TP4 : Improve the Metropolis-Hastings Algorithm.

Exercice 1 : Metropolis-Hastings within Gibbs sampler

$$(x,y) \longmapsto \pi(x,y) \propto \exp\left(-\frac{x^2}{2} - y^2 - \frac{1}{4}\left(\frac{x^2}{2} - y^2\right)^2\right)$$

Q1)

Markov transition kernel $P = \frac{1}{2}(P_1 + P_2)$.

$P_i((x,y); dx' \times dy')$ for $i = 1,2$ is the Markov transition kernel which only updates the ith component, with symmetric random walk proposal mechanism with Gaussian distribution with variance σ_i^2.

Sample from : $\boxed{P_1}$

- Proposal : $X^* = X_n + \sigma_1 \, N(0,1)$
- Ratio : $\alpha_1(X_n, X^*) = \min\left(1, \frac{\pi(X^*, Y_n)}{\pi(X_n, Y_n)}\right)$

$P_1 \Leftrightarrow \begin{cases} Z^* = (X^*, Y_n) \\ Z_n = (X_n, Y_n) \end{cases} \longrightarrow Z_{n+1} = (X_{n+1}, Y_{n+1})$

$= \min\left(1, \frac{\pi_1(X^* | Y_n)}{\pi_1(X_n | Y_n)}\right)$

- $U \sim \mathcal{U}([0,1])$

$$X_{n+1} = \begin{cases} X^* & \text{si } \alpha_1(X_n, X^*) \geq U \\ X_n & \text{sinon} \end{cases}$$

$$Y_{n+1} = Y_n$$

Same for $\boxed{P_2}$

Q2) Sample from kernel P.

- Initialization Z_0
- $Z_n \longrightarrow Z_{n+1}$

$U \sim \mathcal{U}([0,1])$

$$Z_{n+1} \sim P(Z_n) \Longleftrightarrow \begin{cases} Z_{n+1} \sim P_1(Z_n) & \text{si } U \leq \frac{1}{2} \\ Z_{n+1} \sim P_2(Z_n) & \text{sinon} \end{cases}$$

Q3) We can improve the algorithm by :

- adapting (σ_1, σ_2) to the scale of the dimension (a)
- Replace $P = \frac{1}{2}(P_1 + P_2)$ by $P = \alpha P_1 + (1-\alpha) P_2$

Exercise 2) Adaptive Metropolis-Hastings within Gibbs sampler.

Q1) **Algorithm**, NumBatches = 50, Iteration in Batch = 1000

Given $x^{(0)} = (x_1^{(0)}, \dots, x_d^{(0)})$

Set $l = (l_1, \dots, l_d) = (0, \dots, 0)$

$\sigma = (\sigma_1, \dots, \sigma_d) = \exp(l)$

For j in $0, \dots,$ Num_Batches :

for i in $0, \dots,$ Iter_in_Batch :

for k in $1, \dots, d$.

HM to sample from target $x_k^{(t+1)} \sim \pi(x_k \mid x_{-k}^{(t)})$

Proposal:

$x_k^* \sim \mathcal{N}(x_k^{(t)}, \sigma_k^2)$

Acceptance ratio

Symmetric random walk $\rightarrow \alpha_k(x_k^{(t)}, x_k^*) = \dfrac{\pi_k(x_k^* \mid x_{-k}^{(t)})}{\pi_k(x_k^{(t)} \mid x_{-k}^{(t)})}$

when finish the batch update l. $\rightarrow$ for k in $1, \dots, d$:

if acceptance ratio$_k < 0.44$: $l_k = l_k - \delta(j)$

otherwise $l_k = l_k + \delta(j)$

$$\boxed{\text{We define } x_{-i}^{(t)} = (x_1^{(t+1)}, \dots, x_{i-1}^{(t+1)}, x_{i+1}^{(t)}, \dots, x_d^{(t)})}$$

new σ $\rightarrow \sigma = \exp(l)$

The idea of the adaptive Metropolis Hastings is to increase the variance when the acceptance rate is "anormaly" high comparing to 1 dimension and the contrary when the acceptance rate is low. "anormaly".

Exercice 3: Bayesian analyses of a one way random effects model

Inverse Gamma: $x \longmapsto \frac{1}{x^{a+1}} \exp\left(\frac{-b}{x}\right) \mathbb{1}_{\mathbb{R}^+}(x)$

(i) $Y_{i,j} = X_i + \varepsilon_{ij}$

(ii) $X_i \sim N(\mu, \sigma^2)$ There is a typo in the handout

(iii) $\varepsilon_{ij} \sim N(0, \tau^2)$ white noise

Prior: $\pi_{prior}(\mu, \sigma^2, \tau^2) \propto \frac{1}{\sigma^{2(1+\alpha)}} \exp\left(\frac{-\beta}{\sigma^2}\right) \frac{1}{\tau^{2(1+\gamma)}} \exp\left(\frac{-\beta}{\tau^2}\right)$

The prior is not informative on μ [it is degenerated uniform distr. on $\mathbb{R}$].

α, β, γ are known hyperparameters.

1) The density of the a posteriori distribution $(X, \mu, \sigma^2, \tau^2)$ is (up to a normalizing constant)

$$P(X, \mu, \sigma^2, \tau^2 | Y) \propto P(Y | X, \mu, \sigma^2, \tau^2) \times P(X | \mu, \sigma^2) \times \pi_{prior}(\mu, \sigma^2, \tau^2)$$

$$\propto \prod_{i=1}^{N} \prod_{j=1}^{k_i} \underbrace{P(Y_{ij} | X_i, \tau^2)}_{\text{iid} \sim N(X_i, \tau^2)} \times \prod_{i=1}^{N} \underbrace{P(X_i | \mu, \sigma^2)}_{N(\mu, \sigma^2)} \times \underbrace{\pi_{prior}(\mu, \sigma^2, \tau^2)}_{\text{given to us}}$$

2) Gibbs Sampler to update $(\sigma^2, \tau^2, \mu, X)$ at a time

After calculation: we obtain:

- $\forall i \in [1..N]: X_i | \mu, \sigma^2, \tau^2 \sim N\left(\frac{\bar{Y}_i/\tau^2 + \mu/\sigma^2}{\frac{k_i}{\tau^2} + \frac{1}{\sigma^2}}, \frac{1}{\frac{k_i}{\tau^2} + \frac{1}{\sigma^2}}\right)$

And with: $\bar{Y}_i = \sum_{j=1}^{k_i} Y_{ij}$

We can sample the vector X at once using an independent multivariate normal distribution.

- $\mu \mid X, \sigma^2, \tau^2 \sim \mathcal{N}\left(\bar{X}, \frac{\sigma^2}{N}\right)$.

with $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$

- $\tau^2 \mid X, \mu, \sigma^2 \sim \text{Inv-Gamma}\left(\gamma + \frac{Nk}{2}, \beta + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k_i}(Y_{ij} - X_i)^2\right)$

- $\sigma^2 \mid X, \mu, \tau^2 \sim \text{Inv-Gamma}\left(\alpha + \frac{N}{2}, \beta + \frac{1}{2}\sum_{i=1}^{N}(X_i - \mu)^2\right)$

3) In Block Gibbs sampler, we sample the block (X, μ) by a multivariate normal distribution of dimension $N+1$.
Let's see the parameters:

$$P(X, \mu \mid \tau^2, \sigma^2) \propto \prod_{i=1}^{N} \exp\left(\frac{(X_i - \mu)^2}{2\sigma^2}\right) \times \prod_{j=1}^{k_i} \exp\left(\frac{(Y_{ij} - X_i)^2}{2\tau^2}\right)$$

So we have

$$P(X, \mu \mid \tau^2, \sigma^2) \propto \prod_{i=1}^{N} \exp\left(\frac{1}{2}\left[\frac{(X_i - \mu)^2}{\sigma^2} + \frac{1}{\tau^2}\sum_{j=1}^{k_i}(Y_{ij} - X_i)^2\right]\right)$$

$$\propto \prod_{i=1}^{N} \exp\left(\frac{1}{2}\left[\frac{X_i^2}{\sigma^2} + \frac{X_i^2 k_i}{\tau^2} + \frac{\mu^2}{\sigma^2} - \frac{2\mu X_i}{\sigma^2} - \frac{2\bar{Y}_i}{\tau^2} X_i\right]\right)$$

$$\propto \exp \frac{1}{2}\left[\sum_{i=1}^{N} X_i^2\left(\frac{1}{\sigma^2} + \frac{k_i}{\tau^2}\right) - 2\frac{\mu X_i}{\sigma^2} + \frac{N\mu^2}{\sigma^2} - \frac{2\bar{Y}_i}{\tau^2} X_i\right].$$

We conclude that the covariance matrix for the multivariate gaussian is the $(N+1)\times(N+1)$ symetric matrix, we will describe the precision matrix Λ such that

$$\Sigma = (\Lambda)^{-1}$$

$1 \leq i \leq N \Rightarrow$

$$\Lambda_{i,i} = \frac{1}{\sigma^2} + \frac{k_i}{\tau^2}$$

$$\Lambda_{N+1,N+1} = \frac{N}{\sigma^2}.$$

$1 \leq i \neq j \leq N \quad \Lambda_{ij} = 0$

$$1 \leq i \leq N, \quad \Lambda_{i,N+1} = \Lambda_{N+1,i} = \frac{1}{\sigma^2\left(\frac{1}{\sigma^2} + \frac{k_i}{\tau^2}\right)}$$

$$= \frac{1}{1 + k_i\frac{\sigma^2}{\tau^2}}$$

And the mean is: mu the $(N+1)$-vector

with $\quad mu_i = \dfrac{\overline{Y_i}}{\tau^2\left(\frac{1}{\sigma^2} + \frac{k_i}{\tau^2}\right)}$; $1 \leq i \leq N$

and $\quad mu_{N+1} = 0$