Université de Bordeaux
M1 MAS
UE : Simulation stochastique et méthodes bayésiennes pour le traitement du signal
Année 2017-2018

TP2 - Calcul de lois a posteriori

Les TP de cette UE sont inspirés de ceux proposés par François Caron, Jean-François Giovannelli, et Adrien Todeschini qui ont assuré cet enseignement pendant plusieurs années.

---

1 Modèle Gaussien-Gaussien

Le code suivant peut-être utilisé pour tracer la densité d'une loi Gaussienne de moyenne $\theta = 1$ et de variance $\sigma^2 = 2$, et pour simuler un échantillon de taille $n = 100$ selon cette loi.

```
M = 1000
x = seq(-4,6,length.out = M)
theta = 1
sigma2 = 2

f = dnorm(x,mean = theta, sd = sqrt(sigma2))

dev.new()
plot(x,f, type='l', lwd=2, xlab=expression(x),ylab=expression(f(x)))

n = 100
X = rnorm(n,mean = theta, sd = sqrt(sigma2))
```

On considère le modèle bayésien suivant :

$$\begin{array}{rcl} X_1, \dots, X_n | \theta & \sim_{iid} & \mathcal{N}(\theta, 2) \\ \pi(\theta) & \sim & \mathcal{N}(\mu, \tau^2). \end{array}$$

Q. 1 *Soit $1 \leq q \leq n$. Rappeler l'expression de la loi a posteriori $\pi(\theta|X_1, \dots, X_q)$, puis choisissez des valeurs pour les hyperparamètres μ et τ qui vous semblent appropriées.*

Q. 2 *Superposer sur un même graphique la densité de la loi a posteriori $\pi(\theta|X_1, \dots, X_q)$ pour différentes tailles d'échantillon $q \in \{1, \dots, n\}$. Quelle est l'influence de la taille d'échantillon q sur la concentration de la loi a posteriori lorsque le paramètre à estimer est $\theta = 1$?*

Q. 3 *Faites varier les valeurs de μ et τ. Quelle est l'influence de ces hyperparamètres sur la concentration de la loi a posteriori ?*

2 Modèle Beta-Bernoulli

Q. 4 *Utiliser le code ci-dessous pour charger le fichier de données* `betabinom` *associé à ce TP, et visualiser son contenu. Comment pouvez-vous modéliser les éléments du vecteur de données* `data` *?*

```
data = read.table('betabinom.txt')$Y
print(data)
```

Q. 5 *Exécuter le code ci-dessous pour tracer la densité de la loi $Beta(a, b)$ pour différentes valeurs des paramètres a et b.*

```
# Valeurs de a
a_vec <- c(1,.5,2,3,1,3,4)

# Valeurs de b
b_vec <- c(1,.5,2,1,2,8,2)

# Valeurs en abscisse
x_vec <- seq(0,1,by=.01)

# Tracé de la première courbe
plot(x_vec, dbeta(x_vec, a_vec[1], b_vec[1]), type='l', lwd=3, xlab='x',
     ylab='Beta(x;a,b)', ylim=c(0,3))

# Ajout des autres courbes
col <- c(1:7)
for (i in 2:length(a_vec))
{
  lines(x_vec, dbeta(x_vec, a_vec[i], b_vec[i]), type='l', lwd=3, col=col[i])
}
# Ajout d'une légende
legend('top', bty='n',
       legend=paste('a=',a_vec, ', b=', b_vec, sep=''),
       col=col, lwd=3)
```

Dans la suite, on considère le modèle bayésien suivant

$$\begin{array}{rcl} X_1, \ldots, X_n|\theta & \sim_{iid} & Bernoulli(\theta) \\ \pi(\theta) & \sim & Beta(a, b). \end{array}$$

Q. 6 *Rappeler l'expression de la loi a posteriori $\pi(\theta|X_1, \ldots, X_n)$. Pour différentes valeurs du couple de paramètres (a, b) de la loi a posteriori, représenter, sur un même graphique, la densité de la loi a priori ainsi que la densité de la distribution a posteriori pour $n = 5, 10, 20, 100, 1000$.*

Q. 7 *Proposer une estimation de la valeur du paramètre θ pour cet échantillon à partir de l'espérance de la loi a posteriori et le maximum de la loi a posteriori puis la comparer à celle donnée par le maximum de vraisemblance en faisant varier n.*

3 Modèle Gaussien général

On considère les deux modèles bayésiens suivants :

$$\begin{array}{rcl} X_1, \dots, X_n|\theta & \sim_{iid} & \mathcal{N}(\theta, \sigma^2) \\ \pi_(\theta, \sigma) & \sim & 1/\sigma \\ \tilde{\pi}_1(\theta|\sigma^2) & \sim & \mathcal{N}(\theta_0, \sigma^2/n_0) \\ \tilde{\pi}_2(\sigma^2) \sim & IG(\nu/2, s_0^2/2) & \end{array}$$

où π est un *a priori* non informatif et $\tilde{\pi}$ est un a priori conjugué.

Q. 8 *Pour $1 \leqslant q \leqslant n$, rappeler l'expression de la loi* a posteriori $\pi(\theta, \sigma^2|x)$ *et* $\tilde{\pi}(\theta, \sigma^2|x)$ *Pour différents hyperparamètres* $(\theta_0, n_0, \nu, s_0)$ *et différentes tailles d'échantillon* $q \in \{1, ..., n\}$, *tracer sur un même graphique les densitées* a posteriori *non informatives et conjuguées.*

Q. 9 *Générer un échantillon aléatoire de taille* $n = 1000$ *de loi* $\mathcal{N}(1, 1)$. *Étudier l'influence de la loi* a priori *et des hyperparamètres pour l'estimation de* μ *et* σ^2 *à l'aide de l'espérance ou du maximum de la loi* a posteriori.