ST 740: Multiparameter Inference

Alyson Wilson

Department of Statistics
North Carolina State University

September 25, 2013

Less Nice Example

Suppose that our sampling distribution is Gamma(α, β) and we choose independent marginal prior distributions $\pi(\alpha) \sim$ Gamma$(2, 1)$ and $\pi(\beta) \sim$ Gamma$(5, 1)$.

$$
\begin{aligned}
\pi(\alpha, \beta \,|\, \mathbf{y}) &\propto \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \exp(-\beta \sum y_i)(\prod y_i)^{\alpha-1} \exp(-(\alpha+\beta))\alpha\beta^4 \\
&\propto \frac{\beta^{n\alpha+4}}{\Gamma(\alpha)^n} \exp(-\beta(\sum y_i + 1))\alpha \exp(-\alpha)(\prod y_i)^{\alpha-1}
\end{aligned}
$$

We want to make posterior inferences about α, β, and the predictive distribution for the next observation.

Predictive Distribution

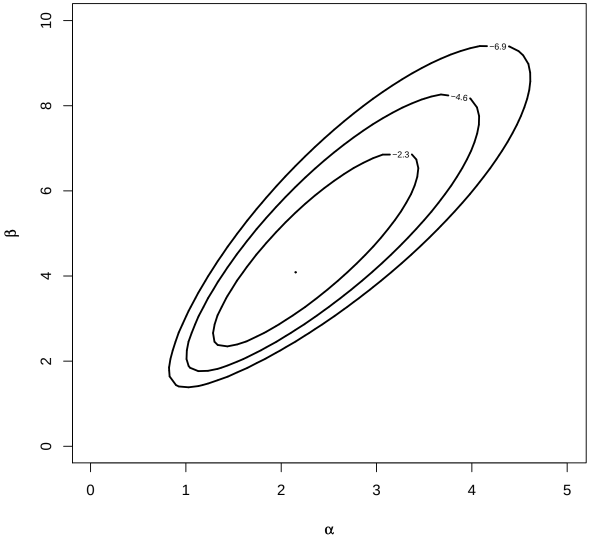
The predictive distribution is the distribution of the next observation conditional on the data observed so far. The uncertainty about the parameters is integrated out.

$$f(y_{n+1} \mid \mathbf{y}) = \int f(y_{n+1} \mid \alpha, \beta)\pi(\alpha, \beta \mid \mathbf{y}) d\alpha d\beta$$

where

$$f(y_{n+1} \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \exp(-\beta y_{n+1})(y_{n+1})^{\alpha-1}$$

Contour Plot of Posterior Distribution



Joint, Marginal, and Conditional Distributions

Posterior?

$$\pi(\alpha, \beta \mid \mathbf{y}) \propto \frac{\beta^{n\alpha+4}}{\Gamma(\alpha)^n} \exp(-\beta(\sum y_i + 1))\alpha \exp(-\alpha)(\prod y_i)^{\alpha-1}$$

Conditionals?

$\pi(\alpha \mid \beta, \mathbf{y})$ hard.

$\beta \mid \alpha, \mathbf{y} \sim \text{Gamma}(n\alpha + 5, \sum y_i + 1)$

Marginals?

$\pi(\beta \mid \mathbf{y})$ hard.

$\pi(\alpha \mid \mathbf{y}) \propto \frac{\alpha \exp(-\alpha)}{\Gamma(\alpha)^n} \frac{\Gamma(n\alpha+5)}{(\sum y_i+1)^{n\alpha+5}} (\prod y_i)^{\alpha-1}$

Now what?

In order to make inference about α and β, we need either an analytic form for the marginal posterior distributions, a mechanism for generating random samples from the marginal posterior distributions, or a mechanism for generating random samples from the joint posterior distribution.

In order to make inference about the predictive distribution, we need either an analytic form for the predictive distribution or a mechanism for generating random samples from the joint posterior distribution.

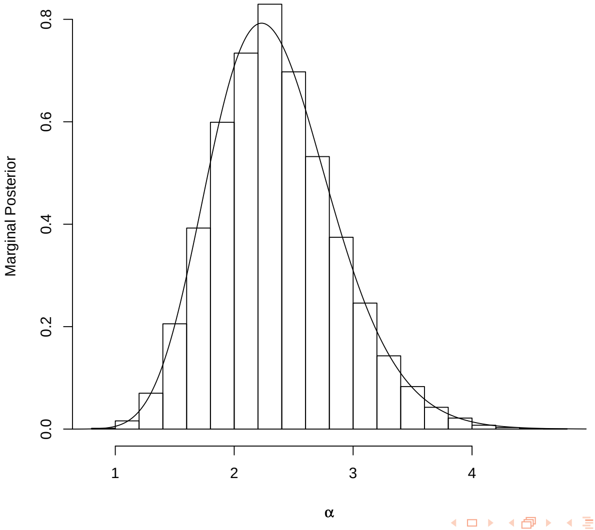
If we can figure out how to draw a sample from $\pi(\alpha \mid \mathbf{y})$, we have an easy expression for $\pi(\beta \mid \alpha, \mathbf{y})$, so we will be able to get samples from the joint posterior distribution and the marginal posterior distribution for β. This will let us calculate everything we're interested in.

Method 1: Brute Force

Illustrate with $\pi(\alpha \mid \mathbf{y}) \propto \frac{\alpha \exp(-\alpha)}{\Gamma(\alpha)^n} \frac{\Gamma(n\alpha+5)}{(\sum y_i+1)^{n\alpha+5}} (\prod y_i)^{\alpha-1}$

- Choose a grid of value of α that covers the posterior density
- Compute the value of the posterior distribution at each point on the grid
- Normalize by dividing each product by the sum of the products. In this step, we are approximating the posterior density by a discrete probability distribution on the grid.
- Take a random sample with replacement from the discrete distribution.

Brute Force



Generalization to Several Parameters

Straightforward: grid things in multiple dimensions. (Implemented in R as `simcontour()` in 2-D).

Suppose that you want to have 1000 grid points for each dimension. Does not scale well—you quickly have to keep track of far too many probabilities and points.

A Little Useful Theory

Theorem: Suppose that $Y_1, \ldots Y_n \sim f(y_i \mid \theta)$, with the Y_i conditionally independent given θ. Then $f(\mathbf{y} \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta)$. Suppose the prior $\pi(\theta)$ and $f(\mathbf{y} \mid \theta)$ are positive and twice differentiable near $\hat{\theta}^{\pi}$, the posterior mode of θ, assumed to exist. Then under suitable regularity conditions, the posterior distribution $\pi(\theta \mid \mathbf{y})$ for large n can be approximated by a normal distribution having mean equal to the posterior mode, and covariance matrix equal to negative the inverse Hessian (second derivative matrix) of the log posterior evaluated at the mode.

Note: In R, the `laplace` function will compute the posterior mode and negative inverse Hessian of the log posterior evaluated at the mode.

Method 2: Posterior Approximations

There are several other forms of the normal approximation, see Berger (1985). These are often referred to as *modal approximations* to the posterior.

There is another set of more accurate posterior estimates that do not require significantly more effort in terms of higher order derivatives or complicated transformations called *Laplace approximations*. See Tierney and Kadane (1986).

Some limitations:

- Posterior must be unimodal, or nearly so
- Accuracy depends on parameterization, and it is hard to figure out the best one
- Result for "large n"

Method 3: Rejection Sampling

We want a random sample from some distribution $\pi(\theta \mid \mathbf{y})$. We may not know this distribution's normalizing constant.

Step 1: Choose another probability density $p(\theta)$ such that

- It is easy to simulate draws from p
- The density of p resembles $\pi(\theta \mid \mathbf{y})$ in terms of location and spread
- For all θ and a constant c, $\pi(\theta \mid \mathbf{y}) \leq cp(\theta)$.

Rejection Sampling

If you can find such a $p(\theta)$, then the algorithm to follow is

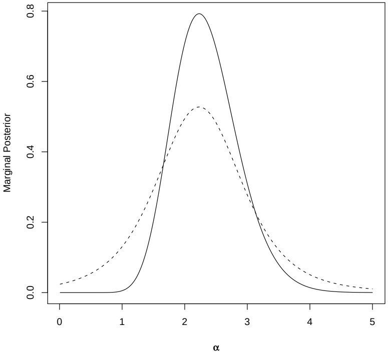
- Simulate independently $\theta^{(i)}$ from $p(\theta)$
- Simulate independently $u^{(i)}$ from a Uniform$(0, 1)$ distribution
- If $u^{(i)} \leq \frac{\pi(\theta^{(i)} \mid \mathbf{y})}{cp(\theta^{(i)})}$ then accept $\theta^{(i)}$ as a draw from the density $\pi(\theta \mid \mathbf{y})$. Otherwise reject $\theta^{(i)}$ (and throw it away).
- Keep going until you get a large enough sample of "accepted" θs

Notes:

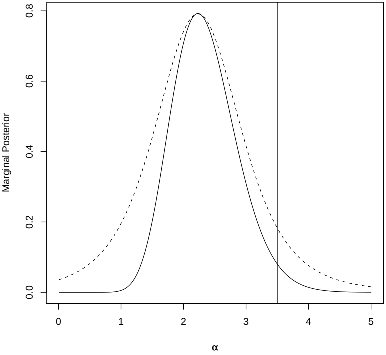
- For a proof that this works, see Devroye (1988) or Ripley (1987).
- The better the "envelope" $(cp(\theta))$ the more efficient the sampling.
- There are a variety of adaptive versions of this algorithm where you revise $p(\theta)$ as you go along, e.g., Gilks (1990).

Rejection Sampling

$\pi(\alpha \,|\, \mathbf{y})$ and t-distribution(df $= 4$, $\mu = 2.2304, \sigma^2 = 0.2555533$)



$\pi(\alpha \,|\, \mathbf{y})$ and 1.5*t-distribution(df $=$ 4, $\mu = 2.2304, \sigma^2 = 0.2555533$)



Method 4: SIR Sampling/Weighted Bootstrap

We want a random sample from some distribution $\pi(\theta \mid \mathbf{y})$. We may not know this distribution's normalizing constant.

Step 1: Choose another probability density $p(\theta)$ such that

- It is easy to simulate draws from p
- The density of p resembles $\pi(\theta \mid \mathbf{y})$ in terms of location and spread

But this time, we're having trouble figuring out what c should be.

SIR Sampling

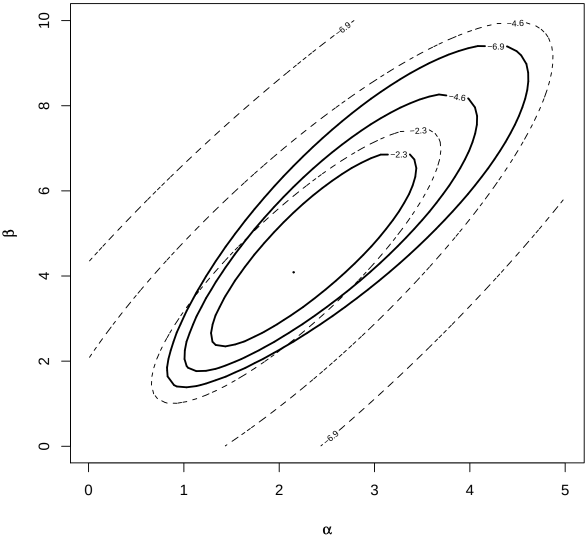
The algorithm to follow is

- Simulate independently $\theta^{(i)}$ from $p(\theta)$, $i = 1, \ldots, m$
- Calculate $w_i = \frac{\pi(\theta^{(i)} \mid \mathbf{y})}{p(\theta^{(i)})}$
- Normalize the weights, $q_i = \frac{w_i}{\sum_{i=1}^m w_i}$
- Resample (with replacement) the $\theta^{(i)}$ with weights q_i

Notes:

1. You want to make sure that the proposal distribution $p(\theta)$ has heavier tails than the posterior distribution $\pi(\theta \mid \mathbf{y})$ or you will have trouble getting samples from the tails.
2. t distributions are again a good choice here.
3. Check your weights and make sure that you don't have one or two large ones and the rest tiny. This depends on how well $p(\theta)$ approximates $\pi(\theta \mid \mathbf{y})$.

SIR Sampling



More Details

Works the same way for the multiparameter posterior distributions, but the proposal densities are usually multivariate t distributions.