General Sampling Methods

Adrian Dobra
adobra@uw.edu

Assistant Professor of Statistics

May 25, 2012

Accept/Reject Methods

Sample from $f(x)$ that is available up to a normalizing constant. Assume we can find another distribution $g(x)$ we know how to sample from such that there exists a constant $M > 0$ with

$$f(x) \leq Mg(x).$$

Rejection Sampling

(a) Draw $x \sim g(\cdot)$.

(b) Draw $U \sim Uniform(0,1)$. If $U \leq \frac{f(x)}{Mg(x)}$, we accept and return x. Otherwise we go back to step (a).

This simulates from $f(x)$ because

$$P(Accept) = \int P(Accept|X = x)g(x)dx = \int \frac{f(x)}{Mg(x)}g(x)dx = \frac{1}{M},$$

$$P(x|Accept) = \frac{P(Accept|x)P(x)}{P(Accept)} = \frac{f(x)}{Mg(x)}g(x)M = f(x).$$

Let $X \sim g(x)$. We want to simulate from the truncated distribution:

$$f(x) = g(x) \cdot I_{\{x \in A\}}.$$

For example, take $A = [c, \infty)$. Then $\{x \in A\} = \{x > c\}$. We have

$$\frac{f(x)}{g(x)} = \frac{1}{I_{\{x \in A\}}} = M.$$

Crude accept/reject:

1. Generate $X \sim g(x)$ until $X \in A$.
2. Return X.

This is very inefficient for simulating truncated Normal random variables.

$$f(x) = \frac{1}{B(\alpha_1, \alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1},\ 0 \leq x \leq 1,$$

has maximum at $x_{\alpha_1,\alpha_2} = \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2}$. Take $M = f(x_{\alpha_1,\alpha_2})$. Then

$$f(x) \leq Mg(x),$$

with $g(x) = 1$. Accept/Reject sampling is:

- Generate $U_1, U_2 \sim \mathit{Uniform}(0,1)$ until $M \cdot U_2 \leq f(U_1)$.
- Return U_1.

The double exponential has density $g(x) = \frac{1}{2}\exp -|x|$. We generate $Y \sim Exp(1)$ and $U \sim Uniform(-0.5, 0.5)$. Then $Y^* = sign(U) \cdot Y \sim g(x)$. The ratio between the $Normal(0,1)$ density and $g(x)$ is

$$\frac{f(x)}{g(x)} = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2}x^2 + |x|\right) \leq \sqrt{\frac{2e}{\pi}} \approx 1.3155 = M.$$

The rejection test is

$$U > \exp\left(-\frac{1}{2}x^2 + |x| - \frac{1}{2}\right) = \exp\left(-\frac{1}{2}(|x|-1)^2\right).$$

The sampling algorithm goes as follows:

1. Generate $U_1, U_2, U_3 \sim Uniform(0,1)$.
2. Take $X = -\log(U_1)$.
3. If $log(U_2) > -\frac{1}{2}(|x|-1)^2$, go to Step 1.
4. If $U_3 \leq 0.5$, take $X \leftarrow -X$.
5. Return X.

The Metropolis-Hastings Algorithm

Want to sample from $X \sim f(x)$. Assume we know how to sample from a conditional density $q(\cdot|x)$ such that $f(y)/q(y|x)$ is known up to a constant independent of x. The M-H algorithm produces a Markov chain with stationary distribution $P(X \leq x)$ as follows:

- Given $x^{(t)}$, generate $Y_t \sim q(y|x^{(t)})$.
- Take

$$X^{(t+1)} = \begin{cases} Y_t, & \text{with probability } \rho(x^{(t)}, Y_t), \\ x^{(t)}, & \text{with probability } 1 - \rho(x^{(t)}, Y_t). \end{cases}$$

  where

$$\rho(x, y) = \min\left\{1, \frac{f(y)}{f(x)} \cdot \frac{q(x|y)}{q(y|x)}\right\}.$$

The independent M-H algorithm is obtained for a proposal independent of the current state, i.e. $q(y|x) = q(y)$.

Let X be a r.v. that takes values $\{1, 2, \ldots, K\}$ with probabilities $p_i = P(X = i)$. An independent M-H algorithm starts with some arbitrary value i_0 and uses a uniform proposal $q(i) = 1/K$, $1 \leq i \leq K$. If the chain is currently at $i^{(t)}$, draw i^* uniformly from $\{1, 2, \ldots, K\}$. Set $i^{(t+1)} = i^*$ with probability

$$\min\{p_{i^*}/p_{i^{(t)}}, 1\}.$$

Otherwise, set $i^{(t+1)} = i^{(t)}$.

Assume a flat prior $\pi(\beta) \propto 1$. The posterior distribution of the coefficients β of the probit model is proportional with the likelihood

$$\pi(\beta|y, X) \propto \prod_{i=1}^{n} \Phi((x^i)^T\beta)^{y_i}[1 - \Phi((x^i)^T\beta)]^{1-y_i},$$

A M-H algorithm for simulating from $\pi(\beta|y, X)$ goes as follows:

- Initialization: compute the MLE $\widehat{\beta}$ and the asymptotic covariance matrix $\widehat{\Sigma}$ of $\widehat{\beta}$.
- Set the starting value at the MLE: $\beta^{(0)} = \widehat{\beta}$.
- At iteration $t \geq 1$ do
  1. Generate $\tilde{\beta} \sim N_k(\beta^{(t-1)}, \tau^2\widehat{\Sigma})$.
  2. Calculate $\rho(\beta^{(t-1)}, \tilde{\beta}) = \min\{1, \pi(\tilde{\beta}|y, X)/\pi(\beta^{(t-1)}|y, X)\}$.
  3. With probability $\rho(\beta^{(t-1)}, \tilde{\beta})$, take $\beta^{(t)} = \tilde{\beta}$. Otherwise take $\beta^{(t)} = \beta^{(t-1)}$.

Assume that the random variables X and Y have a joint distribution with density $f_{X,Y}(x,y)$. The Hammersley-Clifford Theorem says that the conditional densities $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$ uniquely define the joint density of X and Y:

$$f_{X,Y}(x,y) = \frac{f_{Y|X}(y|x)}{\int \left[f_{Y|X}(y|x)/f_{X|Y}(x|y) \right] \, dy}$$

The two-stage Gibbs sampler generates a Markov chain (X_t, Y_t) as follows:

- Choose a starting value $X_0 = x_0$.
- At iteration $t \geq 1$ do
  1. Generate $Y_t \sim f_{Y|X}(\cdot|x_{t-1})$.
  2. Generate $X_t \sim f_{X|Y}(\cdot|y_t)$.

The sequences $(X^{(t)})_t$, $(Y^{(t)})_t$ are Markov chains with stationary distributions $f_X(x) = \int f_{X,Y}(x,y)\, dy$ and $f_Y(y) = \int f_{X,Y}(x,y)\, dx$.

Let

$$(X, Y) \sim N_2\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right)\right).$$

Given y_t, the Gibbs sampler generates:

1. $X_{t+1}|y_t \sim N(\rho y_t, 1-\rho^2)$.
2. $Y_{t+1}|x_{t+1} \sim N(\rho x_{t+1}, 1-\rho^2)$.

The Multi-Stage Gibbs Sampler

Sampling from the joint distribution of p variables based on their full conditionals $f_1, \ldots, f_p$ of each variable given the rest proceeds as follows:

- Start with $x^{(0)} = (x_1^{(0)}, \ldots, x_p^{(0)})$.
- At iteration $t \geq 1$ do
  1. Generate $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \ldots, x_p^{(t)})$.
  2. Generate $X_2^{(t+1)} \sim f_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)})$.

     $\vdots$
  3. Generate $X_p^{(t+1)} \sim f_p(x_2 | x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})$.

Let

$$(X_1, X_2, \ldots, X_p)^T \sim N_p\left(0, (1-\rho)I_p + \rho J_p\right),$$

where I_p is the $p \times p$ identity matrix and J_p is the $p \times p$ matrix with all elements equal to 1. This is the equicorrelation model, i.e.

$$\mathrm{Cor}(X_i, X_j) = \rho, \text{ for any } 1 \leq < j \leq p.$$

The Gibbs sampler proceeds by sequentially sampling from each conditional distribution

$$X_i | x_{-i} \sim N\left(\frac{(p-1)\rho}{1+(p-2)\rho}\bar{x}_{-i}, \frac{1+(p-2)\rho-(p-1)\rho^2}{1+(p-2)\rho}\right).$$

where $\bar{x}_{-i}$ is the mean of $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$.

Let Y_1, Y_2, Y_3 be random variables taking positive real values with joint density

$$f(y_1, y_2, y_3) \sim \exp\{-(y_1 + y_2 + y_3 + \theta_{12} y_1 y_2 + \theta_{13} y_1 y_3 + \theta_{23} y_2 y_3)\}.$$

The full conditionals are exponential:

$$\begin{aligned} Y_1 | y_2, y_3 &\sim Exp(1 + \theta_{12} y_2 + \theta_{13} y_3), \\ Y_2 | y_1, y_3 &\sim Exp(1 + \theta_{12} y_1 + \theta_{23} y_3), \\ Y_3 | y_1, y_2 &\sim Exp(1 + \theta_{13} y_1 + \theta_{23} y_2). \end{aligned}$$