ST 740: Multiparameter Inference

Alyson Wilson

Department of Statistics
North Carolina State University

September 16, 2013

How Many Samples?

Chapter 4 of Hoff covers the “Monte Carlo method,” where we use a random sample from a distribution to approximate moments, cdf, quantiles, density function. Quoting Hoff,

> *Just about any aspect of a posterior distribution we may be interested in can be approximated arbitrarily exactly with a large enough Monte Carlo sample.*

The question is how many samples is “large enough”?

How Many Samples?

We usually pick the number of Monte Carlo samples using a simple heuristic.

- Suppose that we are interested in using a Monte Carlo sample of size S to derive an estimate of a posterior mean $E[\theta \mid \mathbf{y}]$.
- We know from the Central Limit Theorem that $\bar{\theta}$ is approximately normally distributed with mean $E[\theta \mid \mathbf{y}]$ and standard deviation $\sqrt{Var[\theta \mid \mathbf{y}]/S}$.
- Approximate $Var[\theta \mid \mathbf{y}]$ using the sample variance ($\hat{\sigma}^2$) calculated from our Monte Carlo sample.
- An approximate 95% Monte Carlo confidence interval for the posterior mean is $\bar{\theta} \pm 2\sqrt{\hat{\sigma}^2/S}$.
- Pick S so that your Monte Carlo estimate is "close enough."

Roadmap

- Most realistic problems require models with more than one parameter
- When working with multi-parameter models, we almost always summarize our results in terms of the marginal posterior distributions for the parameters
- For the two-parameter normal distribution, this means $\pi(\mu \mid \mathbf{y})$ and $\pi(\sigma^2 \mid \mathbf{y})$
- We might also want to know about the posterior distribution of some function of the parameters: say, $\pi(\frac{\sigma}{\mu} \mid \mathbf{y})$.
- There are two approaches:
  1. Figure the marginals out analytically
  2. Figure out a way to simulate either from the marginals directly or from the posterior

How does simulating help us?

If we have a random sample from the posterior distribution, we also have a sample from the marginals and from the posterior distribution for any function of the parameters.

Suppose we have a sample $(\mu^{(1)}, (\sigma^2)^{(1)}), \ldots, (\mu^{(m)}, (\sigma^2)^{(m)})$ from the joint posterior distribution for μ and σ^2. Then

- $\mu^{(1)}, \ldots, \mu^{(m)}$ is a random sample from the marginal posterior distribution of μ
- $(\sigma^2)^{(1)}, \ldots, (\sigma^2)^{(m)}$ is a random sample from the marginal posterior distribution of σ^2
- $\frac{\sqrt{(\sigma^2)^{(1)}}}{\mu^{(1)}}, \ldots, \frac{\sqrt{(\sigma^2)^{(m)}}}{\mu^{(m)}}$ is a random sample from the posterior distribution of $\frac{\sigma}{\mu}$.

Marginal Distribution

- Consider a model with two parameters (θ_1, θ_2) (e.g., a normal distribution with unknown mean and variance)
- Suppose that we are interested in θ_1
- The marginal posterior distribution of interest is $\pi(\theta_1|y)$
- This can be obtained directly from the *joint posterior density*

$$\pi(\theta_1, \theta_2 \mid \mathbf{y}) \propto \pi(\theta_1, \theta_2) f(\mathbf{y} \mid \theta_1, \theta_2)$$

by integrating with respect to θ_2:

$$\pi(\theta_1 \mid \mathbf{y}) = \int \pi(\theta_1, \theta_2 \mid \mathbf{y}) d\theta_2$$

or by factoring the posterior

$$\pi(\theta_1 \mid \mathbf{y}) = \int \pi(\theta_1 \mid \theta_2, \mathbf{y}) \pi(\theta_2 \mid \mathbf{y}) d\theta_2$$

Calculation Considerations

$$
\begin{aligned}
\pi(\theta_1 \mid \mathbf{y}) &= \int \pi(\theta_1, \theta_2 \mid \mathbf{y}) d\theta_2 \\
&= \int \pi(\theta_1 \mid \theta_2, \mathbf{y}) \pi(\theta_2 \mid \mathbf{y}) d\theta_2
\end{aligned}
$$

Here are four ways to calculate this:

- Do the integration.
- Sample from the joint distribution (more on that later)

Computational Considerations

- A two-step simulation approach:
  1. Marginal simulation step: Draw value $\theta_2^{(k)}$ of θ_2 from $\pi(\theta_2 \mid \mathbf{y})$ for $k = 1, 2, \ldots$.
  2. Conditional simulation step: For each $\theta_2^{(k)}$, draw a value of $\theta_1^{(k)}$ from the conditional density $\pi(\theta_1 \mid \theta_2^{(k)}, \mathbf{y})$

This gives you a sample $\theta_1^{(k)}$ from $\pi(\theta_1 \mid \mathbf{y})$. You can use this sample to calculate a kernel density estimate of the density or statistics (mean, variance, quantiles, etc.)

This will be an effective approach when marginal and conditional are of standard form. We'll see more sophisticated simulation approaches later.

Computational Considerations

- (The Rao-Blackwellized estimator.) Suppose you have a sample $\theta_2^{(1)}, \ldots, \theta_2^{(m)}$ from $\pi(\theta_2 \mid \mathbf{y})$. Use

$$f(\theta_1 \mid \mathbf{y}) \approx \frac{1}{m} \sum_{k=1}^{m} f(\theta_1 \mid \theta_2^{(k)}, \mathbf{y})$$

Example: Normal Model

Assume that we have $Y_i \mid \mu, \sigma^2 \sim \text{Normal}(\mu, \sigma^2)$. Both μ and σ^2 are unknown.

The non-informative prior for (μ, σ^2) assuming prior independence is

$$\pi(\mu, \sigma^2) \propto 1 \times \frac{1}{\sigma^2}$$

The joint posterior distribution is

$$\begin{aligned} \pi(\mu, \sigma^2 \mid \mathbf{y}) &\propto \pi(\mu, \sigma^2) f(\mathbf{y} \mid \mu, \sigma^2) \\ &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \end{aligned}$$

Example: Normal Model

Note that

$$\sum_{i=1}^{n}(y_i-\mu)^2 = \sum_i (y_i^2 - 2\mu y_i + \mu^2)$$
$$= \sum_i y_i^2 - 2\mu n\bar{y} + n\mu^2$$
$$= \sum_i (y_i-\bar{y})^2 + n(\bar{y}-\mu)^2$$

by adding and subtracting $2n\bar{y}^2$.

Let

$$s^2 = \frac{1}{n-1}\sum_i (y_i-\bar{y})^2$$

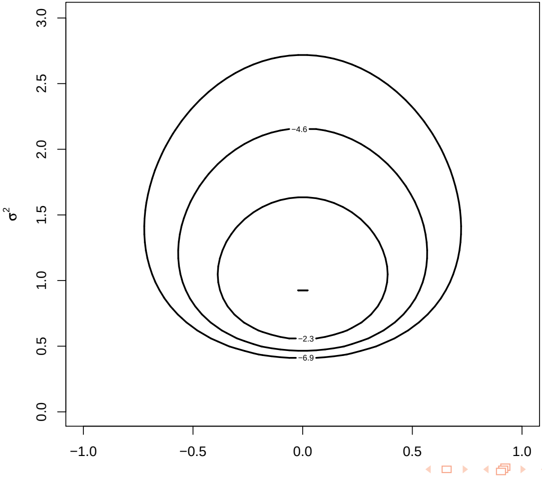
Then we can write posterior for (μ,σ^2) as

$$\pi(\mu,\sigma^2 \,|\, \mathbf{y}) \propto \sigma^{-n-2}\exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2])$$

Contour Plot of Posterior Distribution

$n = 30, \bar{y} = 0, s^2 = 1$

```
mycontour()
```



An Aside on Noninformative Priors for Multiparameter Models

Jeffreys Rule does extend to multiparameter models.

$$\pi(\theta) \propto |I(\theta)|^{1/2}$$

where

$$I_{ij}(\theta) = -E\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(\mathbf{y} \mid \theta)\right]$$

An Aside on Noninformative Priors for Multiparameter Models

This is seldom used in practice.

1. In high dimensions, this is cumbersome to calculate.
2. For situations where you are considering scale and location parameters simultaneously, it produces inappropriate results (see Box and Tiao 1973).

Instead, the general practice is to obtain the Jeffreys' prior for each parameter individually and form the joint prior distribution as the product of the marginal prior distributions.

Example: Normal Model

Conditional posterior $\pi(\mu \mid \sigma^2, \mathbf{y})$

Conditional on σ^2:

$$\pi(\mu \mid \sigma^2, \mathbf{y}) = \text{Normal}(\bar{y}, \sigma^2/n)$$

We can see this by noting that, viewed as a function of μ only:

$$\begin{aligned}
\pi(\mu \mid \sigma^2, \mathbf{y}) &\propto \sigma^{-n-2} \exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]) \\
&\propto \exp(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2)
\end{aligned}$$

that we recognize as the kernel of a Normal$(\bar{y}, \sigma^2/n)$.

Example: Normal Model

Marginal posterior $\pi(\sigma^2 \mid \mathbf{y})$

To calculate $\pi(\sigma^2 \mid \mathbf{y})$ we need to integrate $\pi(\mu, \sigma^2 \mid \mathbf{y})$ over μ:

$$\begin{aligned}
\pi(\sigma^2 \mid \mathbf{y}) &\propto \int \sigma^{-n-2} \exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]) d\mu \\
&\propto \sigma^{-n-2} \exp(-\frac{(n-1)s^2}{2\sigma^2}) \int \exp(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2) d\mu \\
&\propto \sigma^{-n-2} \exp(-\frac{(n-1)s^2}{2\sigma^2}) \sqrt{2\pi\sigma^2/n}
\end{aligned}$$

Then

$$\pi(\sigma^2 \mid \mathbf{y}) \propto (\sigma^2)^{-(n+1)/2} \exp(-\frac{(n-1)s^2}{2\sigma^2})$$

which is proportional to a *scaled-inverse* χ^2 distribution with degrees of freedom $(n-1)$ and scale s^2. (More, simply, it is an inverse gamma distribution.)

Two-Parameter Normal Model

Conditional posterior $\pi(\sigma^2 \mid \mu, \mathbf{y})$

Conditional on μ:

$$\pi(\sigma^2 \mid \mu, \mathbf{y}) = \text{InverseGamma}(\frac{n}{2}, \frac{(n-1)s^2 + n(\bar{y} - \mu)^2}{2})$$

We can see this by noting that, viewed as a function of σ^2 only:

$$\pi(\sigma^2 \mid \mu, \mathbf{y}) \propto (\frac{1}{\sigma^2})^{\frac{n}{2}+1} \exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2])$$

which we recognize as the kernel of an inverse gamma distribution.

Two-Parameter Normal Model

Marginal posterior $\pi(\mu \mid \mathbf{y})$

$$
\begin{aligned}
\pi(\mu \mid \mathbf{y}) &= \int \pi(\mu, \sigma^2 \mid \mathbf{y}) d\sigma^2 \\
&\propto \int (\frac{1}{2\sigma^2})^{n/2+1} \exp(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]) d\sigma^2
\end{aligned}
$$

Use the transformation

$$z = \frac{A}{2\sigma^2}$$

where $A = (n-1)s^2 + n(\bar{y}-\mu)^2$. Then

$$\frac{d\sigma^2}{dz} = -\frac{A}{2z^2}$$

Two-Parameter Normal Model

Marginal posterior $\pi(\mu \,|\, \mathbf{y})$

$$\begin{aligned}
\pi(\mu \,|\, \mathbf{y}) &\propto \int_0^\infty \left(\frac{z}{A}\right)^{\frac{n}{2}+1} \frac{A}{z^2} \exp(-z) dz \\
&\propto A^{-n/2} \int z^{\frac{n}{2}-1} \exp(-z) dz
\end{aligned}$$

$$\pi(\mu \,|\, \mathbf{y}) \propto A^{-n/2} \int z^{\frac{n}{2}-1} \exp\left(-z\right) dz$$

The integrand is an unnormalized Gamma(n/2, 1), so the integral is constant w.r.t. μ.

Two-Parameter Normal Model

Marginal posterior $\pi(\mu \mid \mathbf{y})$

Recall that $A = (n-1)s^2 + n(\bar{y} - \mu)^2$. Then

$$\begin{aligned}
\pi(\mu \mid \mathbf{y}) &\propto A^{-n/2} \\
&\propto [(n-1)s^2 + n(\bar{y} - \mu)^2]^{-n/2} \\
&\propto [1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}]^{-n/2}
\end{aligned}$$

which is the kernel of a t distribution with $n-1$ degrees of freedom, centered at $\bar{y}$ and with scale parameter s^2/n.

Random Sample from Posterior Distribution

We can generate random samples directly from the marginal posteior distributions.

You cannot compute the posterior distribution of functions of the parameters using draws from the marginal posterior distributions: you've lost the covariance structure. You need draws from the joint posterior distribution.

Recall that

$$\begin{aligned}\pi(\theta_1, \theta_2 \,|\, \mathbf{y}) &= \pi(\theta_1 \,|\, \theta_2, \mathbf{y})\pi(\theta_2 \,|\, \mathbf{y}) \\ &= \pi(\theta_2 \,|\, \theta_1, \mathbf{y})\pi(\theta_1, \mathbf{y})\end{aligned}$$

Random Sample from Posterior Distribution

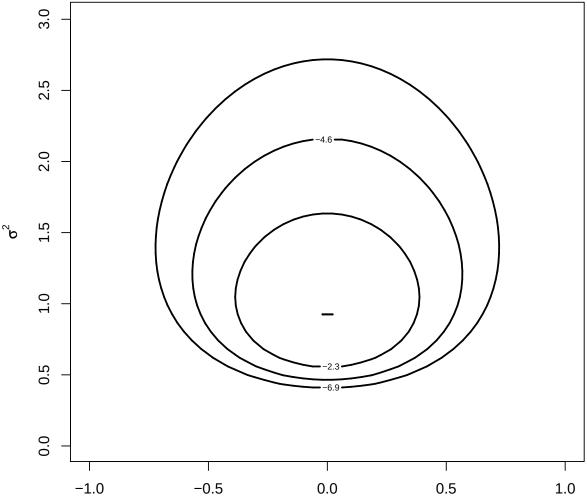
We can get a random sample from the posterior distribution in two ways.

1. - Sample $(\sigma^2)^{(i)}$ from $\left[\sigma^2 \,|\, \mathbf{y}\right] \sim \text{InverseGamma}(\frac{n-1}{2}, \frac{(n-1)s^2}{2})$.
   - Then sample $\mu^{(i)}$ from $\left[\mu \,|\, \sigma^2, \mathbf{y}\right] \sim \text{Normal}(\bar{y}, (\sigma^2)^{(i)}/n)$.
2. - Sample $\mu^{(i)}$ from $[\mu \,|\, \mathbf{y}] \sim \text{t}(n-1, \bar{y}, \frac{s^2}{n})$.
   - Then sample $(\sigma^2)^{(i)}$ from $\left[\sigma^2 \,|\, \mu, \mathbf{y}\right] \sim \text{InverseGamma}(\frac{n}{2}, \frac{(n-1)s^2 + n(\bar{y}-\mu^{(i)})^2}{2})$.

Notice the relationship of this to our strategy for sampling from the marginal distributions.

Contour Plot of Posterior Distribution

$n = 30, \bar{y} = 0, s^2 = 1$



Random Sample from Posterior Distribution

Marginal Distributions