Rejection Sampling

Metropolis-Hastings Algorithm

- Target distribution on $\mathbb{X} = \mathbb{R}^d$ of density $\pi(x)$.
- Proposal distribution: for any $x, x' \in \mathbb{X}$, we have $q(x'|x) \geq 0$ and $\int_{\mathbb{X}} q(x'|x)\, dx' = 1$.

Algorithm

1. Starting with $X^{(1)}$, for $t = 2, 3, \dots$
2. Sample $X^\star \sim q(\cdot|X^{(t-1)})$.
3. Compute $\alpha(X^\star|X^{(t-1)}) = \min\left(1, \frac{\pi(X^\star)q(X^{(t-1)}|X^\star)}{\pi(X^{(t-1)})q(X^\star|X^{(t-1)})}\right)$.
4. Sample $U \sim \mathcal{U}_{[0,1]}$. If $U \leq \alpha(X^\star|X^{(t-1)})$, set $X^{(t)} = X^\star$, otherwise set $X^{(t)} = X^{(t-1)}$.

Metropolis-Hastings Algorithm

- The proposal distribution q can be chosen quite freely, but it should be easy to sample from and to evaluate.
- The acceptance probability $\alpha(x'|x)$ ensures that the chain has π as its stationary distribution.
- If q is symmetric, i.e., $q(x'|x) = q(x|x')$, then the acceptance probability simplifies to $\alpha(x'|x) = \min\left(1, \frac{\pi(x')}{\pi(x)}\right)$. This is also known as Metropolist Random Walk.
- If q does not depend on the current state, i.e., $q(x'|x) = q(x')$, then the algorithm reduces to the Independent Metropolis-Hastings algorithm.
- The choice of q affects the efficiency of the algorithm. A poorly chosen q can lead to slow mixing and high autocorrelation in the samples.

Role of $\alpha(X^\star|X^{(t-1)})$

- The acceptance probability $\alpha(X^\star|X^{(t-1)})$ is crucial for ensuring that the Markov chain has the desired stationary distribution π.
- It corrects for the discrepancy between the proposal distribution q and the target distribution π.
- If the proposed state $X^\star$ has a higher density under π than the current state $X^{(t-1)}$, it is always accepted ($\alpha = 1$).
- If $X^\star$ has a lower density, it may still be accepted with a probability proportional to the ratio of densities, allowing exploration of the state space.
- This mechanism helps to avoid getting stuck in local modes and promotes better mixing of the chain.

Role of $\alpha(X^{\star}|X^{(t-1)})$

- If $\pi(x^{*}) > \pi(x)$, then the proposed state has higher probability than current state which favors acceptance
- If $\pi(x^{*}) < \pi(x)$, then the proposed state has lower probability than current state which favors rejection
- Intuition: want to spend time in high-probability regions, so moves towards them should be favored
- Reverse proposal ratio: how easy is it to propose reverse move compared to forward move?
- Intuition: if proposal mechanism makes it easy to reacg certain states, we need to be more selective accepting such moves. otherwise the chain will be biased towards such easy-to-propos-to regions rather than high probability regions

Transition Kernel

Lemma

The kernel of the Metropolis–Hastings algorithm is given by

$$K(x, y) = \alpha(y \mid x)q(y \mid x) + (1 - a(x))\delta_x(y).$$

Proof: We have

$$\begin{aligned}
K(x, y) &= \int q(x^* \mid x)\{\alpha(x^* \mid x)\delta_{x^*}(y) + (1 - \alpha(x^* \mid x))\delta_x(y)\}\, dx^* \\
&= q(y \mid x)\alpha(y \mid x) + \left\{\int q(x^* \mid x)(1 - \alpha(x^* \mid x))\, dx^*\right\}\delta_x(y) \\
&= q(y \mid x)\alpha(y \mid x) + \left\{1 - \int q(x^* \mid x)\alpha(x^* \mid x)\, dx^*\right\}\delta_x(y) \\
&= q(y \mid x)\alpha(y \mid x) + \{1 - a(x)\}\delta_x(y).
\end{aligned}$$

Reversibility

Proposition

The Metropolis–Hastings kernel K is π-reversible and thus admits π as invariant distribution.

Proof: For any $x, y \in \mathbb{X}$, with $x \neq y$

$$
\begin{aligned}
\pi(x)K(x,y) &= \pi(x)q(y \mid x)\alpha(y \mid x) \\
&= \pi(x)q(y \mid x)\left(1 \wedge \frac{\pi(y)q(x \mid y)}{\pi(x)q(y \mid x)}\right) \\
&= (\pi(x)q(y \mid x) \wedge \pi(y)q(x \mid y)) \\
&= \pi(y)q(x \mid y)\left(\frac{\pi(x)q(y \mid x)}{\pi(y)q(x \mid y)} \wedge 1\right) = \pi(y)K(y,x).
\end{aligned}
$$

If $x = y$, then obviously $\pi(x)K(x,y) = \pi(y)K(y,x)$.