Direct Simulation Methods

Econ 690

Purdue University

Outline

1. Monte Carlo Integration
2. The Method of Composition
3. The Method of Inversion
4. Acceptance/Rejection Sampling

Monte Carlo Integration

Suppose you wish to calculate a posterior moment of the form:

$$E[g(\theta)|y] = \frac{\int_\Theta g(\theta)p(y|\theta)p(\theta)d\theta}{\int_\Theta p(y|\theta)p(\theta)d\theta}.$$

With Monte Carlo Integration, we assume that we can draw directly from the posteiror $p(\theta|y)$.

If this is true, then, under reasonable conditions, a law of large numbers guarantees that

- In the above $\theta^{(i)} \stackrel{iid}{\sim} p(\theta|y)$.
- Note that n here is under our control, as it refers to the Monte Carlo sample size rather than the number of observations.
- Thus, we can estimate the desired (finite-sample) moment with arbitrary accuracy.
- This technique has a very demanding prerequisite - that we can draw directly from $p(\theta|y)$.

Method of Composition

- The method of composition provides a convenient way of drawing from $p(\theta|y)$ when the joint distribution is decomposed into a product of marginals and conditionals, and each of these component pieces can be easily drawn from.

- For example, suppose $\theta = [\theta_1' \;\; \theta_2']'$ and that $p(\theta_2|y)$ and $p(\theta_1|\theta_2, y)$ are well known densities that are easily sampled.

Under these conditions, one can obtain a draw from $p(\theta|y)$ in the following way:

1.
2.

Why this works is, perhaps, obvious, but consider $A \times B \subseteq \Theta_1 \times \Theta_2$. Then,

-

Consider the linear regression model:

$$y = X\beta + u, \quad u|X \sim N(0, \sigma^2 I_n)$$

under the prior

$$p(\beta, \sigma^2) \propto \sigma^{-2}.$$

In our linear regression model notes, we showed

$$\beta|\sigma^2, y \sim N[\hat{\beta}, \sigma^2 (X'X)^{-1}]$$

and

$$\sigma^2|y \sim IG\left[\frac{n-k}{2}, 2[(y - X\hat{\beta})'(y - X\hat{\beta})]^{-1}\right].$$

Thus, we can sample from the joint posterior $p(\beta, \sigma^2|y)$ by first sampling σ^2 from its marginal posterior, and then sampling β from the conditional normal posterior.

The method of composition can also prove to be a very valuable tool for problems of (posterior) prediction.

To this end, consider an out-of-sample value y_f which is presumed to be generated by our regression model:

$$y_f = X_f \beta + u_f, \quad u_f | X_f \sim N(0, \sigma^2).$$

1. Note that $y_f | \beta, \sigma^2$ does not depend on y. (But does through β and σ^2.)
2. The goal is to simulate draws from the posterior predictive:

$$p(y_f | y),$$

which does not depend on any of the model's parameters.

To generate draws from this posterior predictive, we first consider the joint posterior distribution:

$$p(y_f, \beta, \sigma^2 | y).$$

If we can draw from this distribution, we can use only the y_f draws (and ignore those associated with β and σ^2) as draws from the marginal $p(y_f | y)$.

How can we do this?

Note

-

This suggests that draws from the marginal posterior predictive distribution can be obtained by

1.
2.
3.
4. Note, of course, this requires that X_f is known.
5. Doing this many times will produce a set of draws from the posterior predictive $y_f|y$.

- Let's apply this method to generate draws from the posterior predictive using our log wage example:

$$\log(wage)_i = \beta_0 + \beta_1 Ed_i + u_i.$$

- The method just described could be applied directly to sample from the predictive distribution of (log) hourly wages.
- However, the wage density itself is actually more interpretable.
- To sample from the posterior predictive of wages (in levels), we can consider drawing from an augmented density of the form:

$$p(w_f, y_f, \beta, \sigma^2 | y)$$

where

$$w_f = \exp(y_f).$$

$$p(w_f, y_f, \beta, \sigma^2 | y)$$

We can write this joint distribution as follows:

-

where the last line follows since the distribution of w_f only depends on y_f and, in fact,

-

Thus, within the context of our example, we can generate draws from the posterior predictive distribution of *hourly wages* w_f as follows:

1. Generate

2. Generate

3. Generate

4. Generate

- We apply this technique to our data set and generate 10,000 draws from the posterior predictive distribution of hourly wages for two cases: Ed $=$ 12 and Ed $=$ 16.
- The 10,000 draws are then smoothed nonparametrically via a kernel density estimator. [I will provide a MATLAB file for you that does these calculations].
- Graphs of these densities are provided on the following page.

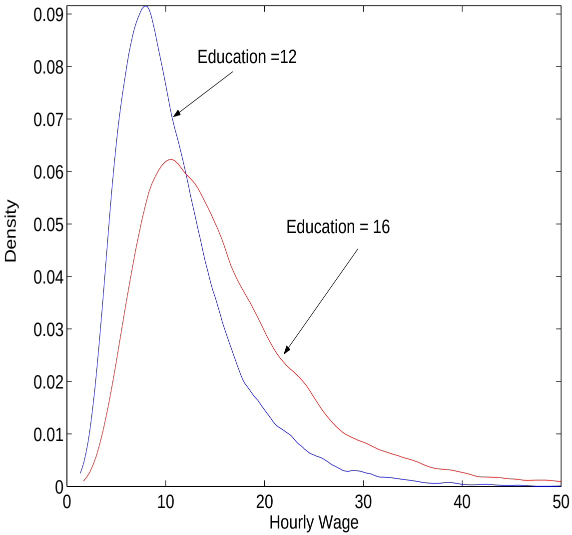

Figure: Posterior Predictive Hourly Wage Densities

- The (posterior predictive) mean hourly wage for high school graduates is (approximately) \$11.07
- The mean hourly wage for those with a BA is (approximately) \$15.88
- The posterior probability that a high school graduate will receive an hourly wage greater than \$15 is

$$\Pr(w_f > 15 | Ed_f = 12, y) \approx .19$$

- The posterior probability that an individual with a BA will receive an hourly wage greater than \$15 is

$$\Pr(w_f > 15 | Ed_f = 16, y) \approx .44$$

- If you are curious, doing the same exercise for someone with a Ph.D., i.e., Ed $= 20$, gives $\Pr(w_f > 15 | Ed_f = 20, y) \approx .72$

The Method of Inversion

- Suppose that X is a continuous random variable with distribution function F and density f. Further, assume that the distribution function F can be easily calculated.
- Let $U \sim \mathcal{U}(0,1)$, a Uniform random variable on the unit interval, and define $Y = F^{-1}(U)$.
- Derive the distribution of the random variable Y.

$$U \sim \mathcal{U}(0,1), \quad Y = F^{-1}(U).$$

- We can establish the desired result using a change of variables.
- First, note that

  with $I(\cdot)$ denoting an indicator function and $U = F(Y)$.
- Thus,
- Therefore Y has distribution function F.

- How is the result we just established useful?
- This result is extremely useful in cases where the cdf F and its inverse are easily calculated because it provides a way to generate draws from f.
- Specifically, we can:
  1.
  2.
- It follows that Y is a draw from f. We now provide several examples of this method.

Consider an exponential random variable with density function

$$p(x|\theta) = \theta^{-1} \exp(-x/\theta), \quad x > 0.$$

Show how the inverse transform method can be used to generate draws from the exponential density.

Note that, for $x > 0$,

$$\begin{aligned} F(x) &= \int_0^x \frac{1}{\theta} \exp\left(-\frac{t}{\theta}\right) \, dt \\ &= 1 - \exp\left(-\frac{x}{\theta}\right). \end{aligned}$$

The results of our previous derivation show that if we can solve for x in the equation

-

with u denoting a realized draw from a $\mathcal{U}(0,1)$ distribution, then x has the desired exponential density.

A little algebra provides

-

as the solution.

- Let $x \sim TN_{[a,b]}(\mu, \sigma^2)$ denote that x is a *truncated Normal* random variable.
- Specifically, this notation indicates that x is generated from a Normal density with mean μ and variance σ^2, which is truncated to lie in the interval $[a, b]$. The density function for x in this case is given as
- Show how the inverse transform method can be used to generate draws from this truncated Normal density.

For $a \leq x \leq b$, the c.d.f. of the truncated Normal random variable is

-

Therefore, if x is a solution to the equation

$$u = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)},$$

where u is realized draw from a $\mathcal{U}(0,1)$ distribution, then $x \sim TN_{[a,b]}(\mu, \sigma^2)$.

With a little algebra it follows that

$$x = \mu + \sigma\Phi^{-1}\left(\Phi\left(\frac{a-\mu}{\sigma}\right) + u\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]\right)$$

is a solution.

When:

- $b = \infty$, so that the random variable is truncated from below only, we substitute $\Phi[(b-\mu)/\sigma]$ with 1 in the above expression.
- $a = -\infty$, so that the random variable is truncated from above only, we substitute $\Phi[(a-\mu)/\sigma]$ with 0 in the above expression.

Suppose $y|\mu, \sigma^2 \sim LN(\mu, \sigma^2)$, implying

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp\left(-\frac{1}{2\sigma^2}[\ln y - \mu]^2\right), \quad y > 0.$$

We seek to use inversion to generate draws from the lognormal distribution.

Note, for $c > 0$,

-

Let

-

Then

-

Thus, setting

-

produces a draw from the desired lognormal distribution.

- Rejection Sampling provides another way of obtaining draws from some density of interest.
- Generally, it proceeds as follows:
  1. Draw from some *approximating density*
  2. Compute a statistic, like a "critical" value.
  3. If some condition is met related to the size of the critical value, then keep the draw as a draw from the target density. Otherwise, start over until the needed condition is satisfied.

Consider the following strategy for drawing from a density $f(x)$ defined over the compact support $a \leq x \leq b$ with

$$M \equiv \max_{a \leq x \leq b} f(x):$$

1. Generate two independent Uniform random variables U_1 and U_2 as follows:

$$U_i \stackrel{iid}{\sim} \mathcal{U}(0,1), \quad i = 1, 2.$$

2. If

$$MU_2 > f(a + [b-a]U_1),$$

   start over. That is, go back to the first step and generate new values for U_1 and U_2, and again determine if $MU_2 > f(a + [b-a]U_1)$.

3. When

$$MU_2 \leq f(a + [b-a]U_1))$$

   set

$$x = a + (b-a)U_1 \quad \text{as a draw from } f(x).$$

We will answer the following two questions regarding this algorithm:

- (a) What is the probability that any particular iteration in the above algorithm will produce a draw that is accepted?
- (b) Sketch a proof as to why x, when it is accepted, has distribution function $F(x) = \int_a^x f(t)\ dt$.

First, we will consider question (a) and investigate the probability of acceptance. Note that

-

- The third line uses the fact that U_1 and U_2 are independent,
- The fourth and fifth lines follow from the fact that $U_i \sim \mathcal{U}(0,1), \;\; i = 1, 2,$
- The fifth line also applies a change of variable, setting $t = a + (b - a)U_1$.
- Thus the probability of accepting a candidate draw in the algorithm is $[M(b-a)]^{-1}$.
- Note that, when using this method to sample from a Uniform distribution on $[a, b]$, all candidates from the algorithm are accepted.

Let us now move on to part (b), which seeks to establish why this algorithm works: Note that

-

Therefore, a candidate draw which is accepted from the acceptance/rejection method has distribution function F, as desired.

Let us now consider an applicaton of this rejection sampling algorithm.

Consider the *triangular density* function, given as

$$p(x) = 1 - |x|, \quad x \in [-1, 1].$$

Describe how the rejection sampling algorithm can be used to generate draws from this density function.

- For this simple example, note that $M = 1$ and $b - a = 2$, so that the overall acceptance rate is one-half.
- That is, we would expect that, say, 50,000 pairs of independent Uniform variates in the acceptance/rejection algorithm would be needed in order to produce a final sample of 25,000 draws.
- Here is a small MATLAB program that does this:

```
iter = 10000;
U2 = rand(iter,1);
U1 = rand(iter,1);
fff = U2 - 1 + abs(2*U1-1);
points = find(fff <= 0);
draws = -1 + 2*U1(points);
```