Asymptotic Posterior Approximation

PUBH 8442: Bayes Decision Theory and Data Analysis

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

03/17/2021

Example: Basketball shooting

- Consider the shooting percentage for a basketball team over n games: $\mathbf{y} = (y_1, \ldots, y_n)$
- Model $y_i \overset{iid}{\sim} \text{Beta}(\theta, 2)$

$$p(y_i \mid \theta) = \theta(1+\theta)y_i^{\theta-1}(1-y_i)$$

For $\theta > 0$.

- Use a Gamma(a, b) prior for θ
- Then,

$$p(\theta \mid \mathbf{y}) \propto \theta^{n+a-1}(\theta+1)^n e^{-b\theta}\left(\prod_{i=1}^n y_i\right)^\theta$$

$$\propto p(\theta)\, p(\vec{y} \mid \theta) = p(\theta) \prod_{i=1}^{n} p(y_i \mid \theta)$$

Example: Basketball shooting

- $p(\theta \mid \mathbf{y})$ does not correspond to a well-known pdf.
- The integral

$$\int \theta^{n+a-1}(\theta+1)^n \exp\left\{-\theta(b+\sum_{i=1}^{n}\log(1/y_i)\right\} d\theta \qquad (1)$$

does not have a simple closed form.

- Difficult to find posterior expectations, probabilities, posterior posterior predictive, etc. without advanced computing
- Finding mode $\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ p(\theta \mid \mathbf{y})$ is simpler.
  - Does not depend on normalizing constant (1)

Example: Basketball shooting

- Assume $a = b = 1$. The mode of $p(\theta \mid \mathbf{y})$ is given by the quadratic formula

$$\frac{d}{d\theta} \log p(\theta \mid \mathbf{y}) = 0$$
$$\to A\theta^2 + B\theta + C = 0$$

where

- $A = \sum_{i=1}^{n} \log y_i - 1$
- $B = 2n - 1 + \sum_{i=1}^{n} \log y_i$
- $C = n$

$$\log p(\theta|\vec{y}) = n\log\theta + n\log(\theta+1) - \theta + \theta\left(\sum \log y_i\right) + C$$

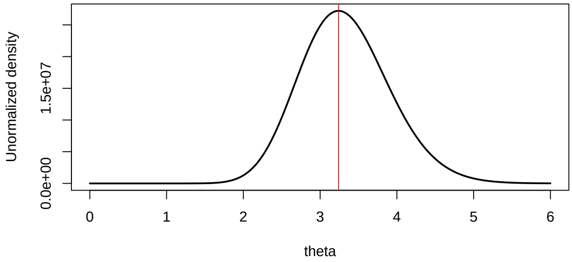
$$\frac{d}{d\theta} = \frac{n}{\theta} + \frac{n}{\theta+1} - 1 + \sum \log y_i = 0$$

$$\longrightarrow n(\theta+1) + n\theta + \theta(\theta+1)\left(\sum \log y_i - 1\right) = 0$$

$$\theta^2\left(\sum \log y_i - 1\right) + \theta\left(2n - 1 + \sum \log y_i\right) + n = 0$$

Example: Basketball shooting

- Assume for $n = 20$ games $\sum_{i=1}^{20} \log y_i = -9.89$
- Then, the posterior mode is $\hat{\theta} = 3.24$
- Plot of "unnormalized" posterior density:



http://www.ericfrazerlock.com/Asymptotic_Posterior_Approximations_Rcode1.r

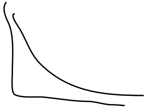
Bayesian Central Limit Theorem

- *Theorem*: Assume $y_1, \ldots, y_n$ are iid with pdf $p(\cdot \mid \theta)$ and θ has prior p_θ. Then, under general conditions,

$$p(\theta \mid \mathbf{y}) \approx \text{Normal}(\hat{\theta}, (I(\mathbf{y}))^{-1}) \text{ as } n \to \infty$$

where $\hat{\theta}$ is the posterior mode, and I is the Fisher information matrix for $p(\theta \mid \mathbf{y})$:

$$I_{ij}(\mathbf{y}) = -\left[\frac{d^2}{d\theta_i d\theta_j} \log(p(\mathbf{y} \mid \theta) p(\theta))\right]_{\theta = \hat{\theta}}.$$



- Sometimes called the *Bayesian Central Limit Theorem*
- Conditions are that the mode $\hat{\theta}$ exists, $p(\mathbf{y} \mid \theta)$ and $p(\theta)$ are positive and twice differentiable at $\hat{\theta}$, and assumptions to assure $\hat{\theta}$ is not a "boundary point".

Let $l(\theta) = \log p(\theta|y)$

- The Bayesian CLT is given by a second order Taylor expansion about $\hat{\theta}$.

$$l(\theta) \approx \underbrace{l(\hat{\theta})}_{L} + (\theta - \hat{\theta})\underbrace{l'(\hat{\theta})}_{0} + \frac{1}{2}(\theta - \hat{\theta})^2 l''(\hat{\theta})$$

$$= C - \frac{1}{2} I_{\theta|y}(y)(\theta - \hat{\theta})^2$$

$$p(\theta|y) = \exp(l(\theta)) \approx C \cdot \exp\left(-\frac{1}{2}\frac{(\theta - \hat{\theta})^2}{I_{\theta|y}(y)^{-1}}\right)$$

Alternative versions

- Alternatively we could simply use the mean (μ) and variance V of $p(\theta \mid \mathbf{y})$ in the approximation:

$$p(\theta \mid \mathbf{y}) \approx \text{Normal}(\mu, V).$$

  - But if we know μ and V, there is likely no need to approximate $p(\theta \mid \mathbf{y})$.

- Alternatively, we could ignore the prior:

$$p(\theta \mid \mathbf{y}) \approx \text{Normal}(\tilde{\theta}, (\hat{I}(\mathbf{y}))^{-1})$$

where $\tilde{\theta}$ is the MLE, and $\hat{I}$ is observed Fisher information

$$\hat{I}_{ij}(\mathbf{y}) = -\left[\frac{d^2}{d\theta_i d\theta_j} \log(p(\mathbf{y} \mid \theta))\right]_{\theta=\tilde{\theta}}.$$

  - Accuracy for moderate n depends on "flatness" of p_θ.

Example: Basketball shooting

- For $a = b = 1$,

$$p(\theta \mid \mathbf{y}) \approx \text{Normal}\left(\hat{\theta}, \frac{\hat{\theta}^2(1+\hat{\theta})^2}{n((1+\hat{\theta})^2+\hat{\theta}^2)}\right)$$

where $\hat{\theta}$ solves the quadratic given earlier:

$$\hat{\theta} = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

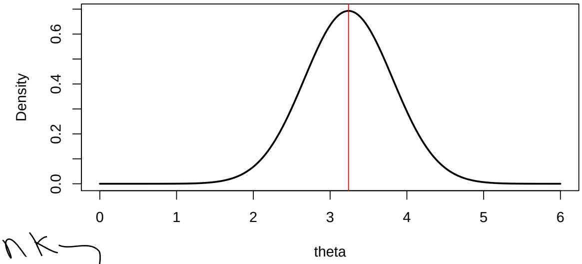
$$\frac{d^2}{d\theta^2} l(\theta) = -\frac{n}{\theta^2} - \frac{n}{(\theta+1)^2}$$

$$\hat{I}(\theta) = \left(-\frac{d^2}{d\theta^2} l(\theta)\right)\Big|_{\theta=\hat{\theta}}$$

$$\frac{1}{\hat{I}(\theta)}$$

Example: Basketball shooting

- Assume for $n = 20$ games $\sum_{i=1}^{20} \log y_i = -9.89$
- Then, the posterior approximation is Normal$(3.24, 0.33)$
- Plot of approximated posterior density:



fix link

http://www.tc.umn.edu/~elock/Asymptotic_Posterior_Approximations_Rcode1.r

Beta-binomial approximation

- Let $y \sim \text{Binomial}(n, \theta)$ and $\theta \sim \text{Beta}(a, b)$.

- Then,

$$p(\theta \mid \mathbf{y}) \approx \text{Normal}\left(\hat{\theta}, \left[\frac{a+y-1}{\hat{\theta}^2} + \frac{b+n-y-1}{(1-\hat{\theta})^2}\right]^{-1}\right)$$

where $\hat{\theta} = \frac{a+y-1}{a+b+n-2}$

$$\theta \mid y \sim \text{Beta}(a+y,\ b+n-y)$$

$$l(\theta) = (a+y-1)\log\theta + (b+n-y-1)\log(1-\theta) + C$$

$$\frac{d}{d\theta} = \frac{a+y-1}{\theta} - \frac{b+n-y-1}{1-\theta} = 0$$

$$\Rightarrow \hat{\theta} = \frac{a+y-1}{a+b+n-2}$$

$$\frac{d^2}{d\theta^2} = -\frac{a+y-1}{\theta^2} - \frac{b+n-y-1}{(1-\theta)^2}$$

- If $a = b = 1$ then

$$p(\theta \mid \mathbf{y}) \approx \text{Normal}\left(\hat{\theta}, \frac{\hat{\theta}(1-\hat{\theta})}{n}\right)$$

where $\hat{\theta} = y/n$

$$\left(\frac{y}{\hat{\theta}^2} + \frac{n-y}{(1-\hat{\theta})^2}\right)^{-1} = \dots =$$

- This corresponds to the standard frequentist approximation.