REJECTION SAMPLING

Assume you want to sample from a distribution $p$ but you don't know how to. There are many reasons why you would want to sample. For instance, maybe you want to compute $E[X]$

$$E[X] = \int x \, p(x) \, dx$$

But you can't integrate. If you have enough samples $x_1, x_2, \ldots, x_n$ from $p(x)$, then

$$\frac{\sum x_i}{n} \rightarrow E[X]$$

Instead of $p$, assume you know how to sample from another distribution $q$ such that:

$$Cq(x) \rightarrow p(x)$$

For some constant $C$.

REJECTION SAMPLING:

Sample $x \sim q(x)$

Accept sample with prob. $\frac{p(x)}{Cq(x)}$

$$P(x \leq x \leq x+\delta) \propto q(x) \cdot \frac{p(x)}{Cq(x)} \propto q(x)$$
With rejection sampling, we know \( p(x) \) but we don't know how to sample from \( p \). But what if we don't even know \( p \) itself? Why is that a concern? This is typical scenario in Bayesian analysis. Recall

\[
f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}
\]

So \( f(\theta) \) posterior is really not known unless we know how to integrate \( \int f(x|\theta)f(\theta)d\theta \). The Metropolis-Hastings algorithm is a kind of rejection sampling that can overcome this problem. Consider a Markov chain with underlying transitions \( p(y|x) \).

Assume our desired distribution (which we don't know) is \( \pi(x) \). \( p \) and \( \pi \) have the same state space. If \( \pi \) is reversible for our Markov chain, then it will be the stationary distribution.

\[p(y|x)\]

\[X \quad Y\]

Assume on the other hand that

\[\pi(x)p(y|x) > \pi(y)p(x|y)\]

Then multiplying the left hand side by \( \frac{\pi(y)p(x|y)}{\pi(x)p(y|x)} \) makes both sides equal (reversibility condition).

Effectively, we change \( p(y|x) \) to \( \frac{\pi(y)p(x|y)}{\pi(x)p(y|x)} \).
At time $n$, sample $y \sim p(y|x_n)$ with prob $\alpha = \min \left[ \frac{\pi(y) p(x_n|y)}{\pi(x_n) p(y|x_n)}, 1 \right]$.

Make $x_{n+1} = y$ (move) else $x_{n+1} = x_n$ (stay).

This is the Metropolis-Hastings algorithm, which is the basis for Markov Chain Monte Carlo (MCMC).

**Symmetric MCMC**

$p(y|x) = p(x|y)$, then

$\alpha = \min \left[ \frac{\pi(y)}{\pi(x_n)}, 1 \right]$.

**Independent MCMC**

$p(y|x) = p(y)$, then

$\alpha = \min \left[ \frac{\pi(y) p(x)}{\pi(x) p(y)} \right]$.

This form can be used in Bayesian analysis to use the likelihood ratio as a basis for moving in the Markov chain.

**Example:**

$x_i; \mu \sim N(\mu, \sigma^2)$

We observe $x$ and we have a prior $f(\mu) \propto \frac{1}{1 + \mu^2}$.
\[ f(\mu|x) \propto f(x|\mu) f(\mu) \]
\[ \propto e^{-\frac{(x-\mu)^2}{2\sigma^2/n}} \cdot \frac{1}{1+\mu^2} \]

This is our (unknown) desired distribution.

With an independent MCMC with underlying transition \( \alpha \cdot \frac{1}{1+\mu^2} \), we compute \( \alpha \) as

\[ \alpha = \min \left[ \frac{e^{-\frac{(\mu_2-x)^2}{2\sigma^2/n}}}{e^{-\frac{(\mu_1-x)^2}{2\sigma^2/n}}}, 1 \right] \]

\[ \alpha = \min \left[ \frac{1}{1+\mu^2}, 1 \right] \]

where \( \mu_1 \) is current sample and \( \mu_2 \) is the next sample to be accepted with prob. \( \alpha \)

So the basic idea here is to use an independent MCMC with underlying transitions based on the prior.
GIBBS SAMPLER

Gibbs sampler is a special case MCMC where Q = 1 (always accept). So the chain is already reversible. It is useful when sampling multiple values.

Assume \( (x_1^t, \ldots, x_n^t) \) is our sample at time \( t \). To obtain a sample at time \( t+1 \):

1. Sample \( x_1^{t+1} \sim p(x_1^t \mid x_2^t, \ldots, x_n^t) \)
2. Sample \( x_2^{t+1} \sim p(x_2^t \mid x_1^{t+1}, x_3^t, \ldots, x_n^t) \)
3. Sample \( x_3^{t+1} \sim p(x_3^t \mid x_1^{t+1}, x_2^{t+1}, x_4^t, \ldots, x_n^t) \)
   
   : 

   Sample \( x_n^{t+1} \sim p(x_n^t \mid x_1^{t+1}, \ldots, x_{n-1}^{t+1}) \)

We can show that \( p \) is the joint distribution is reversible for the underlying Markov chain.

\[
\begin{align*}
p(x_j = b \mid x_{j+1}, \ldots, x_n) & = p(x_j = a \mid x_{j-1}, x_{j+2}, \ldots, x_n) \\
p(x_j = b \mid x_{j-1}, \ldots, x_n) & = p(x_j = a \mid x_{j-1}, x_{j+2}, \ldots, x_n) \\
p(x_j = b \mid x_{j-1}, \ldots, x_n) & = \frac{p(x_j = a \mid x_{j-1}, x_{j+2}, \ldots, x_n) p(x_{j-1}, \ldots, x_{j+1}) p(x_{j+2}, \ldots, x_n)}{p(x_j = b \mid x_{j-1}, x_{j+2}, \ldots, x_n)} \\
& = p(x_j = a \mid x_{j-1}, x_{j+2}, \ldots, x_n) p(x_{j-1}, \ldots, x_{j+1}) p(x_{j+2}, \ldots, x_n)
\end{align*}
\]
The Gibbs sampler is useful when the joint probability is hard to deal with, but marginal prob. are easy.

**Example:**

\[ X_i | \mu, \sigma^2 \sim N(\mu, \sigma^2) \]

**Assume** \( \frac{S}{\sigma^2} \sim \chi_k^2 \)

**And** \( \mu \sim N(\beta, \sigma^2) \)

We know that \( \frac{\mu}{\bar{x}, \sigma^2} \sim N\left(\frac{\frac{1}{n} \mu^2 + \frac{1}{n} \bar{x}^2}{\frac{1}{n} \sigma^2 + \frac{1}{n} \sigma^2}, \frac{\sigma^2}{\frac{1}{n} \sigma^2 + \frac{1}{n} \sigma^2}\right) \)

**And** \( \frac{\frac{S}{\sigma^2} + \frac{S}{\sigma^2}}{\sigma^2} \sim \chi_{k+n}^2 \) where \( S = \sum (x_i - \mu)^2 \)

So we can sample \((\mu, \sigma)\) from posterior distribution by using a Gibbs sampler.

- **Start with arbitrary** \((\mu^0, \sigma^0)\)
- **Repeat for** \(t=1,2,3,\ldots\)
  - Sample \(\mu^{t+1}\) given \(\sigma^t\)
  - Sample \(\sigma^{t+1}\) given \(\mu^{t+1}\)