Metropolis-Hastings Algorithm

**Strength of the Gibbs sampler**

- Easy algorithm to think about.
- Exploits the factorization properties of the joint probability distribution.
- No difficult choices to be made to tune the algorithm.

**Weakness of the Gibbs sampler**

- Can be difficult (impossible) to sample from full conditional distributions.

**Idea:** Use acceptance-rejection method instead.

**Metropolis-Hastings Algorithm**

**Aim:**

- Easy algorithm to think about.
- No difficult choices to be made to tune the algorithm.

**Proposal step:** Sample “candidate” \( X \) from the proposal distribution, \( Z \sim q(z|Y^{(t)}) \).

**Acceptance step:** With probability

\[
\alpha(Y^{(t)}, Z) = \min \left\{ \frac{f(Z)}{f(Y^{(t)}) q(Z|Y^{(t)})} \right\}
\]

set

\[
Y^{(t+1)} = Z \quad \text{(acceptance)}
\]

and otherwise set

\[
Y^{(t+1)} = Y^{(t)} \quad \text{(rejection)}.
\]

**Remarks:**

- Suppose we want to sample from the posterior distribution

\[
\pi(\theta|Y) = \frac{f(Y|\theta) \pi(\theta)}{f(Y)} \quad \text{with} \quad f(Y) = \int f(Y|\theta) \pi(\theta) d\theta
\]

Then

\[
\frac{\pi(\theta')|Y)}{\pi(\theta|Y)} = \frac{f(Y|\theta') \pi(\theta')}{f(Y|\theta) \pi(\theta)}
\]

that is, the normalising constant is not required to run the algorithm.

- Usually the proposal distribution \( q \) is chosen such that it is easy to sample from it.
- If the proposal distribution is symmetric,

\[
q(z|y) = q(y|z)
\]

we obtain the Metropolis algorithm. In this case

\[
\alpha(Y^{(t)}, Z) = \min \left\{ \frac{f(Z)}{f(Y^{(t)})} \right\}
\]

**Interpretation:**

- Proposal state \( Z \) with higher probability are always accepted.
- Change to state with lower probability possible with probability \( \alpha \).

**Special case:** Random-walk Metropolis

\[
q(z|y) = q(j = \pm 1)
\]

- Any density \( q \) that has the same support should work.
- **However:** Some distributions are better than others.

**Tuning Metropolis-Hastings**

We need to find a good proposal distribution

- with high acceptance rate,
- which allows to reach all states frequently (good mixing).

**Example:** Binomial distribution with non-standard prior

\[ Y = (Y_1, \ldots, Y_n)^T \text{ with } Y_i \sim \text{Bin}(1, \theta) \]

\[ S_n = \sum_{i=1}^n Y_i \]

\[ \pi(\theta) = 2 \cos(4\pi \theta) \]

Then the posterior is

\[ \pi(\theta|Y) \sim f(Y|\theta) \pi(\theta) = 2 \theta^n (1 - \theta)^{S_n - S_n} \cos(4\pi \theta) \]

**Proposal distribution:**

\[ q(\theta'|\theta) \sim \exp \left( \frac{1}{2\sigma^2} (\theta - \theta')^2 \right) \]

**Acceptance probability:**

\[
\alpha(\theta, \theta') = \min \left\{ \frac{\pi(\theta'|Y) q(\theta|\theta')}{\pi(\theta) q(\theta'|\theta)} \right\} = \min \left\{ \frac{\theta'^n (1 - \theta')^{S_n - S_n} \cos(4\pi \theta)}{\theta^n (1 - \theta)^{S_n - S_n} \cos(4\pi \theta')} \right\}
\]

Note: Proposal distribution does not depend on \( \theta' \)

\[ \Rightarrow \text{independence sampler} \]
Metropolis-Hastings Algorithm

Example: Bivariate normal distribution
Sample from bivariate normal distribution:
- \( Y = (Y_1, Y_2)^T \sim \mathcal{N}(0, \Sigma) \)
- \( \text{corr}(Y_1, Y_2) = 0.99 \)

Proposal distribution
\( q(Y, Y') \sim \exp \left( -\frac{1}{2\sigma^2} |Y - Y'|^2 \right) \)

Results:

Simulated Annealing

Problem:
For large \( k \)
- \( f^{(k)} \) concentrates on \( x^* \)
- transition between states can be extremely difficult
- chain might become trapped in local mode.

Idea: gradually “cool down” - simulated annealing
- start with \( k < 1 \)
- increase \( k \) slowly (“temperature” \( \tau = k^{-1} \) decreases)
- keep track of the best \( x \) (since we might leave it and not come back)
- optimal cooling scheme
\( \tau(t) = \frac{c}{\log(1 - t)} \)
guarantees that chain converges to maximum: \( Y^{(t)} \to y^* \) with probability 1.

Example: t distribution

MCMC Summary

Software
BUGS (Bayesian inference Using Gibbs Sampling)
http://www.mrc-bsu.cam.ac.uk/bugs/

Strength of MCMC
- Freedom in modelling
- Freedom in inference
- Opportunities for simultaneous inference
- Allows sensitivity analysis
- Model comparison/criticism/choice

Weaknesses of MCMC
- Order \( N^{-\frac{1}{2}} \) precision
- Possibility of slow convergence
- Difficulty in detecting slow convergence