ST 740: Multiparameter Inference

Alyson Wilson

Department of Statistics
North Carolina State University

September 30, 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$(5, 1)$.

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

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

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

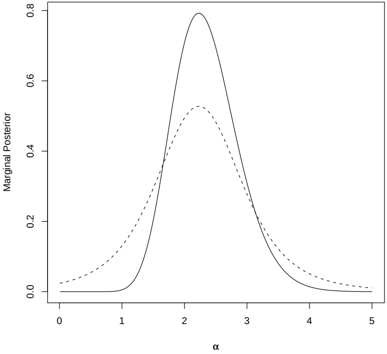
- Simulate independently $\theta^{(i)}$ from $p(\theta)$
- Simulate independently $u^{(i)}$ from a Uniform$(0, 1)$ distribution
- If $u^{(i)} \leq \frac{\pi(\theta^{(i)} \mid \mathbf{y})}{cp(\theta^{(i)})}$ then accept $\theta^{(i)}$ as a draw from the density $\pi(\theta \mid \mathbf{y})$. Otherwise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

Notes:

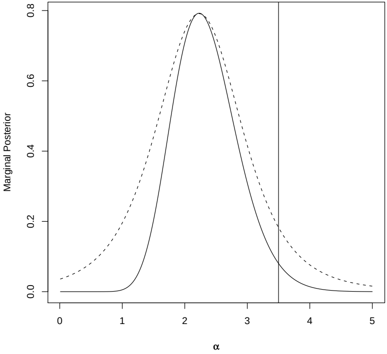
- For a proof that this works, see Devroye (1988) or Ripley (1987).
- The better the "envelope" ($cp(\theta)$) the more efficient the sampling.
- There are a variety of adaptive versions of this algorithm where you revise $p(\theta)$ as you go along, e.g., Gilks (1990).

Rejection Sampling

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



$\pi(\alpha \mid \mathbf{y})$ 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

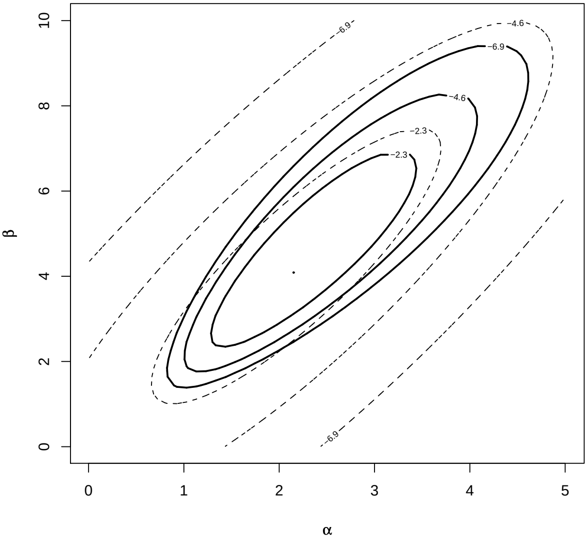
The algorithm to follow is

- 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.

Metropolis-Hastings Algorithm

The algorithm to generate a draw $\theta^{(i+1)}$:

1. Given a draw $\theta^{(i)}$ in iteration i, sample a *candidate* draw θ^* from a *proposal distribution* $J(\theta^* \mid \theta^{(i)})$
2. Calculate
$$r = \frac{\pi(\theta^* \mid \mathbf{y})/J(\theta^* \mid \theta^{(i)})}{\pi(\theta^{(i)} \mid \mathbf{y})/J(\theta^{(i)} \mid \theta^*)}$$
3. If $r \geq 1$, accept the draw θ^* and set $\theta^{(i+1)} = \theta^*$.
4. If $r < 1$, accept the draw and set $\theta^{(i+1)} = \theta^*$ with probability r.
5. Stay in place (do not accept the draw) with probability $1 - r$. Then $\theta^{(i+1)} = \theta^{(i)}$.

Random Walk Metropolis

- Most popular, easy to use.
- Proposal is a normal centered at the current draw.

$$J(\theta^{(*)} \mid \theta^{(i)}) = N(\theta^{(i)}, V)$$

- This proposal is symmetric in $\theta^{(*)}$ and $\theta^{(i)}$, so r simplifies. (When the proposal is symmetric in $\theta^{(*)}$ and $\theta^{(i)}$, the algorithm is called the *Metropolis* algorithm.)

Choosing V

1. V too small: takes long to explore parameter space
2. V too large: jumps to extremes are less likely to be accepted. Stay in the same place too long.

Do some experimentation to get V set right. The optimal acceptance rate (from some theory results) is between 25% and 50% for this type of proposal distribution. Gets lower with higher dimensional problem.

Independence Sampler

- Proposal does not depend on $\theta^{(i)}$.
- Just find a distribution $g(\theta)$ and sample from it.
- Can work well if $g(\theta)$ is a good approximation to $\pi(\theta \mid \mathbf{y})$ and has heavier tails.
- Think normal approximations to posterior (or t distributions) with inflated variances.

Metropolis-Hastings Algorithm

In general, you want to choose a $J(\theta^{(*)} \mid \theta^{(i)})$ so that starting from any $\theta^{(i)}$ you can move to any point in the support of $\pi(\theta \mid \mathbf{y})$. (The support of $J(\cdot \mid \theta)$ contains the support of $\pi(\theta \mid y)$ for every θ.)

Metropolis-Hastings Algorithm

The acceptance probability r is the product of the ratio of the target density evaluated at the candidate and current parameter values

$$\frac{\pi(\theta^* \mid \mathbf{y})}{\pi(\theta^{(i)} \mid \mathbf{y})}$$

and the ratio of the proposal distribution of the current and candidate point

$$\frac{J(\theta^{(i)} \mid \theta^*)}{J(\theta^* \mid \theta^{(i)})}$$

- The first ratio encourages the algorithm to move to parameter values that have high posterior probability.
- The second ratio accounts for the fact that the proposal distribution might favor some values of the parameter over others.