Monte Carlo Simulation

Econ 5170

Computational Methods in Economics

2022-2023 Spring

Outline

1. Random Number Generation

2. A Simple Example of Simulation

3. Monte Carlo Integration

4. Bootstrap

Pseudorandom Number Generation

- Random numbers are seldom used
  - Possible methods
    - Flipping coins
    - Geiger counters measuring radioactivity
  - Disadvantages
    - Expensive given RA salaries
    - RA's complain about radiation risk
- Monte Carlo propagandists
  - Use deterministic sequences
  - Act as if they are random sequences

Pseudorandom Number Generation

- Pseudorandom numbers are used instead
  - They are deterministic sequences, $X_{k+1} = f\left(X_k, X_{k-1}, X_{k-2}, \ldots\right)$
  - They pass *some* randomness tests, such as
    - Unbiasedness

$$\frac{1}{N}\sum_{k=1}^{N} X_k = \mu \equiv E\{X\}$$

    - Zero serial correlation

$$0 = \sum_{k=1}^{N} \left(X_k - \mu\right) X_{k+1}$$

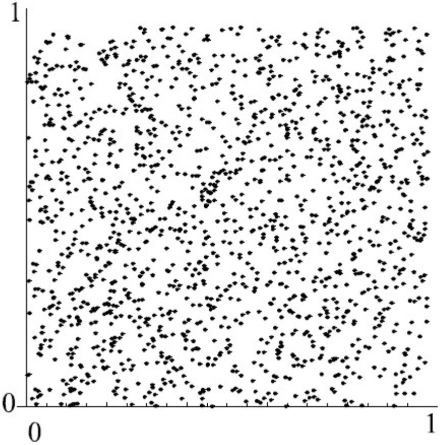
Uniform Random Number Generation

- Linear congruential method (LCM):

$$X_{k+1} = (aX_k + c) \bmod m$$

  - Will generate pseudorandom sequence if parameters chosen well
  - Will eventually cycle; choose parameters to get long cycle
  - $Y_n \equiv (X_{2n+1}, X_{2n+2})$ is a pseudorandom two-dimensional set of points. Similar for R^d
  - LCM generators have fallen into disfavor since they lie on a finite set of hyperplanes.

Uniform Random Number Generation



1500 Points generated by LCM

Uniform Random Number Generation

- Nonlinear schemes
  - $X_{k+1} = f(X_k) \mod m$
  - Fibonacci generator $X_k = (X_{k-1} + X_{k-2}) \mod m$. This sequence has a number of poor properties. In particular, if X_{k-1} and X_{k-2} are small relative to m, so will be X_k.
  - The Fibonacci-like scheme

    $$X_k = (X_{k-24} \cdot X_{k-55}) \text{ mod } 2^{32}$$

    has a period $\sim 10^{25}$ and passes many randomness tests.

Nonuniform Random Number Generation

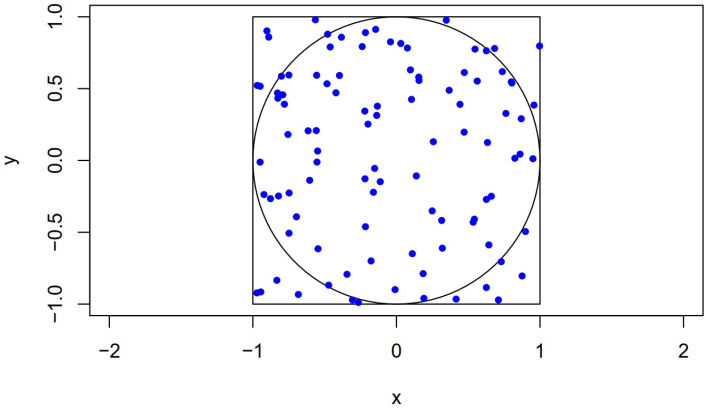
- Need to generate nonuniform random numbers
- Inversion:
  - Suppose X has distribution $F(x)$
  - Then $F^{-1}(U)$ has distribution $F(x)$
  - To approximate X, we compute $y_k = F^{-1}(x_k)$ where x_k is a uniform pseudorandom sequence
- Normal random variables: A special method
  - Suppose U_1 and $U_2 \sim U[0,1]$
  - Then $X_1, X_2 \sim N(0,1)$ are independent where

$$X_1 = \cos(2\pi U_1)\sqrt{-2\ln U_2}$$
$$X_2 = \sin(2\pi U_1)\sqrt{-2\ln U_2}$$

How to Approximate π

- Zu Chongzhi (429–500 AD), an ancient Chinese mathematician, calculated π being between 3.1415926 and 3.1415927, which for 900 years held the world record of the most accurate π.
- He used a deterministic approximation algorithm.
- Now imagine that we present to Zu Chongzhi, with full respect and admiration, a modern computer. How can we achieve a better approximation?

How to Approximate π



How to Approximate π

$$\frac{\text{area of a circle}}{\text{area of the smallest encompassing square}} = \frac{\pi r^2}{(2r)^2} = E[\{x^2 + y^2 < 1\}]$$

It implies that $\pi = 4E[\{x^2 + y^2 < 1\}]$.

Monte Carlo and Simulation Methods

- Gaussian, monomial, and Newton-Cotes formulas
  - use predetermined nodes
  - aim at high accuracy
  - need many nodes
- Sampling methods
  - Generate a sequence of points
  - Short sequence produces rough approximation
  - Longer sequences produce better approximations
- Monte Carlo sampling methods
  - Use law of large numbers “intuition”
  - Order $N^{-1/2}$ convergence
  - Use probability theory to prove theorems
  - Use number-theoretic methods to generate deterministic sequences which appear random

Monte Carlo Integration

- Probability theory
  - If X_i are i.i.d. r.v.'s, density q(x), and support [0, 1], then

$$\bar{X} \equiv \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i = \int_0^1 xq(x)dx, \text{ a.s.}$$

$$\text{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{\sigma_x^2}{N}$$

  - If σ_x^2 is not known a *priori*, an unbiased estimator is $\hat{\sigma}_x^2 \equiv (N-1)^{-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$

Monte Carlo Integration

- LLN suggests Monte Carlo quadrature:
  - If X $\sim$ U[0, 1], then $I_f = \int_0^1 f(x)dx = E\{f(X)\}$
  - The *crude Monte Carlo* method makes N draws from U[0, 1], $\{x_i\}_{i=1}^N$ and defines

$$\hat{I}_f \equiv \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

$$\hat{\sigma}_f^2 = (N-1)^{-1} \sum_{i=1}^{N} \left(f(x_i) - \hat{I}_f \right)^2$$

- $\hat{I}_f$ is a statistical estimate of $\int_0^1 f(x)dx$
  - $\hat{I}_f$ is an unbiased estimate of $\int_0^1 f(x)dx$
  - The variance of the $\hat{I}_f$ estimate is

$$\sigma_{\tilde{I}_f}^2 = N^{-1} \int_0^1 (f(x) - I_f)^2 \, dx = N^{-1} \sigma_f^2$$

Variance Reduction Techniques

- Monte Carlo estimates have high variance; need to reduce variance
- Antithetic Variates
  - Induce negative correlation in $f(x)$ values
  - Form the estimate

$$\hat{I}_f = \frac{1}{2N} \sum_{i=1}^{N} \left(f\left(x_i\right) + f\left(1 - x_i\right) \right)$$

  - If f is monotone, I_f has smaller variance than crude estimate
- Control Variates
  - Suppose φ is similar to f but easily integrated.
  - The identity $\int f = \int \varphi + \int (f - \varphi)$ reduces the problem to
    - a Monte Carlo integration of $\int (f - \varphi)$
    - plus the known integral $\int \varphi$
  - If $\text{cov}(f, \varphi)$ is large, variance is reduced

Variance Reduction Techniques

- Importance Sampling
  - Sample $f(x)$ where its value is most important
  - If $p(x) > 0$, and $\int_0^1 p(x)dx = 1$, then

$$I = \int_0^1 f(x)dx = \int_0^1 \frac{f(x)}{p(x)} p(x)dx$$

  - If x_i is drawn with density $p(x)$, then

$$\hat{I} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}$$

is an unbiased estimate of I, and variance of $\hat{I}$ is

$$\sigma_{\hat{I}}^2 = \frac{1}{N} \left(\int_0^1 \frac{f(x)^2}{p(x)} dx - \left(\int_0^1 f(x)dx \right)^2 \right)$$

Variance Reduction Techniques

- Importance Sampling
  - If $f(x) > 0$ and $p(x) = f(x)/\int_0^1 f(x)$, then $f(x) = Ip(x)$ and $\sigma_{\tilde{I}}^2 = 0$
  - Aim: find a $p(x)$ similar to $f(x)$
  - Thin tails problem
    - In $\sigma_{\tilde{I}}^2$ formula, key term is $f(x)^2/p(x)$
    - When $p(x)$ is small, variance is large.

Standard Optimization Methods with Simulated Objectives

- Consider optimization problem:

$$\min_{x \in U} E\{g(x, Z)\} = f(x) \tag{1}$$

  for some random variable Z.

  - For many problems of the form in (1), the objective $f(x)$ and its derivatives can be computed only with nontrivial error.
  - When solving problems of the form (1) we need to determine how well we need to approximate the integral.

- Idea: take a sample of Z of size N, and replace $E\{g(x, Z)\}$ in (1) with its sample mean $\frac{1}{N}\sum_{i=1}^{N} g\left(x, Z_i\right)$.

Standard Optimization Methods with Simulated Objectives

- For example, suppose that we want to solve

$$\min_{x\in[0,1]} E\left\{(Z-x)^2\right\} \tag{2}$$

where $Z \sim U[0,1]$.

- To solve (2), we take, say, three draws from $U[0,1]$; suppose they are $0.10, 0.73$, and 0.49. We then minimize the sample average of $(Z-x)^2$

$$\min_{x\in[0,1]} \frac{1}{3}\left((0.10-x)^2+(0.73-x)^2+(0.49-x)^2\right) \tag{3}$$

The solution to (3) is 0.43, a rough approximation of the true solution to (2) of 0.5.

Standard Optimization Methods with Simulated Objectives

- Simple portfolio problem. $u(c) = -e^{-c}$. safe asset has total return $R = 1.01$, and the risky asset has return $Z \sim N\left(\mu, \sigma^2\right)$ with $\mu = 1.06$ and $\sigma^2 = 0.04$. The portfolio problem reduces to

$$\max_{\omega} -E\left\{e^{-((1-\omega)R+\omega Z)}\right\} \tag{4}$$

- Optimal ω, denoted ω^*, and equals 1.25.
- The Monte Carlo approach to solve (4) is to use Monte Carlo integration to evaluate the integral objective.
  - Take N draws of $Z \sim N\left(\mu, \sigma^2\right)$, and replace (4) by

$$\max_{\omega} -\frac{1}{N}\sum_{i=1}^{N} e^{-((1-\omega)R+\omega Z_i)} \tag{5}$$

  - Solution to (5) is $\hat{\omega}^*$; hopefully approximates ω^*.
  - Quality of this procedure depends on N.

Exercise

Solve the following two optimization problems

- $\min_{x\in[0,1]} E\left\{(Z-x)^2\right\}$, where $Z \sim U[0,1]$ (hint: use *fmincon* command in Matlab).
- $\max_{\omega} -E\left\{e^{-((1-\omega)R+\omega Z)}\right\}$, where $R = 1.01$, $Z \sim N\left(\mu, \sigma^2\right)$ with $\mu = 1.06$ and $\sigma^2 = 0.04$ (hint: use *fminsearch* command in Matlab).

Quasi-Monte Carlo Methods

- Observation:
  - MC uses "random" sequences to satisfy i.i.d. premise of LLN
  - Integration only needs sequences which are good for integration
  - Integration does not care about i.i.d. property
- Idea of quasi-Monte Carlo methods
  - Explicitly construct a sequence designed to be good for integration
  - Do not leave integration up to mindless random choices
- Pseudorandom sequence are not random.
  - Von Neumann: "Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin."
  - Neither LLN nor CLT apply

Quasi-Monte Carlo Methods

- *All* sampling methods use deterministic sequences
- Probability theory does not apply to any practical sampling scheme
- Pseudorandom schemes seem to have $O\left(N^{-1/2}\right)$ convergence
- qMC methods have done well in some problems with hundreds of dimensions
- Pseudorandom sequences appear to have finite sample advantages for very high dimension problems

Example: Roy Model

- Structural econometric estimation starts from economic principles.
- In an economic model, some elements unobservable to the econometrician dictate an economic agent's decision.
- Roy (1951) proposes such a structural model with latent variables, and the Roy model is the foundation of self-selection in labor economics.

Example: Roy Model

- In the original paper of the Roy model, an economic agent must be either a farmer or a fisher.
- The utility of being a farmer is $U_1^* = x'\beta_1 + e_1$ and that of being a fisher is $U_2^* = x'\beta_2 + e_2$, where U_1^* and U_2^* are latent (unobservable). The econometrician observes the binary outcome $y = 1\{U_1^* > U_2^*\}$. If (e_1, e_2) is independent of x, and

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{bmatrix} \right)$$

  where σ_2 is normalized to be 1.

Example: Roy Model

- We can write down the log-likelihood as

$$L(\theta) = \sum_{i=1}^{n} \{y_i \log P(U_{i1}^* > U_{i2}^*) + (1 - y_i) \log P(U_{i1}^* \leq U_{i2}^*)\}$$

- Let $\theta = (\beta_1, \beta_2, \sigma_1, \sigma_{12})$. Given a trial value θ, we can compute

$$p(\theta | x_i) = P(U_{i1(\theta)}^* > U_{i2(\theta)}^*) = P(x_i'(\beta_1 - \beta_2) > e_{i2} - e_{i1})$$

- Under the joint normal assumption, $e_{i2} - e_{i1} \sim N(0, \sigma_1^2 - 2\sigma_{12} + 1)$ so that

$$p(\theta | x_i) = \Phi(\frac{x_i'(\beta_1 - \beta_2)}{\sqrt{\sigma_1^2 - 2\sigma_{12} + 1}})$$

Example: Roy Model

- However, notice that the analytical form depends on the joint normal assumption and cannot be easily extended to other distributions.
- As long as we know the joint distribution of (e_{i1}, e_{i2}), no matter it is normal or not, it can be generated from the computer, and we can use the stochastic method.
- We estimate

$$\hat{p}(\theta|x_i) = \frac{1}{S}\sum_{i=1}^{S} 1(U_{i1}^{s*}(\theta) > U_{i2}^{s*}(\theta))$$

where $s = 1, ..., S$ is the index of simulation and S is the total number of simulation replications.

Example: Roy Model

- Next, we match moments generated the theoretical model with their empirical counterparts.
- A set of valid choice for the Roy model example is

$$g_1(\theta) = n^{-1}\sum_{i=1}^{n}(y_i - \hat{p}(\theta|x_i)) = 0$$

$$g_2(\theta) = n^{-1}\sum_{i=1}^{n}[(y_i - \bar{y})^2 - \bar{\hat{p}}(\theta|x_i)(1 - \bar{\hat{p}}(\theta|x_i))] = 0$$

where $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ and $\bar{\hat{p}}(\theta) = n^{-1}\sum_{i=1}^{n} p(\theta|x_i)$.

Example: Roy Model

- The first set of moments matches the mean of y_i.
- The second set matches the variance of y_i.
- Since the moment conditions equals the number of unknown parameters, these moment conditions just-identifies the parameter $\theta = (\sigma_1, \sigma_{12})$.
- If we have more moment conditions, we get over-identification.

Bootstrap

- Let $X_1, X_2, ..., X_n \sim F$ be an i.i.d. sample of n observations following a distribution F. The finite sample distribution of a statistic $T_n(\theta) \sim G_n(\cdot, F)$ usually depends on the sample size n, as well as the known true distribution F. Asymptotic theory approximates $G_n(\cdot, F)$ by its limit

$$G(\cdot, F) = \lim_{n\to\infty} G_n(\cdot, F)$$

.

- Instead of referring to the limiting distribution, Bootstrap replaces the unknown distribution F in $G_n(\cdot, F)$ by a consistent estimator F_n of the true distribution, for example, the empirical distribution $\hat{F}_n(\cdot) = n^{-1} \sum_{i=1}^{n} 1\{\cdot \leq X_i\}$. Bootstrap inference is drawn from the bootstrap distribution

$$G_n^*(\cdot) = G_n(\cdot, \hat{F}_n)$$

Bootstrap

- Implementation of bootstrap is a simulation exercise.
- In an i.i.d. environment we sample over each observation with equal weight, which is called nonparametric bootstrap.
- In a dependent dataset such as time series (Chang 2004), clustering data or networks, we must adjust the sampling schedule to preserve the dependence structure.
- In regression context if we hold the independent variables fixed and only bootstrap the residual, we call it parametric bootstrap.

Bootstrap

- Bootstrap is particularly helpful in inference.
- Bootstrap can be used to conduct hypothesis testing.
- Bootstrap is useful when the analytic formula of the variance of an econometric estimator is too complex to derive or code up.
- The standard errors from the analytical expression and those from bootstrap are comparable. Both are asymptotically consistent.
- The bootstrap estimates can also be used to directly compute the confidence intervals.

Bootstrap: Example

One of the most popular estimators for a sample selection model is Heckman (1977)'s two-step method. Let the outcome equation be

$$y_i = x_i \beta + u_i$$

and the selection equation be

$$D_i = z_i \gamma + v_i$$

To obtain a point estimator, we simply run a Probit in the selection model, predict the probability of participation, and then run an OLS of y_i on x_i and $\lambda(\hat{D}_i)$ in the outcome model, where $\lambda(\cdot)$ is the inverse Mill's ratio.

Bootstrap: Example

However, as we can see from Heckman (1979)'s original paper, the asymptotic variance expression of the two-step estimator is very complicated. Instead of following the analytic formula, we can simply bootstrap the variance.

1. Generate a bootstrap sample
2. Estimate the coefficients
3. Repeat the bootstrap
4. Collect the bootstrap outcomes