Biostatistics 682: Applied Bayesian Inference
Lecture 9: Monte Carlo Methods

Jian Kang

Department of Biostatistics
University of Michigan, Ann Arbor

Direct sampling

- Monte Carlo integration
  - Suppose $\theta \sim h(\theta)$ and we seek
$$\gamma \equiv E[f(\theta)] = \int f(\theta)h(\theta)\mathrm{d}\theta$$
  - If $\theta_1, \ldots, \theta_N \sim h(\theta)$, we have
$$\widehat{\gamma} = \frac{1}{N}\sum_{j=1}^{N} f(\theta_j).$$
    which converges to $E[f(\theta)]$ which probability 1 as $N \to \infty$, by the Strong Law of Large Numbers.
  - Variance estimate for $\widehat{\gamma}$ is given by
$$\widehat{\mathsf{Var}}(\widehat{\gamma}) = \frac{1}{N(N-1)}\sum_{j=1}^{N}\{f(\theta_j) - \widehat{\gamma}\}^2$$

Direct sampling

- Histogram estimate

$$p \equiv P[a < \theta < b \mid \mathbf{y}] = E[I_{(a,b)}(\theta) \mid \mathbf{y}],$$

where $I_{(a,b)}(\cdot)$ denotes an indicator function of set (a, b). An estimate of p is given by

$$\widehat{p} = \frac{\sum_{j=1}^{N} I_{a,b}(\theta_j)}{N} = \frac{\text{number of } \theta_j \in (a,b)}{N}.$$

What is the distribution of $N\widehat{p}$ and what is the standard error of $\widehat{p}$

- Kernel density estimate

$$\widehat{p}(\theta \mid \mathbf{y}) = \frac{1}{Nh_N} \sum_{j=1}^{N} K\left(\frac{\theta - \theta_j}{h_N}\right),$$

where K is a "kernel" density (typically a normal or rectangular distribution) and h_N is a window width satisfying $h_N \to 0$ and $Nh_N \to \infty$ as $N \to \infty$.

Example

- Let $y_i \sim N(\mu, \sigma^2)$, for $i = 1, \ldots, n$ and suppose we adopt prior
$$\pi(\mu, \sigma^2) \propto \sigma^{-2}.$$
- Estimate $E(\mu \mid y)$ and $E(\mu/\sigma \mid y)$
- The joint posterior of μ and σ^2 is given by
$$\mu \mid \sigma^2, y \sim N(\overline{y}, \sigma^2/n)$$
and
$$\sigma^2/s^2 \mid y \sim \chi^{-2}_{n-1}$$
where $s^2 = (n-1)^{-1} \sum_{i=1}^{n} (y_i - \overline{y})^2$.
- Suppose $y^* \sim \mathrm{N}(\mu, \sigma^2)$. What is the posterior predictive probability $\Pr(y^* > c \mid y)$?

Direct sampling: Grid Approximations

- Example: Binomial model with non-conjugate prior.

$$y \sim \text{Binomial}(n, \theta)$$

- The triangle prior

$$\pi(\theta) = \begin{cases} 8\theta & \text{if } 0 \leq \theta < 0.25 \\ \frac{8}{3} - \frac{8}{3}\theta & \text{if } 0.25 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- The posterior will no longer be in a convenient distributional form.
- To find the posterior, we can use grid approximations

Grid Approximation

- The unnormalized posterior = the binomial likelihood $\times$ the non-conjugate prior
  $\pi(\theta \mid y) = \pi(y \mid \theta) \times \pi(\theta)/c$ with $c = \int \pi(y \mid \theta)\pi(\theta)\mathrm{d}\theta$
- How to find the normalized posterior?
- How to sample from the posterior?
- Grid approximation
  1. Divide the region with positive unnormalized posterior into m grids, each of width k, starting from θ_0.
  $$\theta_0 + k, \theta_0 + 2k, \ldots, \theta_0 + mk$$
  2. Evaluate each grid point $(\theta_0 + ik)$ at the unnormalized posterior: $\pi(y \mid \theta_0 + ik)\pi(\theta_0 + ik)$ for $i = 1, \ldots, m$.

Find the normalized posterior

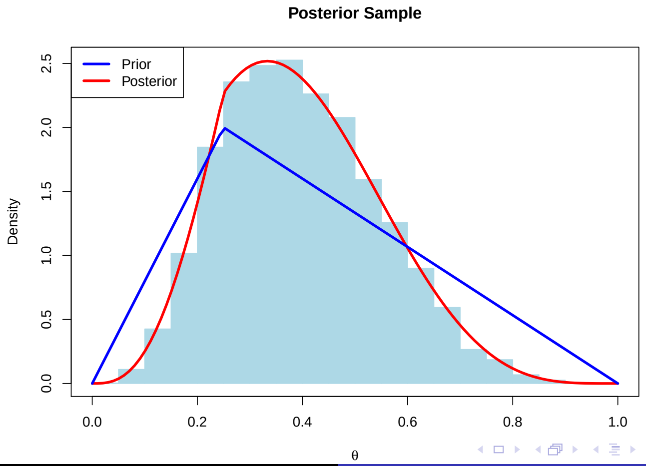
- Estimate c (area under the unnormalized posterior) as the sum of the areas of rectangles of width k and height at the unnormalized posterior ordinates:

$$c \approx \sum_{i=1}^{m} k\pi(y \mid \theta_0 + ik)\pi(\theta_0 + ik)$$

- Divide the unnormalized posterior ordinates by c to get the normalized posterior ordinates.

Sample from the posterior

- We can also sample from our normalized posterior. Note that the normalized posterior ordinates are not probabilities, but rather the heights of the density. However, the "sample()" function in R will normalize the ordinates into probabilities.



Importance sampling

- Suppose we wish to estimate the posterior mean

$$E\{h(\theta) \mid y\} = \frac{\int h(\theta)\pi(y \mid \theta)\pi(\theta)\mathrm{d}\theta}{\int \pi(y \mid \theta)\pi(\theta)\mathrm{d}\theta}$$

- Working density $g(\theta)$ (importance function)
  - Roughly estimate $\pi(\theta \mid y)$
  - Can be easily sampled
-
$$E\{h(\theta) \mid y\} = \frac{\int h(\theta)w(\theta)g(\theta)\mathrm{d}\theta}{\int w(\theta)g(\theta)\mathrm{d}\theta} \approx \frac{\sum_j h(\theta_j)w(\theta_j)}{\sum_j w(\theta_j)}$$

  where $\theta_j \sim g(\theta)$ and $w(\theta) = \pi(y \mid \theta)\pi(\theta)/g(\theta)$ is a weight function.
- How closely $g(\theta)$ resembles $\pi(\theta \mid y)$ controls how good the approximation
  - Good approximation: weights are roughly equal
  - Bad approximation: some of weights are extremely high and some are closed to zero.

Example

- Binomial model with non-conjugate prior (revisit)

$$y \sim \text{Binomial}(n, \theta)$$

- The triangle prior

$$\pi(\theta) = \begin{cases} 8\theta & \text{if } 0 \leq \theta < 0.25 \\ \frac{8}{3} - \frac{8}{3}\theta & \text{if } 0.25 \leq \theta \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Suppose we are interested in estimating posterior mean of $h(\theta) = \theta(1-\theta)$
- Importance sampling: how to choose $g(\theta)$?

Rejection Sampling

- Suppose we wish to sample from the target density: $f(x)$ (could be unnormalized density)
- We need to pick a candidate density $g(x)$ such that $f(x) \leq Mg(x)$ for all x, where M is a constant.
- Repeat the following steps until we get m accepted draws:
  1. Draw a candidate x_c from $g(x)$
  2. Calculate an acceptance probability α for x_c

$$\alpha = \frac{f(x_c)}{Mg(x_c)}$$

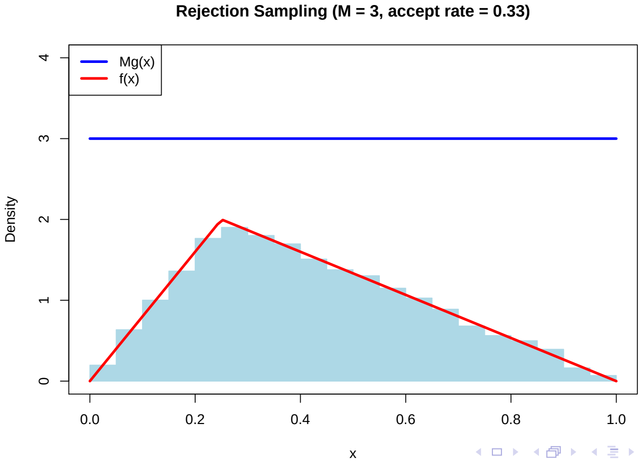
  3. Draw a value u from the Uniform (0,1) distribution
  4. Accept x_c as a draw from $f(x)$ if $\alpha \geq u$. Otherwise, reject x_c and go back to step 1

Example

- The target density: (the triangle density)

$$f(x) = 8xI[0 \le x \le 0.25] + 8/3(1-x)I[0.25 \le x \le 1].$$

- The candidate density: $g(x) = 1$ if $0 \le x \le 1$, $g(x) = 0$ otherwise.
- Set $M = 3$. Find a better M?



Weighted Bootstrap

- Similar to *sampling-importance resampling algorithm* (Rubin, 1988)
- Unnormalized targe density: $f(\theta)$
- Suppose an M appropriate for the rejection method is not readily available, but that we do have a sample $\theta_1, \ldots, \theta_N$ from some approximating density $g(\theta)$. Define

$$w_i = \frac{f(\theta_i)}{g(\theta_i)} \quad \text{and} \quad q_i = \frac{w_i}{\sum_{j=1}^N w_i}$$

- Now draw θ^* from the discrete distribution over $\{\theta_1, \ldots, \theta_N\}$ which places mass q_i at θ_i

$$\theta^* \sim \frac{f(\theta)}{\int f(\theta) \mathrm{d}\theta}$$

- Why this is true?

Example

- The target density: (the triangle density) (revisit)

$$f(x) = 8xI[0 \leq x \leq 0.25] + 8/3(1-x)I[0.25 \leq x \leq 1].$$

- The working density: $g(x) = \pi^{-1}x^{0.5}(1-x)^{0.5}$ if $0 \leq x \leq 1$, $g(x) = 0$ otherwise.