

Lecture XIII: Introduction to Posterior Simulation

Charles B. Moss

June 4, 2013

- I. Bayesian estimation requires "access" or the ability to work with the posterior distribution.
 - A. "In all except simple illustrative cases this cannot be done analytically (Geweke)."
 - B. One way to derive the posterior distribution apart from these potentially infeasible analytical solutions is to simulate the sequence of interest (i.e., integration by simulated annealing).
 - 1. For example, suppose that we are interested in estimating the mean of a standard normal distribution with an unknown mean and variance. Consider the corn production data for Georgia presented in Table 1. The sixth column represents corn yields detrended with a base year of 2010. Calculating the simple population statistics, the mean of the detrended yields is 135.99 and the variance is 191.46.
 - 2. Suppose that we are interested in estimating a production function for corn production in Georgia using a Bayesian approach.
 - 3. As a starting point, consider estimating the mean of the detrended series.
 - a. The simplest way would be to assume that yields are normally distributed. The prior would then consist of a mean and the parameters of the variance (which could either be a χ^2 times a constant or a Gamma distribution).

Table 1: Corn Production Data for Georgia

Year	Nitrogen	Phosphorous	Potash	Yield	Detrended
1964	72	43	52	39	145.02
1965	77	43	56	49	152.72
1966	95	49	67	43	144.41
1967	114	49	71	58	157.11
1968	114	53	73	40	136.80
1969	119	50	74	35	129.50
1970	118	51	74	31	123.19
1971	117	50	72	54	143.89
1972	115	52	74	52	139.58
1973	117	52	75	48	133.28
1974	124	49	75	56	138.97
1975	111	47	68	55	135.67
1976	124	51	75	62	140.36
1977	149	61	99	24	100.06
1978	139	56	98	50	123.76
1979	151	54	89	65	136.45
1980	160	60	87	42	111.15
1995	135	80	88	90	124.57
2001	127	53	79	134	154.74
2005	147	70	104	128	139.52
2010	177	68	99	145	145.00

- b. For our analysis, let's twist the prior and assume that the detrended yield is distributed Gamma

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp\left[-\frac{x}{\beta}\right] \text{ s.t. } 0 < x < \infty, \alpha > 0, \beta > 0. \quad (1)$$

- c. The question is then: What are my priors for α and β . For purpose of convenience I assume that both α and β are distributed based on a χ^2 distribution

$$\left. \begin{array}{l} \alpha \Rightarrow \tilde{\alpha}z \\ \beta \Rightarrow \tilde{\beta}z \end{array} \right\} \Leftarrow z \sim \chi_1^2 \Rightarrow f(z) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} z^{\frac{k}{2}-1} \exp\left(-\frac{z}{2}\right). \quad (2)$$

We start with $k = 1$, $\tilde{\alpha} = 96.60$, and $\tilde{\beta} = 1.41$ (the values for α and β are computed by equating the sample moments with the theoretical moments for the Gamma distribution).

- d. The likelihood function given a value of α and β are then given by

$$\begin{aligned} L(\alpha, \beta | z^0) &= \prod_{i=1}^{21} f(z_i | \alpha, \beta) \\ &= \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right)^{21} \exp\left[(\alpha-1) \sum_{j=1}^{21} \ln(z_j) - \sum_{j=1}^{21} \frac{z_j}{\beta}\right] \end{aligned} \quad (3)$$

Using R we can write the negative of the natural logarithm of the likelihood function in Equation 3 as:

```
like <- function(b) {
  ll <- log(dgamma(ycorn[1], b[1], 1/b[2]))
  for (j in 2:nrow(dta)) {
    ll <- ll + log(dgamma(ycorn[j], b[1], 1/b[2]))
  }
  return(-ll)}

```

(We compute the negative of the natural logarithm of the likelihood function so that we can estimate the same function using maximum likelihood for comparison). The posterior simulation can then be written as

```
li <- 0.0
while (li <= 0.0) {
  rdraw <- rchisq(2,1,ncp=0)
  astar <- 96.60*rdraw[1]
  bstar <- 1.41*rdraw[2]
  li <- exp(-like(cbind(astar,bstar))) }

drw <- cbind(astar,bstar,li,astar*li,bstar*li)

for (j in 2:5000) {
  li <- 0
  while (li <= 0.0) {
    rdraw <- rchisq(2,1,ncp=0)
    astar <- 96.60*rdraw[1]
    bstar <- 1.41*rdraw[2]
    li <- exp(-like(cbind(astar,bstar))) }
  drw <- rbind(drw,cbind(astar,bstar,li,astar*li,bstar*li)) }
```

- e. Taking this code a piece at a time – The first group of commands (the first while statement) draws the first observation.
- f. In drawing this observation, we first draw two χ^2 variables with $k = 1$. We call the first one α and the second one β by scaling them to our prior.
- g. Given these draws, we compute the likelihood value given our sample (again, we will want to compute the negative of the natural logarithm of the likelihood function for maximum likelihood purposes).
- h. Implicitly, the while loop then checks whether this likelihood value is numerically zero. If the value is numerically zero, the code draws another value for α and β .
- i. Given a non-zero likelihood value, we then save the value for α , β , and the likelihood value along with the likelihood

Table 2: First 15 Draws

α	β	$L(\alpha, \beta z^0)$	$\alpha \times L(\alpha, \beta z^0)$	$\beta \times L(\alpha, \beta z^0)$
16.366036	3.6063712	5.161736e-112	8.447715e-111	1.861513e-111
44.257463	2.0001953	6.988166e-82	3.092785e-80	1.397770e-81
126.944588	1.4577102	1.242922e-87	1.577823e-85	1.811821e-87
198.703487	0.6398556	1.481007e-43	2.942812e-41	9.476306e-44
444.959460	0.4103511	1.060205e-207	4.717484e-205	4.350565e-208
47.109087	11.0106761	8.827614e-297	4.158608e-295	9.719800e-296
277.763761	0.5465063	1.923570e-56	5.342981e-54	1.051243e-56
32.504004	1.4533630	4.603350e-283	1.496273e-281	6.690338e-283
142.223638	1.3999225	1.681271e-121	2.391164e-119	2.353648e-121
54.902357	1.8708224	5.719333e-60	3.140049e-58	1.069986e-59
27.319398	2.8014054	1.829368e-90	4.997723e-89	5.124802e-90
1.032492	8.2863558	1.079069e-168	1.114130e-168	8.941546e-168
68.133911	1.8793790	4.518720e-39	3.078780e-37	8.492386e-39
49.453893	1.3698303	1.481504e-178	7.326613e-177	2.029409e-178
99.463607	0.9639298	3.507438e-100	3.488624e-98	3.380924e-100

value times α and the likelihood value times β .

4. The estimates for α and β are then computed as

$$\hat{\alpha} = \frac{\sum_j \alpha_j \times L(\alpha_j, \beta_j | z^0)}{\sum_j L(\alpha_j, \beta_j | z^0)} = 90.5006$$

$$\hat{\beta} = \frac{\sum_j \beta_j \times L(\alpha_j, \beta_j | z^0)}{\sum_j L(\alpha_j, \beta_j | z^0)} = 1.6394$$
(4)

5. As a point of comparison, we compute the maximum likelihood estimates for the same parameters

$$\max_{\alpha, \beta} \ln(L(\alpha, \beta | z^0))$$
(5)

which yields $\hat{\alpha} = 92.0027$ and $\hat{\beta} = 1.4781$.

Table 3: Different Estimates of the Gamma Distribution

Parameter	α	β
Method of Moments	96.5928	1.4079
Bayesian	90.5006	1.6394
Maximum Likelihood	92.0027	1.4781

6. Consider the different estimates presented in Table 3. They are all in the same neighborhood.
- C. Geweke's discussion follows this approach.
1. We are interested in generating a sequence $\{\omega^{(m)}\}$ that "closely resembles" the theoretical distribution $\omega | (y^0, A)$.
 2. This sequence $\{\omega^{(1)}, \dots, \omega^{(M)}\}$ can be used to approximate the moments of $E[h(\omega) | y^0, A]$ which yields the Bayes estimator

$$\operatorname{argmin}_{a \in A} E[L(a, \omega) | y^0, A]. \quad (6)$$