Importance Sampling

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

03/24/2021

Importance Sampling

- Recall: Direct sampling can be used to estimate

$$\int c(\theta)p(\theta)\,d\theta \approx \frac{1}{N}\sum_{1}^{N} c(\theta_i)$$

  by simulating draws from $p(\theta)$

- What if normalizing constant for $p(\theta)$ is unknown?
- Several "indirect" simulation methods, including *importance sampling*
- Basic idea of importance sampling:
  - Simulate θ from some other distribution g, but weigh each simulation by its relative likelihood under p.

Importance Sampling

- Consider the posterior $p(\theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \theta)p(\theta)$
- For a given function c,

$p(\theta \mid \vec{y})$

$$E(c(\theta) \mid \mathbf{y}) = \frac{\int c(\theta) p(\mathbf{y} \mid \theta)p(\theta) \, d\theta}{\int p(\mathbf{y} \mid \theta)p(\theta) \, d\theta}$$

- Simulate $\theta_1, \ldots, \theta_N$ iid from density $g(\theta)$
- Define the weights $w(\theta) = p(\mathbf{y} \mid \theta)p(\theta)/g(\theta)$:

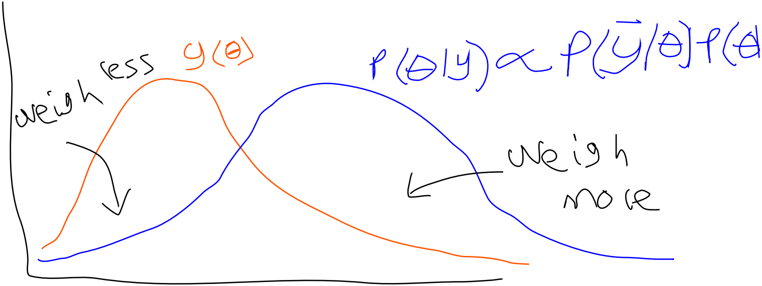
$p(\vec{y} \mid \theta)p(\theta)$

$$E(c(\theta) \mid \mathbf{y}) = \frac{\int c(\theta) w(\theta)g(\theta) \, d\theta}{\int w(\theta)g(\theta) \, d\theta}$$

$$\approx \frac{\frac{1}{N}\sum_{j=1}^{N} c(\theta_j)w(\theta_j)}{\frac{1}{N}\sum_{j=1}^{N} w(\theta_j)}$$

Comments

- $g(\theta)$ is called the *importance function*
- $w(\theta)$ weighs values according to there (relative) posterior probability, and probability of being simulated from g
  - Higher posterior probability → higher weight
  - Higher probability of being simulated from g → lower weight
- Ideally, g closely resembles $p(\theta \mid \mathbf{y})$
  - $g = p(\theta \mid \mathbf{y})$ minimizes simulation variability (equal weights)
  - If $g(\theta)$ and $p(\theta \mid y)$ are very different, many weights will be close to 0 so need large N for good approximations
  - g and $p(\theta \mid \mathbf{y})$ should (at least) have common support.


weigh less
g(θ)
p(θ|y) ∝ p(y⃗|θ)p(θ)
weigh more

Density approximation

- Plugging in $c(\theta) = \mathbb{1}_{\theta \in (a,b)}$ gives

$$P(\theta \in (a,b)) \approx \frac{\sum_{\theta_j \in (a,b)} w(\theta_j)}{\underbrace{\sum_{j=1}^{N} w(\theta_j)}}$$

- Can use this to approximate the posterior density with weighted histograms.

bin height = $\frac{1}{b-a} \cdot$

Simulation precision

- Define

$$\widehat{c(\theta)} = \frac{\frac{1}{N}\sum_{j=1}^{N} c(\theta_j)w(\theta_j)}{\frac{1}{N}\sum_{j=1}^{N} w(\theta_j)}$$

- Define $x_j = c(\theta_j)w(\theta_j)$ and $y_j = w(\theta_j)$
- We can compute the (simulation) variability in $\widehat{c(\theta)}$ as

$$\text{Var}\left(\frac{\bar{x}}{\bar{y}}\right) \approx \frac{1}{N}\left(\frac{\hat{\sigma}_X^2}{\bar{y}^2} + \frac{\bar{x}^2\hat{\sigma}_y^2}{\bar{y}^4} - \frac{2\bar{x}\hat{\sigma}_{xy}}{\bar{y}^3}\right)$$

where $\hat{\sigma}_x^2 = \frac{1}{N-1}\sum(x_j - \bar{x})^2$, $\hat{\sigma}_y^2 = \frac{1}{N-1}\sum(y_j - \bar{y})^2$, and

$$\hat{\sigma}_{xy} = \frac{1}{N-1}\sum(x_j - \bar{x})(y_j - \bar{y})$$

- *The expression in Carlin&Louis p.116 is incorrect*

Example: Normal-Gamma model

- Suppose $\mathbf{y} = y_1, \ldots, y_n \overset{iid}{\sim} \text{Normal}(0, \theta)$
- Prior $\theta \sim \text{Gamma}(3, 0.5)$[1]
- The posterior is

$$p(\theta \mid \mathbf{y}) \propto \theta^{(4-n)/2} \text{exp} \left\{ -\frac{1}{2\theta} \left(\theta^2 + \sum_{i=1}^{n} y_i^2 \right) \right\}$$

  - Not a well-known or easily integrable pdf (unlike IG prior)

---

[1] Example credit – D. Bandyopadhyay

Example: Normal-Gamma model

- Importance sample with $g = \text{Gamma}(\alpha = cs^2, \beta = c)$
  - s^2 is the sample variance of $\mathbf{y}$
  - Expected value under g is s^2
  - Higher c → less variance
- Importance weight function is $w(\theta) = \dfrac{f(\theta)\, p(\vec{y}|\theta)}{g(\theta)}$

$$w(\theta) = \theta^{\frac{4-n}{2} - cs^2 + 1} \exp\left\{ -\frac{1}{2\theta}\left(\theta^2 + \sum_{i=1}^{n} y_i^2\right) + c\theta \right\}$$

- Consider posterior simulation precision for different c.

Example: Normal-Gamma model

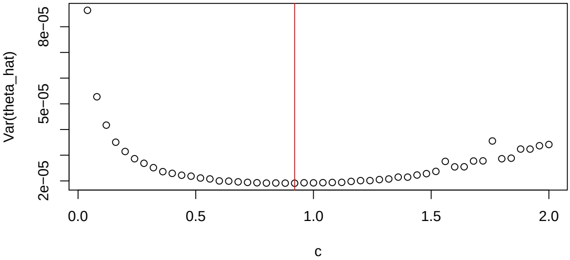
- For different values of c:
- Simulate $N = 100000$ draws from $g(\theta)$
- Compute $w(\theta)$ from each draw
- For the estimated expected value

$$\hat{\theta} = \frac{\frac{1}{N}\sum_{j=1}^{N} \theta_j w(\theta_j)}{\frac{1}{N}\sum_{j=1}^{N} w(\theta_j)}$$

compute the simulation uncertainty $\text{Var}(\hat{\theta})$ using formula.

Example: Normal-Gamma model

- Simulation variance of $\hat{\theta}$ for c for given data $y_1, \ldots, y_{20}$



http://www.ericfrazerlock.com/More_on_Importance_Sampling_Rcode1.r

- Minimized at $c = 0.92$

Example: Normal-Gamma model

- Using draws under $c = 0.92$,

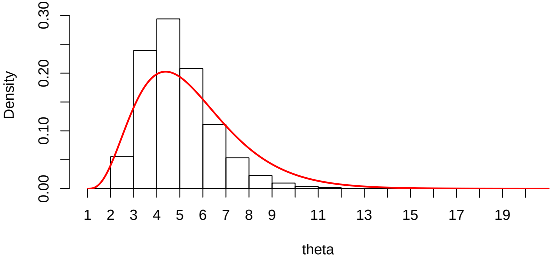
$$\hat{\theta} = 4.890$$

- We can also compute posterior probabilities, e.g.,

$$P(\theta > 5 \mid \mathbf{y}) = \frac{\sum_{\theta_j > 5} w(\theta_j)}{\sum_{j=1}^{N} w(\theta_j)} = 0.410$$

Example: Normal-Gamma model

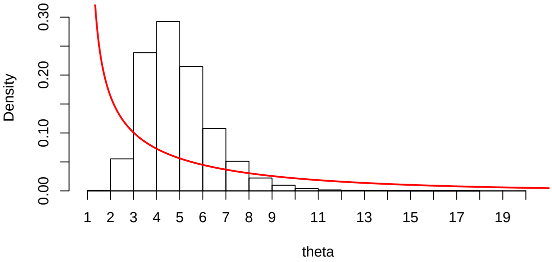
- Weighted histogram approximating posterior $p(\theta \mid \mathbf{y})$, for $c = 0.92$. $g(\theta)$ is shown in red



http://www.ericfrazerlock.com/More_on_Importance_Sampling_Rcode1.r

Example: Normal-Gamma model

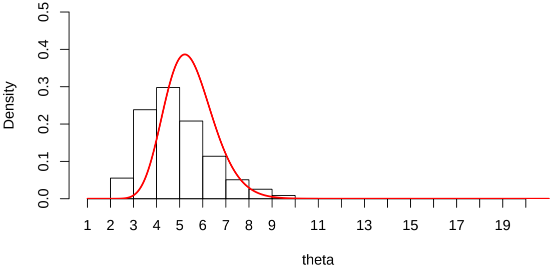
- Weighted histogram approximating posterior $p(\theta \mid \mathbf{y})$, for $c = 0.10$. $g(\theta)$ is shown in red



http://www.ericfrazerlock.com/Importance_Sampling_Rcode1.r

Example: Normal-Gamma model

- Weighted histogram approximating posterior $p(\theta \mid \mathbf{y})$, for $c = 4$. $g(\theta)$ is shown in red



http://www.ericfrazerlock.com/Importance_Sampling_Rcode1.r