Cursus HPC-Big Data 2023 / Analyse bayésienne :

TP n°2 – Algorithme de Metropolis-Hastings

<u>Rappels</u> :

On considère une distribution d'échantillonnage $f(\mathbf{y}|\theta)$ et un prior $f(\theta)$. Le posterior $f(\theta|\mathbf{y})$ n'étant pas toujours calculable, l'algorithme de Metropolis-Hastings permet de générer une chaîne de Markov $\{\theta_1,\ldots,\theta_i,\ldots\}$ ayant $f(\theta|\mathbf{y})$ comme distribution invariante.

L'idée de base est que si Na et Nb sont les nombres d'éléments θ_i de la chaîne de Markov égaux respectivement à $\theta_i = \theta_a$ et $\theta_i = \theta_b$, le rapport Na / Nb doit être proche du rapport $f(\theta_a|\mathbf{y}) / f(\theta_b|\mathbf{y})$.

Si θ_i est le dernier élément connu de la chaîne, soit θ^* la valeur candidate de θ_{i+1} ,valeur tirée aléatoirement selon une distribution de transition instrumentale conditionnelle $g(\theta^*|\theta_i)$.

Le critère d'acceptation de θ^* comme nouvel élément de la chaîne est basé sur la valeur calculable du rapport $r = f(\theta^*|\mathbf{y}) / f(\theta_i|\mathbf{y}) = f(\mathbf{y}|\theta^*) f(\theta^*) / [f(\mathbf{y}|\theta_i) f(\theta_i)]$.

- Si $r > 1$: $f(\theta^*|\mathbf{y}) > f(\theta_i|\mathbf{y})$, la proportion d'éléments de la chaîne égaux à θ^* doit être supérieure à celle relative à θ_i.
  θ_i venant d'être intégrée à la chaîne, il nous faut alors intégrer θ^*.
  → On accepte la valeur candidate : $\theta_{i+1} = \theta^*$.

- Si $r < 1$: $f(\theta^*|\mathbf{y}) < f(\theta_i|\mathbf{y})$, la proportion d'éléments de la chaîne égaux à θ^* doit être inférieure à celle relative à θ_i, et ce dans le rapport r. Il nous faut donc accepter θ^* avec une probabilité égale à r et la rejeter (et donc reconduire la valeur précédente θ_i) avec une probabilité de 1- r.

La valeur candidate θ^* est générée selon une distribution $g(\theta^*|\theta_i)$ en général très simple centrée sur θ_i comme par exemple une loi uniforme sur $[\theta_i - \delta, \theta_i + \delta]$ ou une loi normale $N(\theta_i , \delta^2)$. Le choix du paramètre δ détermine l'efficacité de l'algorithme.

Connaissant θ_i , l'algorithme de Metropolis-Hastings générant θ_{i+1} est donc le suivant :

1) Tirage de θ^* selon la loi de densité $g(\theta^*|\theta_i)$
2) Calcul du rappport $r = f(\mathbf{y}|\theta^*) f(\theta^*) / [f(\mathbf{y}|\theta_i) f(\theta_i)]$
3) $\theta_{i+1} = \theta^*$ avec la probabilité $\min(r,1)$
   $= \theta_i$ avec la probabilité $1 - \min(r,1)$

A noter que le point 3 peut être effectué simplement ainsi :
- tirer une valeur u selon la loi uniforme sur [0,1]
- choisir $\theta_{i+1} = \theta^*$ si $u < r$, sinon rejeter θ^* et choisir $\theta_{i+1} = \theta_i$.

<u>Remarque</u> : Pour des raisons d'instabilité numérique, il est préférable de calculer $\ln(r)$ plutôt que r. On acceptera alors θ^* si $\ln(u) < \ln(r)$.

<u>Exercice</u> : Modèle Normal avec variance connue.

L'algorithme de Metropolis-Hastings va ici être exploité dans un cas où la distribution a posteriori est connue pour pouvoir comparer la chaîne de Markov générée avec sa distribution invariante théorique.

On considère le modèle bayésien normal de distribution d'échantillonnage normale $N(\theta,\sigma^2)$ avec σ^2 connue et de prior normal $N(b,d^2)$. Soit **y** un n-échantillon iid selon $N(\theta,\sigma^2)$.

On utilisera les valeurs numériques suivantes :
y = (9.37 , 10.18 , 9.16 , 11.60 , 10.33) , $\sigma^2 = 1$, $d^2 = 10$, $b = 5$.

a) Déterminer la distribution a posteriori $f(\theta|\mathbf{y})$.

b) Coder l'algorithme de Metropolis-Hastings puis générer une chaîne de Markov à 10000 éléments, de distribution invariante $f(\theta|\mathbf{y})$ et d'élément initial nul.
On utilisera comme distribution de transition instrumentale conditionnelle $g(\theta^*|\theta_i)$ la loi normale $N(\theta_i , \delta^2)$ avec $\delta^2 = 2$.
Calculer la fréquence d'acceptation de la valeur candidate θ^*.

c) Comparer graphiquement les résultats obtenus aux questions précédentes.

d) Illustrer graphiquement l'efficacité de l'algorithme de Metropolis-Hastings selon les valeurs suivantes du paramètre δ^2 = 2, 64 puis 1/32.

e) Analyser l'autocorrélation de la chaîne selon la valeur de δ^2 ?

f) Comparer les fréquences d'acceptation de la valeur candidate θ^*.

g) Interpréter les différents résultats obtenus.
Quelle est la meilleure valeur de δ^2 parmi les trois valeurs testées ?

h) Tester l'algorithme avec une distribution instrumentale conditionnelle uniforme de support $[\theta_i-\delta , \theta_i+\delta]$ avec δ= 1, 100 puis 1/10.