb{R}$.
   ii) Pour $n \geq 1$, si X_n, Y_n ont été générées, poser

$$X_{n+1} \sim f_{X \mid Y}(x \mid Y_n)$$
$$Y_{n+1} \sim f_{Y \mid X}(y \mid X_{n+1}).$$

   Soit $(X_n, Y_n)_{n \geq 0}$ générée suivant l'algorithme de Gibbs. Dans cette question, on suppose que pour un n donné, la loi du couple (X_n, Y_n) est de densité précisément $f_{X,Y}$.
   (a) Quelle est alors la loi de Y_n, de $X_{n+1} \mid Y_n$?
   (b) Montrer que $\mathcal{L}((X_{n+1}, Y_n)) = \mathcal{L}((X_n, Y_n))$.
   (c) En déduire que $\mathcal{L}((X_{n+1}, Y_{n+1})) = \mathcal{L}((X_n, Y_n))$.
3. Pour $(X_n, Y_n)_{n \geq 0}$ générée suivant l'algorithme de Gibbs, montrer que la suite $(Z_n)_{n \geq 0} = ((X_n, Y_n))_{n \geq 0}$ est une chaîne de Markov, et que $\mathcal{L}((X,Y))$ en est une loi stationnaire.
4. Proposer une généralisation de l'algorithme de Gibbs à p variables notées $X^{(1)}, \dots, X^{(p)}$.
5. *Application aux statistiques bayésiennes.* Considérons un a priori hiérarchique $\theta \mid \alpha \sim \Pi_\alpha$ et $\alpha \sim Q$, dans le cadre et notations de l'exercice 1 du TD7. On souhaite simuler suivant une approximation de la loi a posteriori $\mathcal{L}(\theta \mid X)$. On suppose que l'on sait facilement simuler suivant les lois $\mathcal{L}(\theta \mid \alpha, X)$ et $\mathcal{L}(\alpha \mid \theta, X)$.
   (a) Montrer qu'il suffit de savoir simuler suivant (une approximation de) $\mathcal{L}((\theta, \alpha) \mid X)$.
   (b) Proposer un algorithme de type Gibbs pour simuler suivant la loi $\mathcal{L}((\theta, \alpha) \mid X)$.