Multiple-try MCMC without Rejection

Quan Zhou

Department of Statistics
Texas A&M University
In this note, I explain how the famous multiple-try Metropolis (MTM) algorithm [5] can be turned into a rejection-free MCMC method without extra computational cost.

For details, see our paper [arxiv] (see Algorithm 7 therein) or my other slides at https://web.stat.tamu.edu/~quan/.
MTM algorithm

- $\mathcal{X}$: a general state space.
- $Q(x, \cdot)$: a proposal distribution given current state $x \in \mathcal{X}$.
- $q(x, \cdot)$: density function of $Q(x, \cdot)$; we assume $q(x, y) = q(y, x)$.
- $\pi$: density function of the target distribution.

MTM is essentially a Metropolis-Hastings algorithm with a complicated proposal scheme. Instead of simply proposing one state from $Q(x, \cdot)$, MTM proposes multiple candidate moves (i.e., multiple “tries”) and then assign larger proposal probabilities to states with larger $\pi$. We assume the weight of $y$ given current state $x$ is proportional to

$$h \left( \frac{\pi(y)}{\pi(x)} \right)$$

for some function $h: (0, \infty) \rightarrow (0, \infty)$. 
**MTM algorithm**

An iteration of MTM at state $x$ with $m$ tries:

1. Draw $y_1, \ldots, y_m$ from $Q(x, \cdot)$.
2. Select $y$ from $y_1, \ldots, y_m$ with probability $\propto h \left( \frac{\pi(y)}{\pi(x)} \right)$.
3. Draw $x_1, \ldots, x_{m-1}$ from $Q(y, \cdot)$. Set $x_m = x$.
4. Accept $y$ with probability

   $$\min \left\{ 1, \frac{Z_h(x, y_1, \ldots, y_m)}{Z_h(y, x_1, \ldots, x_m)} \right\},$$

   where $Z_h(x, y_1, \ldots, y_m) = \sum_{k=1}^m h \left( \frac{\pi(y_k)}{\pi(x)} \right)$. 


Choice of $h$

The recent studies [2, 1] suggest that one wants to choose $h$ such that

$$h(u) = u h(u^{-1}), \quad \forall u > 0.$$  

Such a function is called a balancing function. Examples include

$$h(u) = 1 + u, \quad h(u) = \sqrt{u}, \quad h(u) = \min\{1, u\}.$$  

We assume $h$ is a balancing function henceforth.
Our Multiple Try Importance Tempering algorithm

An iteration of MT-IT at state $x$ with $m$ tries, $y_1, \ldots, y_m$.

1. Select $y$ from $y_1, \ldots, y_m$ with probability $\propto h\left(\frac{\pi(y)}{\pi(x)}\right)$.

2. Accept $y$. Assign to the previous state $x$ (un-normalized) importance weight $Z_h(x, y_1, \ldots, y_m)^{-1}$.

3. Draw $x_1, \ldots, x_{m-1}$ from $Q(y, \cdot)$. Set $x_m = x$. In the next iteration, we use $x_1, \ldots, x_m$ as the $m$ tries at state $y$.

The only differences from MTM are that (1) we always accept $y$, (2) we calculate importance weight instead of acceptance probability (note both rely on evaluating the function $Z_h$).
Why is it correct?

Let’s consider the dynamics of the state \((x, \{y_1, \ldots, y_m\}) \in \mathcal{X} \times \mathcal{X}_m\), where \(\mathcal{X}_m\) is the collection of all unordered subsets of \(\mathcal{X}\) with \(m\) elements. It is a Markov chain with transition density

\[
p((x, \{y_1, \ldots, y_m\}), (y, \{x_1, \ldots, x_m\})) = A_1 A_2,
\]

where

\[
A_1 = \frac{h \left( \frac{\pi(y)}{\pi(x)} \right)}{Z_h(x, y_1, \ldots, y_m)}, \quad A_2 = \prod_{k=1}^{m-1} q(y, x_k).
\]

\(A_1\) corresponds to how we select \(y\) from \(\{y_1, \ldots, y_m\}\), and \(A_2\) corresponds to how we generate \(x_1, \ldots, x_m\). Note that \(x_m\) is fixed to be \(x\)!
**Why is it correct?**

$p$ satisfies the detailed balance condition w.r.t. the stationary distribution

\[
\pi_h(x, \{y_1, \ldots, y_m\}) \propto \pi(x) Z_h(x, y_1, \ldots, y_k) \prod_{k=1}^{m} q(x, y_k).
\]

Comparing this to a reference distribution

\[
\bar{\pi}(x, \{y_1, \ldots, y_m\}) = \pi(x) \prod_{k=1}^{m} q(x, y_k),
\]

we see that $Z_h(x, y_1, \ldots, y_k)^{-1}$ is the importance weight we need.
Why is it correct?

Here is the proof of the detailed balance condition. Since \((y_1, \ldots, y_m)\) is treated as an unordered set, we can assume \(y = y_k\). Then,

\[
\pi_h(x, \{y_1, \ldots, y_m\})p((x, \{y_1, \ldots, y_m\}), (y, \{x_1, \ldots, x_m\})) = \pi(x)q(x, y)h\left(\frac{\pi(y)}{\pi(x)}\right) \prod_{k=1}^{m-1} q(y, x_k) \prod_{k=1}^{m-1} q(x, y_k).
\]

It only remains to show that

\[
\pi(x)q(x, y)h\left(\frac{\pi(y)}{\pi(x)}\right) = \pi(y)q(y, x)h\left(\frac{\pi(x)}{\pi(y)}\right),
\]

which holds since \(q\) is assumed symmetric and \(h\) is a balancing function.
Why is it correct?

Caveat: Fixing $x_m = x$ is important! This guarantees that the transition density from $(x, \{y_1, \ldots, y_m\})$ to $(y, \{x_1, \ldots, x_m\})$ is nonzero only if

$$y \in \{y_1, \ldots, y_m\} \text{ and } x \in \{x_1, \ldots, x_m\}.$$ 

Without this symmetry, reversibility fails and the stationary distribution of the chain is unclear.
References


