STAT 534
Lecture

2019.5.28

mmp@stat.washington.edu
Scribes: Xinyun Dai, Cherie Wu

1 Rejection Sampling

- Goal: Get independent samples from $f(x)$ on domain Ω
- Given: $g(x)$, $\frac{f(x)}{g(x)} \leq M$, $g >> f <=> f > 0 => g > 0$
- Algorithm

Algorithm 1 Rejection sampling algorithm

procedure REJECTION SAMPLING()
  `sample x` $\sim$ `g,`
  `sample u` $\sim$ `uniform[0,1]`
  `until u` $\leq \frac{f(x)}{g(x)} \leq 1$
  `return x`

Notice that the acceptance probability $P[accept] \propto \frac{1}{M}$:

$$P[accept] = \int_\Omega g(x) P[accept|x] dx = \int_\Omega \frac{f(x)}{Mg(x)} = \frac{1}{M}$$

This property has implication of how to choose proper $g(x)$. A proper $g(x)$ will yield a smaller M value so that the acceptance probability will not be too small.

Also, It can be confirmed that the sample generated using the procedure above is actually from the true $f(x)$ distribution:

$$\hat{f}(x) = \frac{g(x)a(x)}{P[accept]} = g(x)\frac{f(x)}{Mg(x)} = f(x)$$

- Example: Sampling from "tail" of $f(x)$
  Define $I_{(x \in A)}$ where $A = \{x||x| > c\}$ for some constant c. Then from the relationship between $f(x)$ and $g(x)$:

$$f(x) \propto g(x) I_{[x \in A]}$$

So the target $f(x)$ can be calculated as below:

$$f(x) = \frac{g(x)I_{x \in A}}{P_g[A]}$$

Then M can be calculated as:

$$M = \frac{1}{P_g[A]}$$

It can be seen that if the tail probability of $f(x)$ is very small, then M is a very large number which means that the acceptance probability will be very small. The algorithm will not be efficient.

2 Importance Sampling

- Goal: $E_f[h(x)] \quad (x \in \Omega, h : \Omega \in \mathbb{R})$
- Given: $g(x)$, $g > 0$ whenever $f > 0$
- Algorithm

Algorithm 2 Importance sampling algorithm

procedure IMPORTANCE SAMPLING()
  for `t=1:T` **do**
    sample $x_t \sim g$
    $\bar{\bar{h}} = \frac{1}{T} \sum_{t=1}^{T} h(X_t) \frac{f(x)}{g(x)}$

It can be confirmed that $E_g[\bar{\bar{h}}]$ is actually $E_f[h]$:

$$E_f[\bar{h}] = \int_{\Omega} h(x) f(x) dx \approx \bar{h} = \frac{1}{T} \sum_{t=1}^{T} h(X_t)$$

$$\begin{aligned} E_g[\bar{\bar{h}}] &= \frac{1}{T} \sum_{t=1}^{T} E_g[h(X_t) \frac{f(x)}{g(x)}] \\ &= E[h(X_t) \frac{f(x)}{g(x)}] \\ &= \int_{\Omega} h(x) \frac{f(x)}{g(x)} dx \\ &= E_f[h] \end{aligned}$$

Notice that $Var(\bar{\bar{h}}) \propto \frac{1}{T}$:

$$Var(\bar{\bar{h}}) = \frac{1}{T^2} \times T \times Var(\frac{h(x)f(x)}{g(x)}) \propto \frac{1}{T}$$

Since we want the variance to be small, then ideally: $g(x) \propto h(x)f(x)$

3 Metropolis-Hastings Sampling

- Goal: Simulate samples from target distribution π
- Given: Proposal distribution $g(x^{'}|x)$
- Algorithm

Algorithm 3 Metropolis-Hastings algorithm

procedure METROPOLIS-HASTINGS()
  `Initialize` $x^{(0)} \sim q(x)$
  for `iteration t=1,2,3...` **do**
    sample $x^{'} \sim q(x^{'}|x_{t-1})$ ▷ proposal distribution
    $a = min(1, \frac{\pi(x^{'})q(x_{t-1}|x^{'})}{\pi(x_{t-1})q(x^{'}|x_{t-1})})$ ▷ Acceptance Probability
    $x_t = \begin{cases} x^{'} & \text{with probability } a \\ x_{t-1} & \text{with probability } 1-a \end{cases}$

- Example: Bayesian estimation for Probit Model
  Given that $\mathbf{x} \in \mathbb{R}^d, y \in \{0,1\}$, and $P[y=1|\mathbf{x}] = \Phi(\beta^T\mathbf{x})$ where Φ is the cdf of standard normal distribution.

  The goal is to find the posterior distribution of β

  $$P(\beta|(x_i, y_i)_{i=1:n}, \pi_0)$$

  where $(x_i, y_i)_{i=1:n}$ is the data and π_0 is the prior distribution.

  Notice that choosing the uniform distribution as the prior distribution is not a good idea. The better choice is $\pi_0 \sim N(0, \Sigma)$ with large Σ

  First, we initiate the proposal distribution

  $$q(\beta^{'}|\beta) = N(\beta, \tau^2\Sigma)$$

  Compute the maximum likelihood estimator $\hat{\beta}$ and $var(\hat{\beta}) = \Sigma$.

Algorithm 4 Bayesian Estimation for Probit Model

procedure METROPOLIS-HASTINGS()
  $\beta_0 = \hat{\beta}$
  for `iteration t=1,2,3...` **do**
    sample $\beta' \sim N(\beta, \tau^2\Sigma)$ ▷ proposal distribution
    $a = min(1, \frac{\pi(\beta^{'})N(\beta_{t-1}, \beta', \tau^2\Sigma)}{\pi(\beta_{t-1})N(\beta', \beta_{t-1}, \tau^2\Sigma)})$ ▷ Acceptance Probability
    $\beta_t = \begin{cases} \beta^{'} & \text{with probability } a \\ \beta_{t-1} & \text{with probability } 1-a \end{cases}$