

Lecture XIV: Direct Sampling, Acceptance and Importance Sampling

Charles B. Moss

June 21, 2013

I. Generalization of Notation

- A. Our primary interest is in simulating the posterior distribution for a parameter of interest – say $h(\omega)$.
- B. However, the methods can be used to develop any distribution.
- C. In this chapter Geweke develops the distribution function of ω – $\omega \sim p(\omega | \theta, I)$ and $\theta \sim p(\theta | I)$ where I denotes the distribution of interest.

II. Direct Sampling

- A. In this section, we develop the simulation of a closed model like the normal linear regression model.
 - 1. In our previous discussion of the normal linear regression model, we developed

$$\bar{s}^2 h | (y^0, X, A) \sim \chi^2(\bar{\nu}) \tag{1}$$

$$\beta | (h, y^0, X, A) \sim N(\bar{\beta}, h^{-1} \bar{H}^{-1})$$

”The following result shows that it is possible to use the sequence $\{\theta^{(m)}\}$ to approximate several aspects of the distribution of ω , including moments.

- B. **Theorem 4.1.1 Approximation of Distributions and Moments by Direct Sampling** Suppose that the sequence $\{\theta^{(m)}, \omega^{(m)}\}$ is i.i.d., with $\theta^{(m)} \in \Theta$, $\omega^{(m)} \in \Omega$, $\omega^{(m)} \sim p(\omega | \theta^{(m)}, I)$, and $\omega^{(m)} | \theta^{(m)} \sim p(\omega | \theta^{(m)}, I)$. Let $h : \Omega \rightarrow R^1$ and consider several conditions:

1. $E[h(\omega) | I] = h$.
2. $\text{var}[h(\omega) | I] = \sigma^2$.
3. For given $p \in (0, 1)$, there is a unique h_p such that the statements

$$P[h(\omega) \leq h_p | I] \geq p \text{ and } P[h(\omega) \geq h_p | I] \geq 1 - p \quad (2)$$

are both true.

4. The pdf $p[h(\omega) | I]$ is continuous, and for the unique h_p corresponding to p , $p[h(\omega) = h_p | I] > 0$.

Then

- a. $\omega^{(m)} \stackrel{i.i.d.}{\sim} p(\omega | I)$, regardless of whether any of conditions 1 - 4 are true.
- b. Given condition 1, $\bar{h}^{(M)} = M^{-1} \sum_{m=1}^M h(\omega^{(m)})$.
- c. Given conditions 1 and 2, $M^{1/2} (\bar{h}^{(M)} - \bar{h}) \xrightarrow{d} N(0, \sigma^2)$, and

$$\hat{\sigma}^{2(M)} = M^{-1} \sum_{m=1}^M [h(\omega^{(m)}) - \bar{h}^{(M)}]^2 \xrightarrow{a.s.} \sigma^2. \quad (3)$$

Let $\hat{h}_p^{(M)}$ be any real number such that

$$M^{-1} \sum_{m=1}^M I_{(-\infty, \hat{h}_p^{(M)})} [h(\omega^{(m)})] \geq p \text{ and} \quad (4)$$

$$M^{-1} \sum_{m=1}^M I_{[\hat{h}_p^{(M)}, \infty)} [h(\omega^{(m)})] \geq 1 - p.$$

- d. Given condition 3 $\hat{h}_p^{(M)} \rightarrow h_p$.
- e. Given conditions 3 and 4

$$M^{1/2} (\hat{h}_p^{(M)} - h_p) \xrightarrow{d} N[0, p(1-p)/p[h(\omega) = h_p | I]^2] \quad (5)$$

C. Given conditions 1 and 2

$$\bar{h}^{(M)} \sim N(h, \hat{\sigma}^{2(M)}/M) \quad (6)$$

is valid for large M .

D. "In this case $(\hat{\sigma}^{2(M)}/M)^{1/2}$ is known as the *numerical standard error* (NSE) of $\bar{h}^{(M)}$ (Geweke)."

E. **Theorem 4.1.2 Approximation of Bayes Actions by Direct Sampling** Suppose that the sequence $\{\theta^{(m)}, \omega^{(m)}\}$ is i.i.d., with $\theta^{(m)} \sim p(\theta|I)$ and $\omega^{(m)} | (\theta^{(m)}, I) \sim p(\omega | \theta^{(m)}, I)$. Let $L(a, \omega) \geq 0$ be a loss function defined on $A \times \Omega$, where A is an open subset of R^m . Suppose that the risk function

$$R(a) = \int_{\Omega} \int_{\Theta} L(a, \omega) p(\theta|I) p(\omega|\theta, I) d\theta d\omega \quad (7)$$

has a strict global minimum at $\hat{a} \in A \subset R^m$. Consier several additional conditions, for a suitably open neighborhodd of $\hat{a}$, $N(\hat{a})$:

1. $M^{-1} \sum_{m=1}^M L(a, \omega^{(m)})$ converges uniformly to $R(a)$ for all $a \in N(\hat{a})$, almost surely.
2. $\partial L(a, \omega) / \partial a$ exists and a continuous for a , for all $\omega \in \Omega$ and all $a \in N(\hat{a})$.
3. $\partial^2 L(a, \omega) / \partial a \partial a'$ exists and is a continuous function of a , for all $\omega \in \Omega$ and all $a \in N(\hat{a})$.
4. $B = \text{var} [\partial L(a, \omega) / \partial a]_{a=\hat{a}} | I$ exists and is finite.
5. $H = E[\partial^2 L(a, \omega) / \partial a \partial a']_{a=\hat{a}} | I$ exists and is finite and nonsingular.
6. For any $\epsilon > 0$, there exists M , such that

$$P \left[\sup_{a \in N(\hat{a})} |\partial^3 L(a, \omega) / \partial a_i \partial a_j \partial a_k| < M_{\epsilon} | I \right] \geq 1 - \epsilon \quad (8)$$

for all $i, j, k = 1, \dots, m$.

Let A_m be the set of all roots $M^{-1} \sum_{m=1}^M \partial L(a, \omega^{(m)}) / \partial a = 0$. Then

- a. Given conditions 1 and 2, for any $\epsilon > 0$

$$\lim_{M \rightarrow \infty} P \left[\inf_{a \in A_M} (a - \hat{a})' (a - \hat{a}) > \epsilon | I \right] = 0. \quad (9)$$

- b. Given conditions 1-6, if $\hat{a}_M$ is any element of A_M such that $\hat{a}_M \xrightarrow{P} \hat{a}$, then

- i. $M^{1/2} (\hat{a}_M - \hat{a}) \xrightarrow{d} N(0, H^{-1} B H^{-1})$.
- ii. $M^{-1} \sum_{m=1}^M \partial L(a, \omega^{(m)}) / \partial a|_{a=\hat{a}_M} \cdot \partial L(a, \omega^{(m)})|_{a=\hat{a}_M} \xrightarrow{P} B$.
- iii. $M^{-1} \sum_{m=1}^M \partial^2 L(a, \omega^{(m)}) / \partial a \partial a'|_{a=\hat{a}_M} \xrightarrow{P} H$.

III. Acceptance and Importance Sampling

A. "Suppose that we cannot derive a method for drawing i.i.d. random vectors $\theta^{(m)}$ directly from the density $p(\theta|I)$ but we can simulate i.i.d. drawings from a density $p(\theta|S)$ that is similar to $p(\theta|I)$. Then it may be possible to learn about many aspects of the vector of interest ω , but the sense in which $p(\theta|S)$ is similar to $p(\theta|I)$ (Geweke)."

B. Acceptance Sampling

1. **Theorem 4.2.1 Acceptance Sampling** Let $k(\theta|I) = c_I \cdot p(\theta|I)$ be a kernel of the density of interest $p(\theta|I)$, and let $k(\theta|S) = c_S \cdot p(\theta|S)$ be a kernel of the source density $p(\theta|S)$. Let $r = \sup_{\theta \in \Theta} k(\theta|I) / k(\theta|S) < \infty$. Suppose that $\theta^{(m)}$ is drawn as follows:
 1. Draw u uniform on $[0, 1]$.
 2. Draw $\theta^* \sim p(\theta|S)$.
 3. If $u > k(\theta^*|I) / [r \cdot k(\theta^*|S)]$, then return to step 1.
 4. Set $\theta^{(m)} = \theta^*$.
2. "The computational efficiency of the algorithm is driven by those θ for which $p(\theta|S)$ has the most severe undersampling relative to $p(\theta|I)$ (Geweke)."

B. Importance Sampling

1. **Theorem 4.2.2 Approximation of Moments by Importance Sampling** Suppose that the sequence $\{\theta^{(m)}, \omega^{(m)}\}$ is independent and identically distributed, with $\theta^{(m)} \sim p(\theta|S)$ and $\omega^{(m)} \sim p(\omega|\theta^{(m)}, I)$. Define the weighting function $w(\theta) = k(\theta|I) / k(\theta|S)$, let $h : \rightarrow R^1$, and consider several additional conditions:
 1. $E[h(\omega)|I] = h$ exists.
 2. $\text{var}[h(\omega)|I] = \sigma^2$ exists.

3. The support of $p(\theta | S)$ includes Θ .
4. $w(\theta)$ is bounded above on Θ .

Then

- a. Given conditions 1 and 3

$$\bar{h}^{(M)} = \frac{\sum_{m=1}^M w(\theta^{(m)}) h(\omega^{(m)})}{\sum_{m=1}^M w(\theta^{(m)})} \xrightarrow{a.s.} \bar{h}. \quad (10)$$

- b. Given conditions 1-4

$$M^{1/2} (\bar{h}^{(M)} - \bar{h}) \xrightarrow{d} N(0, \tau^2) \quad (11)$$

and

$$\hat{\tau}^{2(M)} = \frac{M \sum_{m=1}^M [h(\omega^{(m)}) - \bar{h}^{(M)}]^2 w(\theta^{(m)})^2}{\left[\sum_{m=1}^M w(\theta^{(m)}) \right]^2} \xrightarrow{a.s.} \tau^2. \quad (12)$$

2. The basic concept of importance sampling is to set a lower bound on the draw. In most cases, we draw several outcomes that are numerically zero (i.e., the probability 1.0E-125 is essentially zero).
3. **Theorem 4.2.3 Approximation of Bayes Actions by Importance Sampling** Suppose that the sequence $\{\theta^{(m)}, \omega^{(m)}\}$ is independent and identically distributed, with $\theta^{(m)} \sim p(\theta | S)$ and $\omega^{(m)} | \theta^{(m)} \sim p(\omega | \theta^{(m)}, I)$. Suppose that the support of $p(\theta | S)$ includes Θ , and the weighting function $w(\theta) = k(\theta | I) / k(\theta | S)$ is bounded above. Let $L(a, \omega) \geq 0$ be a loss function defined on $A \times \Omega$, where A is an open subset of R^m . Suppose that the risk function

$$R(a) = \int_{\Omega} \int_{\Theta} L(a, \omega) p(\theta | I) p(\omega | \theta, I) d\theta d\omega \quad (13)$$

has a strict global minimum at $\hat{a} \in A \subset R^m$. Consider several additional conditions, for a suitably defined open neighborhood of $\hat{a}$, $N(\hat{a})$:

1. $M^{-1} \sum_{m=1}^M L(a, \omega^{(m)})$ converges uniformly to $R(a)$ for all $a \in N(\hat{a})$, almost surely.
2. $\partial L(a, \omega) / \partial a$ exists and is a continuous function of a , for all $\omega \in \Omega$ and all $a \in N(\hat{a})$.
3. $\partial^2 L(a, \omega) / \partial a \partial a'$ exists and is a continuous function of a , for all $\omega \in \Omega$ and all $a \in N(\hat{a})$.
4. $B = \text{var} \left[\partial L(a, \omega) / \partial a \partial a' \big|_{a=\hat{a}} w(\theta)^{1/2} | S \right]$ exists and is finite.
5. $H = E \left[\partial^2 L(a, \omega) / \partial a \partial a' \big|_{a=\hat{a}} | S \right]$ exists and is finite and nonsingular.
6. For any $\epsilon > 0$, there exists M_ϵ such that

$$P \left[\sup_{a \in N(\hat{a})} \left| \partial^3 L(a, \omega) / \partial a_i \partial a_j \partial a_k \right| < M_\epsilon | S \right] \geq 1 - \epsilon \quad (14)$$

for all $i, j, k = 1, \dots, m$.

Let A_M be the set of all roots $M^{-1} \sum_{m=1}^M [\partial L(a, \omega^{(m)}) / \partial a] w(\theta^{(m)}) = 0$. Then

- a. Given conditions 1 and 2, for any $\epsilon > 0$, we obtain

$$\lim_{M \rightarrow \infty} P \left[\inf_{a \in A_M} (a - \hat{a})' (a - \hat{a}) > \epsilon | S \right] = 0. \quad (15)$$

- b. Given conditions 1 – 6, if $\hat{a}_M$ is any element of A_M such that $\hat{a}_M \xrightarrow{p} \hat{a}$, then

- i. $M^{1/2} (\hat{a}_M - \hat{a}) \xrightarrow{d} N(0, H^{-1} B H^{-1})$.
- ii. $M^{-1} \sum_{m=1}^M w(\theta^{(m)})^2 \partial L(a, \omega^{(m)}) / \partial a \big|_{a=\hat{a}_M} \cdot \partial L(a, \omega^{(m)}) / \partial a' \big|_{a=\hat{a}_M} \xrightarrow{p} B$.
- iii. $M^{-1} \sum_{m=1}^M w(\theta^{(m)}) \partial^2 L(a, \omega^{(m)}) / \partial a \partial a' \big|_{a=\hat{a}_M} \xrightarrow{p} H$.

IV. Bayesian estimation of the Inverse Hyperbolic Sine Transformation using Simulation

A. Moss and Shonkwiler (1993) propose the use of the Inverse Hyperbolic Sine (IHS) transformation to normality to model non-normal crop yields.

1. The general form of the IHS transformation is

$$g(x, \theta) = \frac{\sinh^{-1}(\theta x)}{\theta} = \frac{\ln(\theta x + \sqrt{1 + \theta^2 x^2})}{\theta} \quad (16)$$

2. In this representation the effect of θ is symmetric (i.e., $g(x, \theta) = g(x, -\theta)$). In addition, as $\theta \rightarrow 0$, the transformation approaches linear.

3. From an econometric perspective, we can envision this as transforming the residual of a regression to normality

$$x = \delta' z. \quad (17)$$

Moss and Shonkwiler propose using a Kalman filter. In this study, we simply assume a constant

$$\tilde{g}(x, \theta, \delta) = \frac{\sinh^{-1}(\theta [x - \delta])}{\theta} = \frac{\ln(\theta [x - \delta] + \sqrt{1 + \theta^2 [x - \delta]^2})}{\theta} \quad (18)$$

4. The distribution function is then completed by integrating the results of 18 into the normal density function along with the Jacobian of the transformation (i.e., the derivative of Equation 18 with respect to x)

$$f(x, \mu, \sigma, \theta, \delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\tilde{g}(x, \theta, \delta) - \mu)^2}{2\sigma^2}\right] \frac{1}{\sqrt{1 + \theta^2 (x - \delta)^2}} \quad (19)$$

5. In completing this formulation we note that x is the sample values (i.e., the Georgia corn data from Lecture 13). Hence, prior distributions are required for μ , σ , θ , and δ .

- a. In this case our prior for δ and σ will be derived from the distribution. Taking δ to be the mean of the sample, we assume $\delta \sim N(135.99, 191.46/21)$. Similarly, the $\sigma^2 \sim 191.46\chi_1^2$.
- b. Given the symmetric formulation for θ and given that several estimates for corn find $\theta = 0.5$, we assume $\theta \sim 0.50\chi_1^2$.
- c. Finally, we assume a prior of a symmetric distribution – $\mu \sim N(0, 20.0)$.