MPA - Ensimag 2A **2012-13**

Séance 6 - Chaînes de Markov et algorithmes MCMC

Exercice 1. Dans cette séance, on considère le modèle de vraisemblance binomiale et de loi a priori choisie dans lafamille de lois Beta (modèle Beta-binomial). On note θ le paramètre correspondant à une probabilité inconnue d'un événement donn, et on suppose que l'on observe $y = 9$ réalisations de cet événement lors de la répétition de $n = 20$ épreuves indépendantes. L'objectif de cette séance est de programmer plusieurs algorithmes de Monte Carlo par chaînes de Markov et d'évaluer la convergence de ces algorithmes.

1. On suppose que la loi a priori est uniforme sur $(0, 1)$. Donner un argument de théorie de l'information pour justifier que ce choix exprime le maximum d'incertitude a priori sur θ. Rappeler les expressions mathématiques de la vraisemblance, de la loi Beta et de la loi a posterior du paramètre θ, aux constantes près.
2. Rappeler l'algorithme d'estimation de Metropolis-Hasting (MH) pour un noyau instrumental $Q(\theta^*|\theta)$.
3. On choisit ce noyau de transition égal à la loi a priori (les tirages ne dépendent pas de la valeur en cours θ). Vérifier que le rapport de MH est égal à :

$$r = \left(\frac{\theta^*}{\theta^t}\right)^y \times \left(\frac{1-\theta^*}{1-\theta^t}\right)^{(n-y)}$$

4. Ecrire un script en `R` implantant la simulation de la loi a posteriori

```
theta.1 <- runif(1)
ratio <- (theta.1/theta.0)^ y *((1 - theta.1)/(1 - theta.0))^ (n-y)
if (runif(1) < ratio) theta.0 <- theta.1
```

5. Afficher un histogramme de la loi a posteriori, calculer numériquement un intervalle de crédibilité à 95% pour θ et afficher un histogramme de la loi predictive sachant y.

6. Testons maintenant l'algorithme implanté dans le script `R` suivant

```
#Metropolis-Hastings
#kernel = random walk
nit = 1000
theta.0 <- 0.1
theta.post <- theta.0
log.likelihood <- NULL

for (i in 1:nit) {
# proposal
b <- 100
theta.1 <- rbeta(1,b*theta.0/(1-theta.0), b)

# MH ratio
ratio1 <- (theta.1/theta.0)^(y)* ((1 - theta.1)/(1 - theta.0))^(n-y)
ratio2 <-  dbeta(theta.0, b*theta.1/(1-theta.1), b)
/dbeta(theta.1, b*theta.0/(1-theta.0), b)
ratio <- ratio1*ratio2

if (runif(1) < ratio) {theta.0 <- theta.1}
theta.post <- c(theta.post,theta.0)
log.likelihood <- c(log.likelihood, y*log(theta.0) + (n-y)*log(1 - theta.0))
}

# show results
u <- seq(0, 1, length= 200)
par(mfrow=c(1,3))
plot(u, dbeta(u, 10, 12), type="n", ylab="Frequency",
 main="Posterior distribution (burnin = 200)", xlab = "x")
hist(theta.post[200:1000], col = "blue" , prob = T, add = T)
lines(u, dbeta(u, 10, 12), col = 2, lwd = 3)
plot(u, dbeta(u, 10, 12), type="n", ylab="Frequency",
main="150 sweeps", xlab = "x")
hist(theta.post[1:150], col = "yellow" , prob = T, add = T)
lines(u, dbeta(u, 10, 12), col = 2, lwd = 3)
plot(log.likelihood[1:100], cex = .5, pch = 19, type = "l")
```

7. Quel est le choix de du noyau de transition instrumental $Q(\theta_1|\theta_0)$ dans l'algorithme précédent ?
8. Donner l'espérance et la variance de θ_1 sachant la valeur en cours θ_0 (les valeurs pour la loi beta(a, b) sont $a/(a+b)$ et $ab/(a+b)^2(a+b+1)$).
9. Commenter les résultats affichés par le script `R`. L'algorithme est il sensible au choix de la valeur initiale (la modifier) ? L'algorithme est-ill sensible au choix de la loi instrumentale ? Comment vos observations se traduisent-elles pour le choix de la période *burn-in*.

Exercice 2. Echantillonnage de Gibbs. On considère un couple (θ_1, θ_2) de loi gaussienne de moyenne nulle et de matrice de covariance

$$\Lambda = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

On pose $\rho = .99$.

1. Ecrire un algorithme de simulation exact pour le couple (θ_1, θ_2). Le programmer en langage `R`. Que peut-on dire des résultats.
2. Rappeler le principe de l'algorithme d'échantillonnage de Gibbs.
3. Montrer que le script `R` donné ci-dessous implante cet algorithme et commenter sa sortie.

```
rho <- .99
nit <- 2000
theta1 <- -3; theta2 <- 3
theta1.post <- theta1
theta2.post <- theta2

for (i in 1:nit){
theta1 <- rnorm(1, rho*theta2, sd = sqrt(1 - rho^2))
theta2 <- rnorm(1, rho*theta1, sd = sqrt(1 - rho^2))
theta1.post <- c(theta1.post, theta1)
theta2.post <- c(theta2.post, theta2)
}

# show results
qqnorm(theta2.post[100:200])
abline(0,1)
plot(theta1.post[-1], theta2.post[-1])
cor(theta1.post, theta2.post)
hist(theta2.post)
```

4. La vitesse de convergence et les propriétés de mélange de la chaîne de Markov sont-elles dépendantes de ρ ?