

Lecture VIII: Credible Sets and First Pass at Metropolis Hastings

Charles B. Moss

May 17, 2013

I. Credible Sets

- A. "A credible set conveys one aspect of uncertainty about a vector of interest $\omega \in \Omega$ with pdf $p(\omega)$. It is a mapping from the distribution ω to a subset of Ω containing ω with given probability (Geweke)."
1. Most of our empirical models in econometrics use the normal distribution in some way. In most of my examples I have started out by assuming that the prior distribution was a normal.
 2. The normal density function is typically for $x \in \{-\infty, \infty\}$. Several draws may yield a fairly small probability.
 - a. Table 1 presents a synopsis of 100 draws for three normal variables. Some of the individual values become numerically small (i.e., the joint probability for the first outcome is 0.0022).
 - b. Table 2 presents selected percentiles for the 100 draws.
 3. Intuitively, these draws contribute little information to estimation.
 4. Hence, we want to concentrate our attention on those points that are more informative – more credible.

B. **Definition 2.5.1** A set $C \subseteq \Omega$ such that

$$P(\omega \in C) = \int_C p(\omega) d\nu(\omega) = 1 - \alpha \quad (1)$$

Table 1: Draws for Three Normal Variables

Draw	z_1	z_2	z_3	$p(z_1)$	$p(z_2)$	$p(z_3)$	Product
1	2.2845	-1.2153	-0.2221	0.0294	0.1906	0.3892	0.0022
2	0.1179	-1.1200	-1.1703	0.3962	0.2131	0.2012	0.0170
3	-2.1012	-0.7230	-0.1354	0.0439	0.3072	0.3953	0.0053
4	-1.3002	1.1657	-2.0551	0.1713	0.2022	0.0483	0.0017
5	0.8870	1.0126	0.6402	0.2692	0.2389	0.3250	0.0209
6	0.8877	1.0833	-0.9622	0.2690	0.2219	0.2511	0.0150
7	-1.3635	0.9516	-0.6716	0.1575	0.2537	0.3184	0.0127
8	-0.1147	-0.7883	-0.8309	0.3963	0.2924	0.2825	0.0327
9	0.0592	-0.4299	-1.8863	0.3982	0.3637	0.0673	0.0098
10	-0.6420	-0.4180	-1.4172	0.3246	0.3656	0.1461	0.0173
11	0.8473	-0.4586	-1.0656	0.2786	0.3591	0.2261	0.0226
12	0.5204	0.1158	-0.0217	0.3484	0.3963	0.3988	0.0551
13	-0.5080	-0.3335	-0.5366	0.3506	0.3774	0.3454	0.0457
14	1.7738	-0.3949	-0.9517	0.0827	0.3690	0.2536	0.0077
15	0.2372	0.1535	0.4016	0.3879	0.3943	0.3680	0.0563
16	0.6908	-1.8439	0.6710	0.3143	0.0729	0.3185	0.0073
17	0.8643	-0.5884	0.7340	0.2746	0.3355	0.3047	0.0281
18	0.2850	-1.5014	-0.2653	0.3831	0.1292	0.3852	0.0191
19	0.5173	-0.3518	0.7960	0.3490	0.3750	0.2906	0.0380
20	1.9970	-0.8639	-0.6354	0.0543	0.2747	0.3260	0.0049
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
100	-0.1368	1.1335	0.5235	0.3952	0.2099	0.3479	0.0289

Table 2: Percentiles for Random Draws

Percentile	Value
0.05	0.0018
0.10	0.0025
0.25	0.0084
0.50	0.0215
0.75	0.0380
0.90	0.0460
0.95	0.0507

is a $100(1 - \alpha)\%$ credible set for ω with respect to $p(\omega)$

- C. **Definition 2.5.2** A $100(1 - \alpha)\%$ highest probability density (HPD) credit set for ω with respect to $p(\omega)$ is a $100(1 - \alpha)\%$ credit set C for ω with the property that if $\omega_1 \in C$ and $\omega_2 \in C$ then $p(\omega_1) > p(\omega_2) \forall \omega_1 \in C$ and all $\omega_2 \in C$.
- D. **Theorem 2.5.1 Optimality of Highest-Density Regions** Suppose that the distribution of ω is absolutely continuous and $p(\omega)$ is the pdf of ω . For $C \subseteq \Omega$, let $\nu(C)$ be the Lebesgue (ordinary) measure of C . Let $C^*(\alpha) = \{C : P(\omega \in C) = 1 - \alpha\}$. Given the loss function $L(C, \omega) = k\nu(C) - I_C(\omega)$, $k > 0$ and defined on $[C^*(\alpha) \times \Omega] \rightarrow R$, C is a solution of

$$\min_{\tilde{C}} E \left[L(\tilde{C}, \omega) \right] \quad (2)$$

if and only if for all $\omega_1 \in C$ and $\omega_2 \in C$, $p(\omega_1) > p(\omega_2)$ except possibly for a collection of (ω_1, ω_2) with probability zero.

II. An introduction to Metropolis – Hastings

- A. Consider the KLEM data for agriculture presented in Table 1 of Lecture 5. Suppose that we want to estimate the equation

$$s_{it} = \alpha_0 + \alpha_1 \ln(Q_t) + \alpha_2 \ln(w_{1t}) + \alpha_3 \ln(w_{2t}) + \alpha_4 \ln(w_{3t}) + \epsilon_t \quad (3)$$

where s_{it} is the share of agricultural cost expended on capital, Q_t is the level of agricultural output, w_{1t} is the price of capital inputs (normalized by the material price), w_{2t} is the price of labor (normalized by the material price), w_{3t} is the price of energy (normalized by the material price), and ϵ_t is the residual.

- B. Slightly extending Equation 19 from Lecture 7

$$E(\alpha | z^0) = \frac{\sum_{s=1}^S \alpha_s \exp \left(- \frac{\sum_{i=1}^N (s_{it} - \alpha_0 - \alpha_1 \ln(Q_t) - \alpha_2 \ln(w_{1t}) - \alpha_3 \ln(w_{2t}) - \alpha_4 \ln(w_{3t}))^2}{2\sigma_x^2} \right)}{\sum_{s=1}^S \exp \left(- \frac{\sum_{i=1}^N (s_{it} - \alpha_0 - \alpha_1 \ln(Q_t) - \alpha_2 \ln(w_{1t}) - \alpha_3 \ln(w_{2t}) - \alpha_4 \ln(w_{3t}))^2}{2\sigma_x^2} \right)}$$

$$\ni: \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} \sim N(\tilde{\alpha}, \Sigma_\alpha)$$

(4)

C. Consider the scenario where we draw an observation from the distribution in Equation 4. We could then compute value empirical value of the kernel

$$k(x^0 | a_1) = \exp \left[\frac{\sum_{i=1}^N (s_{it} - \alpha_0 - \alpha_1 \ln(Q_t) - \alpha_2 \ln(w_{1t}) - \alpha_3 \ln(w_{2t}) - \alpha_4 \ln(w_{3t}))^2}{2\sigma_x^2} \right]$$

(5)

For example, the first draw of from the prior distribution could be

$$a_1 = \begin{pmatrix} -1.8105 \\ 0.7634 \\ 0.2929 \\ 1.6459 \\ 0.6487 \end{pmatrix} \Rightarrow K(x^0 | a_1) = 1.63E - 069$$

(6)

Next, assume that we draw a second observation

$$a_2 = \begin{pmatrix} -0.1732 \\ 1.9317 \\ -0.9525 \\ 0.8670 \\ 0.0655 \end{pmatrix} \Rightarrow K(x^0 | a_2) = 0.0000 \quad (7)$$

(or at least is numerically zero).

1. This last event could be said to lie outside the credible set.
2. To make this draw more credible, we could apply a procedure that we will develop a little later as the Metropolis Hastings Monte Carlo algorithm.
3. Specifically, we will create a new a_2 based on a linear combination of a_1 and the observed a_2 .
 - a. Specifically, suppose that we draw $\phi \in [0, 1]$ using a random uniform variable.
 - b. The new draw $\tilde{a}_2$ then becomes

$$\begin{aligned} \tilde{a}_2 &= \phi a_1 + (1 - \phi) a_2 \\ &= \phi \begin{pmatrix} -1.8105 \\ 0.7634 \\ 0.2929 \\ 1.6459 \\ 0.6487 \end{pmatrix} + (1 - \phi) \begin{pmatrix} -0.1732 \\ 1.9317 \\ -0.9525 \\ 0.8670 \\ 0.0655 \end{pmatrix} \\ &\ni: \phi = 0.905 \Rightarrow \tilde{a}_2 = \begin{pmatrix} -1.6553 \\ 0.8742 \\ 0.1748 \\ 1.5720 \\ 0.5934 \end{pmatrix} \end{aligned} \quad (8)$$

$$\Rightarrow K(x^0 | \tilde{a}_2) = 4.77E - 114.$$

While this adjusted probability is low, it is still higher (more credible) than the original draw in Equation 7.

4. To complete this idea – how do we decide whether the new point is credible? One way is to declare a trigger. Specifically, we say that the new draw is not credible if

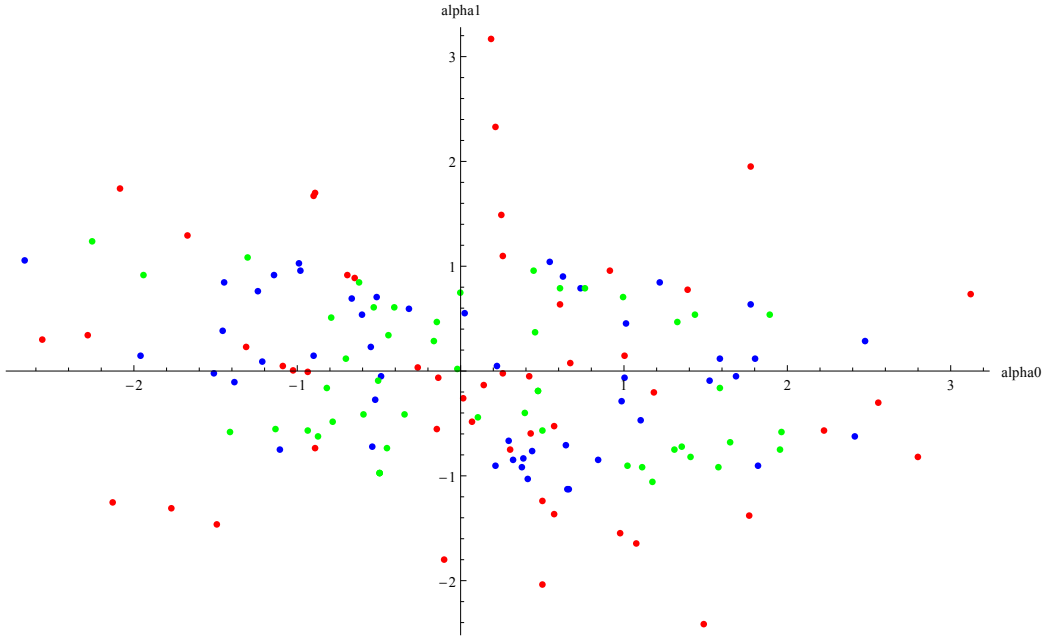


Figure 1: Effect of Metropolis Hastings Monte Carlo Sampling

$$\frac{K(x^0 | a_2)}{K(x^0 | a_1)} \leq c^* \quad (9)$$

for some finite c^* (say 0.75).

5. Figure 1 presents fifty different draws. The red points are without the Metropolis Hastings Monte Carlo adjustment. The blue points set $c^* = 0.25$ while the green dots set $c^* = 0.75$. Obviously increasing c^* concentrates the draws around the center of the distribution.

D. Next, setting the prior distribution to

$$\alpha = \begin{pmatrix} 0.25 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \end{pmatrix} \quad \Sigma_\alpha = 1.50 \times I_{5 \times 5} \quad (10)$$

yields the KLEM estimates for the share of capital presented in

Table 3. A couple of things in the table are worth noting.

1. Because of a mathematical/computing code consideration setting c^* to zero still has a MCMC effect. Specifically, these probability estimates are sensitive to an underflow error – the case where the computed number is less than machine precision. As a result setting c^* to zero still yields a resampling.
2. Table 3 presents the empirical sum and average empirical sum for the likelihood functions. Specifically,

$$\sum_{i=1}^{15,000} K(x^0 | \alpha) = 341.46 \quad (11)$$

for $c^* = 0.00$ and 15,000 draws. Hence, on average each draw represents a weight of 0.0228 in the sample estimation

$$\frac{\sum_{i=1}^{15,000} K(x^0 | \alpha)}{15,000} = 0.0228. \quad (12)$$

This measure says something about the relevency of the area drawn by the simulation. As this number declines – the simulator is drawing from a less credible set. In this case, the credibility is similar across critical values and sample sizes.

2. None of the coefficients have moved very far from the prior values. The α_0 s remain fairly close to 0.25 and the remaining coefficients remain close to zero.
3. The estimated values of standard errors for the coefficient estimates need a little more work. The values presented are the "sample estimates"

$$\hat{\Sigma}_{\alpha} = \sum_{i=1}^N \frac{K(x^0 | \alpha_i)}{\sum_{j=1}^N K(x^0 | \alpha_j)} \begin{bmatrix} \alpha_{0i} - \hat{\alpha}_0 \\ \alpha_{1i} - \hat{\alpha}_1 \\ \alpha_{2i} - \hat{\alpha}_2 \\ \alpha_{3i} - \hat{\alpha}_3 \end{bmatrix} \begin{bmatrix} \alpha_{0i} - \hat{\alpha}_0 & \alpha_{1i} - \hat{\alpha}_1 & \alpha_{2i} - \hat{\alpha}_2 & \alpha_{3i} - \hat{\alpha}_3 \end{bmatrix}. \quad (13)$$

Table 3: Estimates for KLEM Agricultural Sample 1960–1969

Parmater	Critical Value for MCMC (c^*)					
	0.00	0.05	0.10	0.25	0.50	0.75
15,000 Draws						
α_0	0.1643 (1.1391)	0.0655 (1.0768)	0.1706 (1.1147)	0.2008 (1.0732)	0.2485 (1.0793)	0.0425 (1.0739)
α_1	0.0047 (0.1269)	0.0056 (0.1207)	0.0031 (0.1219)	-0.0018 (0.1244)	-0.0073 (0.1313)	0.0122 (0.1276)
α_2	0.0787 (1.1016)	-0.0461 (1.0332)	0.0336 (1.0534)	0.1193 (1.0699)	0.0438 (0.9906)	-0.0603 (1.0924)
α_3	0.0852 (0.9341)	0.0452 (0.9317)	0.0027 (0.8753)	0.0112 (0.9506)	-0.0283 (0.8699)	0.0239 (0.8980)
α_4	-0.0126 (1.2038)	-0.0583 (1.0424)	0.0650 (1.0557)	-0.0241 (1.0574)	0.0189 (1.1403)	0.0497 (1.1654)
$\sum_{i=1}^{15,000} K(x^0 \alpha)$	341.56	356.02	376.44	371.44	373.46	340.81
$\sum_{i=1}^{15,000} K(x^0 \alpha) / 15,000$	0.0228	0.0237	0.0251	0.0248	0.0249	0.0227
30,000 Draws						
α_0	0.1539 (1.1375)	0.1880 (1.1193)	0.1045 (1.1202)	0.1396 (1.1055)	0.1989 (1.1587)	0.1913 (1.0887)
α_1	0.0011 (0.1243)	-0.0019 (0.1236)	0.0026 (0.1280)	0.0021 (0.1291)	-0.0058 (0.1333)	-0.0085 (0.1312)
α_2	0.0360 (1.0767)	0.0529 (1.0473)	-0.0074 (1.0614)	0.0403 (1.1031)	0.0084 (1.0622)	-0.0181 (1.1105)
α_3	0.0148 (0.9383)	0.0014 (0.8923)	-0.0023 (0.8980)	0.0305 (0.9037)	0.0132 (0.9030)	-0.0408 (0.9193)
α_4	-0.0071 (1.0976)	-0.0054 (1.1212)	-0.0401 (1.1114)	-0.0438 (1.0994)	-0.0518 (1.1066)	-0.0466 (1.1272)
$\sum_{i=1}^{15,000} K(x^0 \alpha)$	696.90	735.53	735.65	735.74	720.02	736.31
$\sum_{i=1}^{15,000} K(x^0 \alpha) / 15,000$	0.0232	0.0245	0.0245	0.0245	0.0240	0.0245
45,000 Draws						
α_0	0.1853 (1.1735)	0.1868 (1.1157)	0.1649 (1.1012)	0.1011 (1.0775)	0.1827 (1.0918)	0.1559 (1.1326)
α_1	0.0010 (0.1255)	-0.0060 (0.1248)	-0.0022 (0.1244)	0.0072 (0.1263)	-0.0032 (0.1271)	-0.0001 (0.1394)
α_2	0.0183 (1.0641)	-0.0392 (1.0904)	-0.0159 (1.0728)	0.0425 (1.0563)	-0.0139 (1.0588)	0.0402 (1.0679)
α_3	0.0189 (0.9475)	-0.0266 (0.9237)	-0.0089 (0.8924)	0.0108 (0.9172)	-0.0012 (0.8957)	-0.0045 (0.9121)
α_4	0.0389 (1.1240)	-0.0097 (1.1286)	-0.0040 (1.1378)	0.0383 (1.0790)	-0.0147 (1.1070)	0.0110 (1.1254)
$\sum_{i=1}^{15,000} K(x^0 \alpha)$	1047.32	1105.35	1097.81	1075.02	1097.04	1060.78
$\sum_{i=1}^{15,000} K(x^0 \alpha) / 15,000$	0.0233	0.0246	0.0244	0.0239	0.0244	0.0236

The true information must include information from the prior (both the prior variance matrix for the α s as well as the change between the prior and the new estimates). Using the prior from Equation 10

$$\tilde{\Sigma}_{\alpha} = \phi_1 \Sigma_{\alpha} + \phi_2 \hat{\Sigma}_{\alpha} + \phi_3 \begin{bmatrix} \hat{\alpha}_0 - \alpha_0 \\ \hat{\alpha}_1 - \alpha_1 \\ \hat{\alpha}_2 - \alpha_2 \\ \hat{\alpha}_3 - \alpha_3 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_0 - \alpha_0 & \hat{\alpha}_1 - \alpha_1 & \hat{\alpha}_2 - \alpha_2 & \hat{\alpha}_3 - \alpha_3 \end{bmatrix} \quad (14)$$

where ϕ_1 , ϕ_2 , and ϕ_3 are weights $\phi_1 + \phi_2 + \phi_3 = 1$.