Econ 673: Microeconometrics

Chapter 2: Simulation Tools for Estimation and Inference

Fall 2008

Outline

1. The Role of Simulation in Estimation and Inference
   - Monte Carlo Analysis
   - Simulation Based Integration
   - Historical Perspective
2. Generating Uncorrelated Pseudo-Random Numbers
   - Generating Uniform Pseudo-Random Numbers
   - Generating Non-Uniform Pseudo-Random Numbers
     - Inversion Method
     - Inversion Method
3. Integration by Simulation
   - Antithetic Acceleration
   - Importance Sampling
   - Monte Carlo Markov Chain Methods
   - Quasi-Monte Carlo or Low-Discrepancy Methods

Required Readings

- Judd, K L. (1998) *Numerical Methods in Economics*, Cambridge, MA: The MIT Press, Ch. 8-9.
- Stern, S., (1997) "Simulation Based Estimation," *Journal of Economic Literature*, 35: 2006-2039, Sections 1 through 3.1.
- Train, K., (2003), *Discrete Choice Methods with Simulation*, Cambridge, MA: Cambridge University Press, Ch. 9.

Monte Carlo Analysis

In a standard Monte Carlo exercise, the analyst

1. Specifies the true model:

$$y_i = f\left(X_i, \beta; \epsilon_i\right); \quad \epsilon_i \sim iidN(0, \sigma^2) \tag{1}$$

2. Simulates a series of R datasets of length N using

$$y_i^r = f\left(X_i, \beta; \epsilon_i^r\right); \quad i = 1, \ldots, N; r = 1, \ldots, R \tag{2}$$

   where ϵ_i^r are pseudo-random numbers drawn from $N(0, \sigma^2)$

3. Uses the psuedo-data sets to
   - Test the performance of alternative estimators
   - Measure the impact of model misspecification

Example

- Herriges, J. A., and C. L. Kling (1997), "The Performance of Nested Logit Models When Welfare Estimation is the Goal," *The American Journal of Agricultural Economics*, Vol. 79, No. 3, pp. 792-802.
- Focused on the impact of error misspecification in
  - Modeling recreation site choice
  - Inferring welfare impacts of changing site conditions or availability

The Role of Simulation in Estimation and Inference | Monte Carlo Analysis

Monte Carlo Analysis (cont'd)

- Key to Monte Carlo analysis is the ability to generate sequences that mimic
  - Truly random numbers drawn from the distribution of interest
  - Typically start with *iid* random variates
- Monte Carlo Analysis can also be used informally to test estimation routines

Common Integration Problems

Two basic types of problems arise:

$$E\left[g(x)\right] = \int g(x)f(x)dx$$

$$\begin{aligned} P\left[x \in D\right] &= \int_D f(x)dx \\ &= \int 1(x \in D)f(x)dx \end{aligned}$$

Numerical Approximations

- Methods rely on variants of Riemann sums:

$$\int g(x)f(x)dx \approx \sum_{i=1}^{R} \left[g(x_i)f(x_i)\right](a_i - a_{i-1})$$

  where $\{a_0 < a_1 < \cdots a_R\}$ partitions the domain of $g(x)$ and $x_i \in \left[a_{i-1}, a_i\right]$.
- There are several potential problems with such approximations:
  - Their accuracy depends on the number of partitions R and the spacing used.
  - They become computationally infeasible for high dimensional integrals (> 4).

The Starting Point for Simulation Integration

If $E[g(x)] < \infty$ and one can draw an *iid* sequence from the distribution of x; i.e.,

$$x^r \sim F(x), r = 1, \ldots, R \tag{3}$$

then by the strong law of large numbers

$$g_R \equiv \frac{1}{R}\sum_{r=1}^{R} g(x^r) \xrightarrow{a.s.} E[g(x)] \tag{4}$$

i.e.,

$$Pr\left(\lim_{R\to\infty} g_R = E[g(x)]\right) = 1 \tag{5}$$

and

$$E[g_R] = E[g(x)] \tag{6}$$

The Starting Point for Simulation Integration

If $Var[g(x)] = \Xi < \infty$ then the Lindberg-Levy Central Limit Theorem yields

$$\sqrt{R}\left(g_R - E[g(x)]\right) \xrightarrow{d} N(0, \Xi) \tag{7}$$

This result can be used to assess the precision of the simulation, replacing Ξ with the sample covariance:

$$\hat{\Xi} = \frac{1}{R}\sum_{r=1}^{R} [g(x^r) - g_R]\,[g(x^r) - g_R]' \tag{8}$$

Many of the Simulation Developments Emerged in

- Bayesian Econometrics/Statistics
- Modeling Limited Dependent Variables
- Stochastic Dynamic Programming Models

Bayesian Analysis

- Bayes Theorem provides that

$$p(\theta|y, X) \propto p(\theta)p(y|\theta, X) \tag{9}$$

- Traditionally hampered by the need to rely on *conjugate* priors, yielding tractable posteriors.
- But these priors did not always reflect one's actual priors
- Numerical methods virtually eliminate these restrictions
  - Dorfman, J. H. (1997) *Bayesian Economics Through Numerical Methods*, New York: Springer-Verlag.
  - Gelman, A., J. Carlin, H.S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, London: Chapman & Hall.
  - Geweke, J., "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57(1989a): 1317-1339.
  - Koop, G., D.J. Poirier and J.L. Tobias (2007). *Bayesian Econometric Methods*, Cambridge, Cambridge University Press.

Limited Dependent Variable Models

- Example: The classic RUM model, where individual i's utility from choosing alternative j is given by:

$$U_{ij} = X_{ij}\beta + u_{ij}$$

- The probability of choosing alternative 1 is given by

$$P_{i1} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1\left(u_{ij}^1 < X_{ij}^1, \forall j \neq 1\right) dF\left(u_{ij}^1\right)$$

where

$$u_{ij}^1 \equiv u_{ij} - u_{i1}$$

and

$$X_{ij}^1 \equiv X_{i1} - X_{ij}$$

LDV (cont'd)

- Numerical quadrature methods for integration typically feasible for integrals of dimension 4 or less.
- Analysts have relied instead on convenient, though not necessarily realistic, choices for $F(\cdot)$ yielding
  - Multinomial logit
  - Nested logit
- Simulation integration techniques allow for more realistic choices for $F(\cdot)$
  - Multinomial Probit
  - Mixed Logit

LDV Sources

- Hajivassiliou, V. A., and P. A. Ruud, "Classical Estimation Methods for LDV Models Using Simulation," in R. F. Engle and D. L. McFadden (Eds.) *Handbook of Econometrics*, Volume IV, Amsterdam: Elsevier, 1994, pp. 2383-2441.
- Hajivassiliou, V. A., "Simulation Estimation Methods for Limited Dependent Variable Models, in G. S. Maddala, R. Rao, and H. D. Vinod (Eds.) *Handbook of Statistics*, Volume 11, Amsterdam: Elsevier, 1993, pp. 519-543.

Computation of confidence bounds for $g(\hat{\beta})$

Alternatives

1. Taylor series expansion:

$$g\left(\hat{\beta}\right) \approx g(\beta) + \left(\beta - \hat{\beta}\right)' \frac{\partial g(\beta)}{\partial \beta} \tag{10}$$

$$\begin{aligned} \Rightarrow Var\left[g\left(\hat{\beta}\right)\right] &\approx \left[\frac{\partial g(\beta)}{\partial \beta}\right] Var\left[\hat{\beta}\right] \left[\frac{\partial g(\beta)}{\partial \beta}\right]' \\ &\approx \left[\frac{\partial g(\hat{\beta})}{\partial \beta}\right] Var\left[\hat{\beta}\right] \left[\frac{\partial g(\hat{\beta})}{\partial \beta}\right]' \end{aligned}$$

   - requires computation of derivatives
   - relies on a first order approximation

Computation of confidence bounds for $g(\hat{\beta})$ (cont'd)

Alternatives

2. Bootstrapping:
   - Treats current sample as representative of the population
   - Generate "pseudo-samples" by drawing with replacement from sample
   - Re-estimating the model with each pseudo-sample generates sampling distribution for $g(\hat{\beta})$
   - Drawback is that bootstrapping is computationally intensive.
3. Simulation
   - Draw $\beta^r \sim N(\hat{\beta}, \hat{\Sigma})$
   - Generate $g(\hat{\beta}^r)$, from which various moments and confidence intervals can be constructed.

Generating Uncorrelated Pseudo-Random Numbers

- Constructing Uniform variates
  - Ideal is to generate a sequence of iid $U[0,1]$, from which more complex random variables can be generated.
  - In fact, pseudo-random sequences are used; i.e., deterministic sequences mimicking random variables:
    - Zero serial correlation (at all lags)
    - The correct frequency of sequential patterns (e.g., runs <0.1)
- Sources:
  - Geweke, J., "Monte Carlo Simulation and Numerical Integration," in H. M. Amman, D. A Kendrick, and J. Rust (Eds.) *Handbook of Computational Economics Volume I*, Amsterdam: Elsevier, 1996, pp. 731-800.
  - Judd, K L. *Numerical Methods in Economics*, Cambridge, MA.: The MIT Press, 1998, Ch. 8.

Uniform Pseudo-Random Numbers

- A simple pseudo-random number generator is the linear congruential generator:
$$J_{r+1} = (aJ_r + c) \bmod_m \tag{11}$$
where a, c and m are constants.
- These generators cycle and, if m is prime, m is the cycle length.
- With a 32-bit machine, $m = 2^{31} - 1 = 2,147,483,647$ is usually used.
- Uniform pseudo-variates are then formed using:
$$u_r = \frac{J_r}{m} \tag{12}$$

Alternative Generators

- The *shuffle* generator uses a table of m seeds that are randomly used and updated.
- Judd(1998) suggests, beginning with 55 odd numbers, using
$$J_r = (J_{r-24} \cdot J_{r-55}) \bmod_{2^{32}} \tag{13}$$
This sequence has a period length of 10^{23} and passes many tests of randomness.

Suggested Practices - Geweke (1996, p. 745)

- Be sure pseudo-random number generator is documented.
- Don't skimp on this step – it is usually a small part of overall computational effort.
- Be sure cycle length is sufficient for problem at hand.
- If you are unsure about the generator, check for sensitivity of results to alternative generators.

Constructing Non-Uniform Variates: Inversion Method

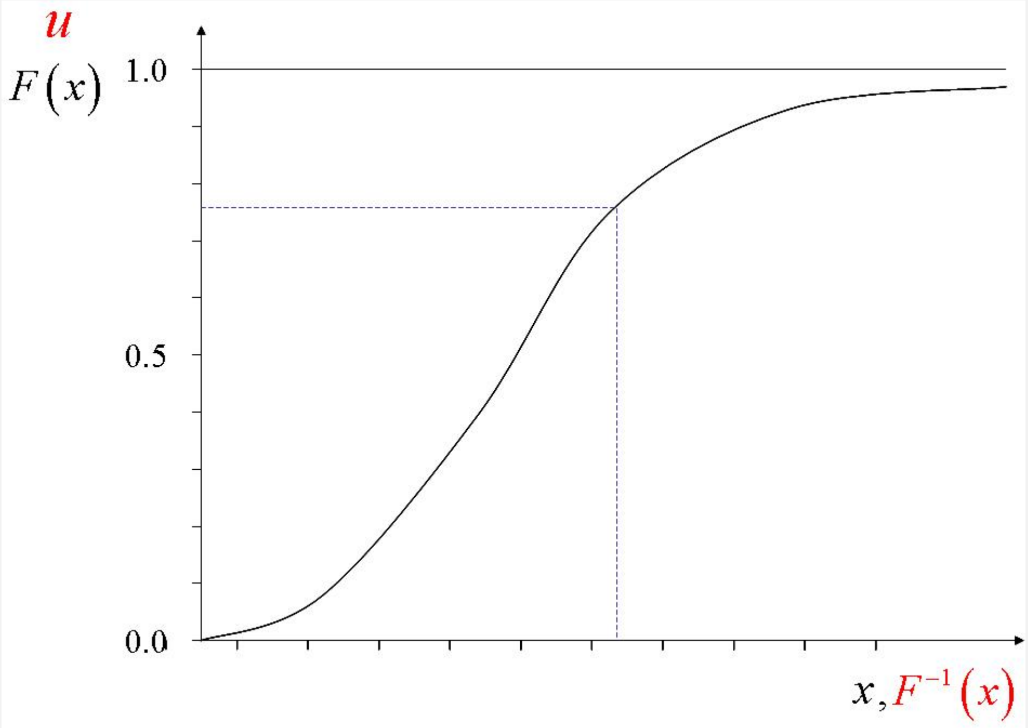
- If $u \sim U[0,1]$ and $F(x)$ is a continuous distribution, then $F^{-1}(u)$ has cumulative distribution function $F(x)$ if $F^{-1}(u)$ exists.

$$P\left[F^{-1}(u) \leq d\right] = P\left[u \leq F(x)\right] = F(d) \tag{14}$$

- If $F^{-1}(u)$ does not exist, then

$$x_r = \min_x \{x | F(x) \geq u_r\} \tag{15}$$

represents a draw from the distribution function $F(x)$.



Inversion Method: Extreme Value

Given

$$F(x) = exp\left[-exp(-x)\right] \tag{16}$$

then

$$x_r = F^{-1}(u_r) = -ln\left[-ln(u_r)\right] \sim F(x) \tag{17}$$

if

$$u_r \sim U[0,1] \tag{18}$$

Inversion Method: Truncated Normal (Version 1)

If $x \sim N(\mu, \sigma^2)$, then a truncated version of x, with truncation points (a, b) (sometimes denotes $y \sim TN_{(a,b)}(\mu, \sigma^2)$ has cdf:

$$F(y) = \frac{\left[\Phi\left(\frac{y-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]}{\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]} \tag{19}$$

then

$$y_r = \sigma\Phi^{-1}\left\{u_r\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] + \Phi\left(\frac{a-\mu}{\sigma}\right)\right\} + \mu \tag{20}$$

if

$$u_r \sim U[0, 1] \tag{21}$$

Inversion Method: Truncated Normal (Version 2)

If $y \sim TN_{(a,b)}(\mu, \sigma^2)$, then

$$y_r = \sigma\Phi^{-1}(\tilde{u}_r) + \mu \tag{22}$$

if

$$\tilde{u}_r \sim U\left[\Phi\left(\frac{a-\mu}{\sigma}\right), \Phi\left(\frac{b-\mu}{\sigma}\right)\right] \tag{23}$$

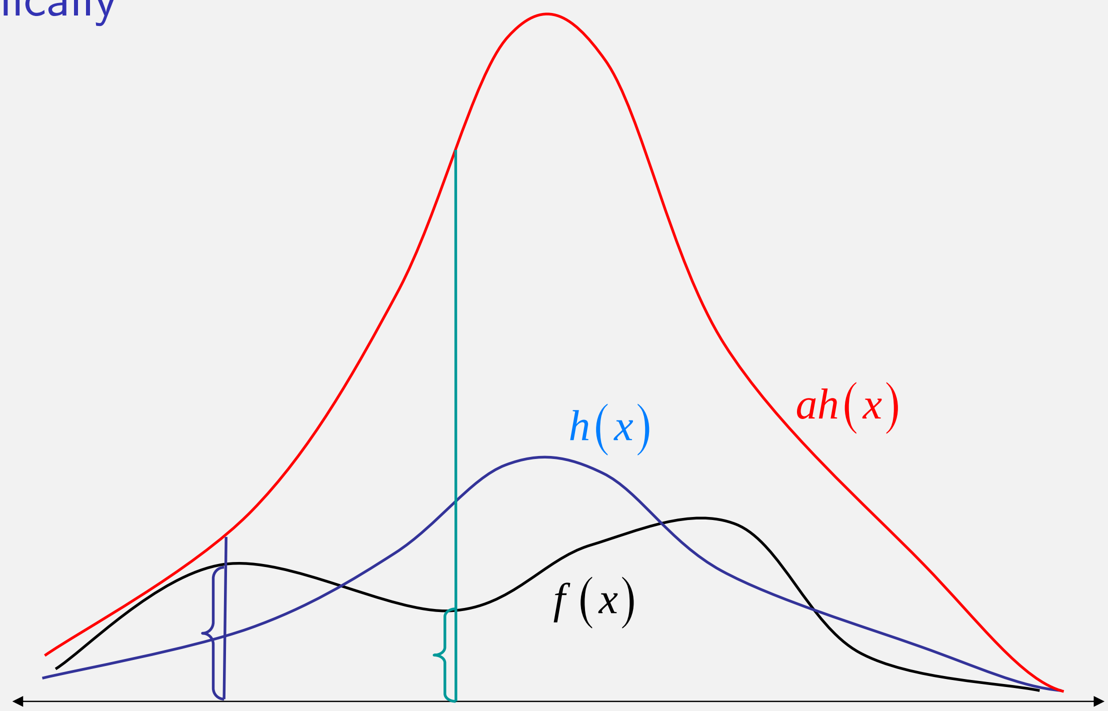
Inversion Method Limitations

- Need inverse function $F^{-1}(\cdot)$ or an efficient approximation.
- For some distributions, faster and more precise algorithms exist
  Example: Geweke's (1991) mixed rejection algorithm for the truncated univariate normal
  - 2.47 to 6.24 times faster than inversion algorithm
  - Avoids inversion problems in tails
- Limited to univariate distributions

Acceptance/Rejection Method

- Suppose that x has a continuous pdf $f(x)$ and support on D
- Let $h(x)$ be an alternative pdf such that
  - one can readily draw *iid* variates from this distribution
  - and
    $$\sup_{x \in D}\left[\frac{f(x)}{h(x)}\right] = a < \infty \tag{24}$$
    where a is a known constant with $a \geq 1$ if $h(\cdot)$ is proper.
- To generate *iid* variates from the distribution $F(x)$:
  1. Generate $u^r \sim U[0,1]$ and $z^r \sim H(z)$
  2. If
     $$u^r \leq \frac{f(z^r)}{ah(z^r)} \tag{25}$$
     then set $x^r = z^r$; otherwise return to step 1.

Graphically



Geweke (1996, p. 747) Notes that:

- The unconditional probability of accepting an outcome is:

$$\int_{-\infty}^{\infty} \left[\frac{f(x)}{ah(x)} \right] h(x)dx = \frac{1}{a}. \tag{26}$$

- The unconditional probability of accepting an outcome with a value of at most c is

$$\int_{-\infty}^{c} \left[\frac{f(x)}{ah(x)} \right] h(x)dx = \frac{F(c)}{a}. \tag{27}$$

- From this, one can deduce that the probability of an outcome with a value of at most c is $F(c)$, which is, of course, what we want.

AR Method (cont'd)

- Advantages:
  - Can cope with an arbitrary pdf
  - Can easily monitor efficiency in terms of %accepted
- Disadvantages:
  - Can be inefficient if a is large
  - Limited use for multivariate distributions

Drawing from Discrete Distributions

- Essentially an extension of the Inverse Method to a step function
- If the distribution has support on the ordered set $\{x_1, x_2, \ldots\}$ with $p_i = Pr(x = x_i)$ and

$$\sum_{i=1}^{\infty} p_i = 1 \tag{28}$$

then the right continuous distribution function for x is

$$F(x) = \sum_{i=1}^{\infty} p_i 1(x_i \leq x) \tag{29}$$

- One can then drawing x^r using $u_r \sim U[0,1]$ with $x_r = x_i$ if $F(x_{i-1}) < u^r \leq F(x_i)$ where $F(x_0) = -0.01$

Special Cases

- If $u_1 \sim U[0,1]$ and $u_2 \sim U[0,1]$ are mutually independent, then

$$\begin{aligned} x_1 &= cos(2\pi u_1)\sqrt{-2log(u_2)} \sim N(0,1) \\ x_2 &= cos(2\pi u_2)\sqrt{-2log(u_1)} \sim N(0,1) \end{aligned}$$

  but x_1 and x_2 will be correlated.
- If $x_k^r \sim iidN(0,1); k = 1, \ldots, K; r = 1, \ldots, R$, then

$$\begin{aligned} y_k^r &= \mu + \sigma x_k^r \sim iidN(\mu, \sigma^2) \\ z^r &= \sum_{k=1}^{K_1} (x_k^r)^2 \sim \chi^2_{K_1} \end{aligned}$$

- Similar transformations exist for Student t, F, and Beta variates.
- Generalizations exist for multivariate Normal and Wishart variates.

Integration by Simulation

- In Monte Carlo Analysis, the key is generating a sequence of *pseudo-random* numbers that are
  - serially uncorrelated
  - random (or at least mimic randomness)
- For the purposes of integration, neither of these characteristics is critical and adherence to these ideals may be counterproductive.

Crude Monte Carlo Integration

If $x^r \sim F(x)$, then

$$g_R^{CMC} \equiv \frac{1}{R}\sum_{r=1}^{R} g(x^r) \xrightarrow{a.s.} \int g(x)f(x)dx = E\left[g(x)\right] \tag{30}$$

If $\sigma_g^2 \equiv Var\left[g(x)\right]$ then by the central limit theorem

$$\sqrt{R}\left(g_R^{CMC} - g\right) \xrightarrow{d} N(0, \sigma_g^2) \tag{31}$$

So that

$$Var\left[g_R^{CMC}\right] = \frac{\sigma_g^2}{R} \rightarrow 0 \text{ as } R \rightarrow \infty \tag{32}$$

Antithetic Acceleration

- The risk with CMC is that a given draw will be unusually large, not necessarily offset by subsequent draws.
- Antithetic acceleration relies on correlated draws to reduce simulation variance.

Antithetic Acceleration (cont'd)

- Consider again estimating $E\left[g(x)\right]$ where $x \sim F(x)$.
- Suppose that $F(x)$ is invertable, so that

$$g_R^{CMC} \equiv \frac{1}{R}\sum_{r=1}^{R} g(x^r) = \frac{1}{R}\sum_{r=1}^{R} g\left(F^{-1}\left[u^r\right]\right) \tag{33}$$

where $u^r \sim U[0,1]$.
- Since $\tilde{u}^r \equiv 1 - u^r \sim U[0,1]$, we can alternatively compute

$$\tilde{g}_R^{CMC} \equiv \frac{1}{R}\sum_{r=1}^{R} g(\tilde{x}^r) = \frac{1}{R}\sum_{r=1}^{R} g\left(F^{-1}\left[\tilde{u}^r\right]\right) \tag{34}$$

Antithetic Acceleration (cont'd)

- The antithetic simulator in this case is simply the average of these two CMC simulators; i.e.,

$$g_R^{AAS} \equiv \frac{1}{2}\left(g_R^{CMC} + \tilde{g}_R^{CMC}\right) = \frac{1}{2R}\sum_{r=1}^{R}\left\{g\left(F^{-1}\left[u^r\right]\right) + g\left(F^{-1}\left[1-u^r\right]\right)\right\} \tag{35}$$

- The key to the AA simulator is that if $x^r = F^{-1}(u^r)$ is large, then $\tilde{x}^r = F^{-1}(1-u^r)$ will be small.

The Variance of the AA Simulator

$$
\begin{aligned}
Var\left[g_R^{AAS}\right] &= \frac{\sigma_g^2 + Cov\left\{g\left(F^{-1}\left[u^r\right]\right), g\left(F^{-1}\left[1-u^r\right]\right)\right\}}{2R} \\
&= \frac{1}{2} Var\left[g_R^{CMC}\right] + \frac{Cov\left\{g\left(F^{-1}\left[u^r\right]\right), g\left(F^{-1}\left[1-u^r\right]\right)\right\}}{2R}
\end{aligned}
$$

As a result, the above covariance term will be negative if $g(x)$ is monotonic.

An Alternative AA Simulator

- The previous AA simulator relied on inversion, which is not always possible.
- An alternative AA simulator pivots the simulator around the mean of x; i.e.,

$$
g_R^{AAS} = \frac{1}{2}\left(g_R^{CMC} + \check{g}_R^{CMC}\right) = \frac{1}{2R}\sum_{r=1}^{R}\left\{g\left(x^r\right) + g\left(\check{x}^r\right)\right\} \tag{36}
$$

where

$$
\check{x}^r = E(x) - \left[x^r - E(x)\right] = 2E(x) - x^r \tag{37}
$$

Importance Sampling

- The CMC and AA simulators require draws from $F(x)$, which may be difficult (particularly for Bayesian analysis).
- Importance Sampling relies on a distribution $h(x)$
- The steps are as follows:
  1. Draw x^r from $h(x)$.
  2. Compute

$$
\begin{aligned}
g_R^{IS} &= \frac{1}{R}\sum_{r=1}^{R}\frac{f(x^r)}{h(x^r)}g(x^r) \\
&= \frac{1}{R}\sum_{r=1}^{R}w(x^r)g(x^r)
\end{aligned}
$$

where

$$w(x^r) \equiv \frac{f(x^r)}{h(x^r)} \tag{38}$$

IS Integration - Intuition

- The basic idea of IS intergration comes from the fact that

$$E\left[g(x)\right] = \int g(x)f(x)dx = \int \frac{g(x)f(x)}{h(x)}h(x)dx \tag{39}$$

- The sampling distribution $h(x)$ should be chosen so that
  - It is easy to draw from
  - f and h have the same support
  - It is easy to evaluate $\frac{g(x)f(x)}{h(x)}$
  - $w(x)g(x) = \frac{g(x)f(x)}{h(x)}$ is bounded and smooth over the support of x. Ideally $w(x)g(x)$ is a constant.

ISS - Notes

- It is good practice to keep track of the largest weights.
  Large weights suggest that there are regions of $f(x)$ that are poorly represented by $h(x)$
- Problems arise when $f(x)$ has fatter tails than $h(x)$. It is common practice to inflate the variance of $h(x)$ to avoid this problem.
- Unlike the AR approach, all of the draws from $h(x)$ are used.

The GHK Simulator

- A particularly useful version of the ISS is the Geweke-Hajivassiliou-Keane (GHK) simulator for multivariate normal probabilities.
- The key insight is to split the joint pdf into a series of conveniently simulated conditional probabilities.
- Consider computing:

$$Pr\left[\mathbf{x} \in D\right] = Pr\left[x_k \in (a_k, b_k), k = 1, \ldots K\right] \tag{40}$$

where $\mathbf{x} \sim N(\mu, \Omega)$

The GHK Simulator (cont'd)

The joint probability can be written as:

$$\begin{aligned} Pr[\mathbf{x} \in D] &= Pr[x_k \in (a_k, b_k), k = 1, \dots K] \\ &= Pr[x_1 \in (a_1, b_1)]\, Pr[x_2 \in (a_2, b_2) | x_1 \in (a_1, b_1)] \\ &\cdots\ Pr[x_K \in (a_K, b_K) | x_j \in (a_j, b_j), j = 1, \dots K-1] \\ &= \prod_{k=1}^{K} Pr[x_k \in (a_k, b_k) | x_j \in (a_j, b_j) \forall j < k] \\ &= \prod_{k=1}^{K} p_{k|<k} \end{aligned}$$

The GHK Simulator (cont'd)

Each of these entries:

$$p_{k|<k} = Pr[x_k \in (a_k, b_k) | x_j \in (a_j, b_j) \forall j < k] \tag{41}$$

is a simple truncated normal probability, since

$$x_k | (x_1, \dots, x_{k-1}) = x_k | \mathbf{x}_{<k} \sim N\left(\mu_{k|<k}[\mathbf{x}_{<k}], \Omega_{k|<k}\right) \tag{42}$$

where the conditional mean is

$$\mu_{k|<k}[\mathbf{x}_{<k}] = \mu_k + \Omega_{k,<k}\Omega^{-1}_{<k,<k}(\mathbf{x}_{<k} - \boldsymbol{\mu}_{<k}) \tag{43}$$

and the conditional variance is

$$\Omega_{k|<k} = \Omega_{kk} - \Omega_{k,<k}\Omega^{-1}_{<k,<k}\Omega_{<k,k} \tag{44}$$

Truncated Normal Variates

Recall that if $u^r \sim U[0,1]$, then

$$y_r = \sigma\Phi^{-1}\left\{u_r\left[\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right]+\Phi\left(\frac{a-\mu}{\sigma}\right)\right\}+\mu \quad (45)$$

is drawn from a truncated normal distribution.

Steps in the GHK Simulator

For $r = 1, \ldots, R$

1. Set $k = 1$ and compute

$$p_1^r = p_1 = Pr\left[x_1 \in (a_1, b_1)\right] = \Phi\left(\frac{b_1-\mu_1}{\Omega_{11}}\right) - \Phi\left(\frac{a_1-\mu_1}{\Omega_{11}}\right) \quad (46)$$

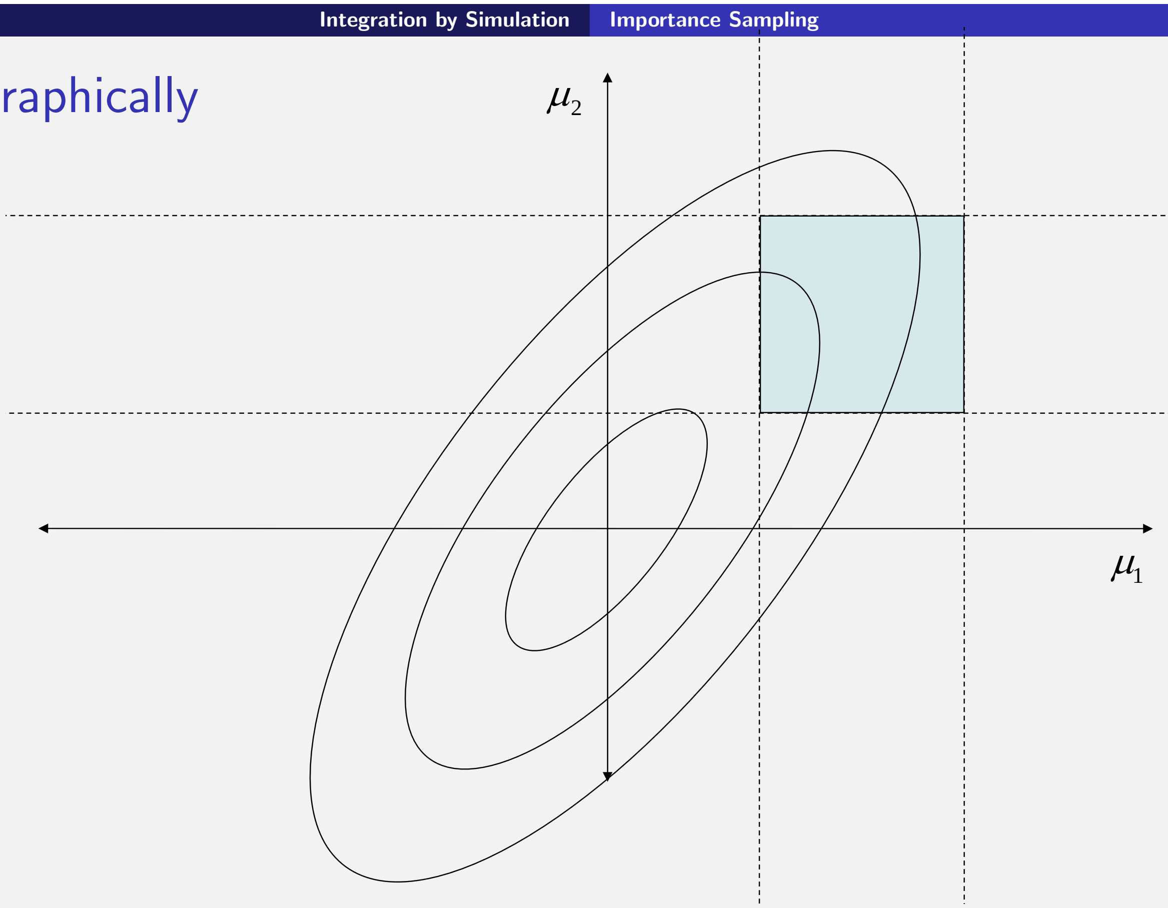
2. Draw x_1^r from the truncated distribution with mean μ_1 and variance Ω_{11}.
3. Increment k by 1 and compute

$$\begin{aligned} p_{k|<k}^r &= Pr\left[x_k \in (a_k, b_k) | \mathbf{x}_{<k}\right] \\ &= \Phi\left(\frac{b_k - \mu_{k|<k}(\mathbf{x}_{<k})}{\Omega_{k|<k}}\right) - \Phi\left(\frac{a_k - \mu_{k|<k}(\mathbf{x}_{<k})}{\Omega_{k|<k}}\right) \end{aligned}$$

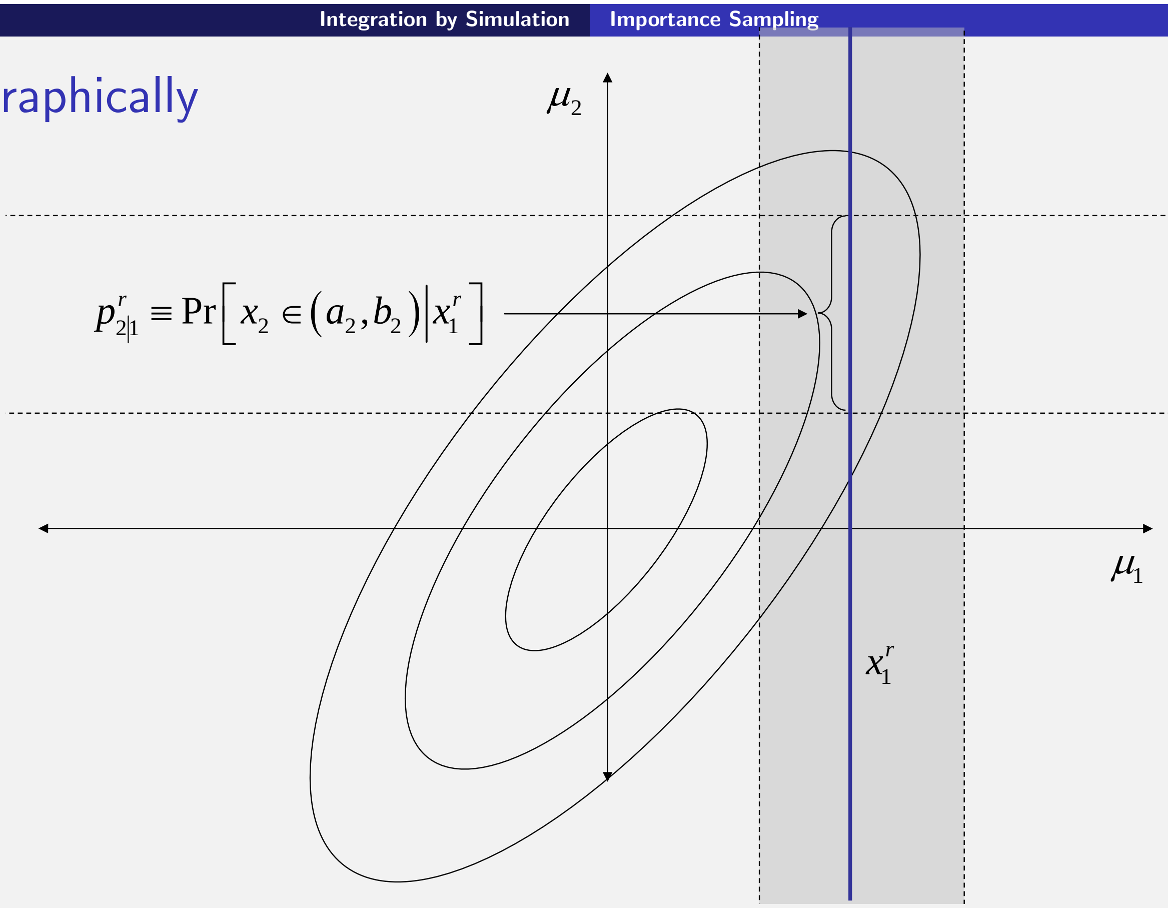
Steps in the GHK Simulator

4. If $k < K$, draw $x^r_{k|<k}$ from the truncated normal distribution with mean $\mu_{k|<k}(\mathbf{x}_{<k})$ and variance $\Omega_{k|<k}$ and return to step 3; otherwise go to Step 5.
5. Compute
$$p^r = \prod_{k=1}^{K} p^r_{k|<k} \tag{47}$$
6. Repeat Steps 1 through 5 from $r = 1, \ldots, R$ and compute
$$p_R^{GHK} = \frac{1}{R}\sum_{r=1}^{R} p^r \tag{48}$$

Graphically



Graphically



Advantages of the GHK Simulator

- Each step of the simulator is easy to compute
- $p_R^{GHK} \in (0, 1)$; i.e., it is strictly bounded between zero and one.
- The variance of p_R^{GHK} is smaller than the CMC simulator

$$p_R^{CMC} = \frac{1}{R} \sum_{r=1}^{R} 1(\mathbf{x}^r \in D) \tag{49}$$

  where $\mathbf{x}^r \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$, since $p^r \in (0, 1)$ has less variance than $1(\mathbf{x}^r \in D)$.
- p_R^{GHK} is continuous and differentiable w.r.t. the parameters.
- Hajivassiliou, McFadden, and Ruud (1992) found the GHK simulator to generally outperform 12 other simulators.

Monte Carlo Markov Chain Methods

- MCMC provides a method of simulator complex nonstandard multivariate distributions.
- Basic steps to drawing K sequences $x_k^r; k = 1, \ldots, K; r = 1, \ldots, R$ from the target distribution $F(x)$:
    1. Choose starting points x_k^0, typically spread out.
    2. For each r, x_k^r is chosen from a transition distribution $T^r\left(x_k^r | x_k^{r-1}\right)$, where this distribution is chosen such that $x_k^r \xrightarrow{d} F(x)$ as $r \to \infty$
    3. Construct

$$g_{R,K}^{MCMC} = \frac{1}{K(R - R_b)} \sum_{k=1}^{K} \sum_{r=R_b+1}^{R} g(x_k^r) \tag{50}$$

Notes on MCMC

- Clearly, the individual elements of the sequences are not uncorrelated.
- Multiple sequences can be used to monitor convergence.
- The initial R_b elements in each sequence are typically dropped to reduce the dependence on the starting values.
- Sources:

    - Gelman, A., J. Carlin, H.S. Stern, and D. B. Rubin (1995), *Bayesian Data Analysis*, London: Chapman & Hall.
    - Carlin, B. P, and T. A. Louis (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall.
    - Koop, G., D.J. Poirier and J.L. Tobias (2007). *Bayesian Econometric Methods*, Cambridge, Cambridge University Press.

The Metropolis Algorithm

1. Choose starting points x_k^0
2. For each r, $\tilde{x}_k^r$ is chosen from the *jumping distribution* $J^r\left(\tilde{x}_k^r | x_k^{r-1}\right)$, where $J^r\left(a|b\right) = J^r\left(b|a\right)$
3. Set
$$x_k^r = \begin{cases} \tilde{x}_k^r & \text{with probability } min(\theta, 1) \\ x_k^{r-1} & \text{otherwise} \end{cases}$$
where
$$\theta = \frac{f(\tilde{x}_k^r)}{f(x_k^{r-1})} \tag{51}$$
4. Construct
$$g_{R,K}^{MS} = \frac{1}{K(R-R_b)} \sum_{k=1}^{K} \sum_{r=R_b+1}^{R} g(x_k^r) \tag{52}$$

The Metropolis-Hastings Algorithm

- Relaxes the restriction that the jumping distribution has to be symmetric.
- Replaces θ with
$$\theta = \frac{f(\tilde{x}_k^r)/J^r\left(\tilde{x}_k^r | x_k^{r-1}\right)}{f(x_k^{r-1})/J^r\left(x_k^{r-1} | \tilde{x}_k^r\right)} \tag{53}$$

Quasi-Monte Carlo or Low-Discrepancy Methods

- "Quasi-Monte Carlo Methods are sampling methods that do not rely on probabilistic ideas and pseudorandom sequences..." Judd(1998, p. 310)
- Draws primarily on number theory.
- Sources:
  - Judd, K L. (1998) *Numerical Methods in Economics*, Cambridge, MA: The MIT Press, Ch. 9.
  - Geweke, J., "Monte Carlo Simulation and Numerical Integration," in H. M. Amman, D. A. Kendrick, and J. Rust (Eds.), *Handbook of Computational Economics Volume I*, Amsterdam: Elsevier, 1996, pp. 731-800.

Fundamental Theorem - Hlawka

Under certain regularity conditions

$$\left| \frac{1}{N} \sum_{j=1}^{N} f(x^j) - \int_{I^d} f(x) dx \right| \leq V^{HK}(f) D_N^* \tag{54}$$

where, intuitively,

- $V^{HK}(f)$ measures the overall variability of the function f
- D_N^* is the *discrepancy* of the sequence $x^j, j = 1, \ldots, N$, which is small if the sequence evenly fills up the space D

Halton Sequences

- Halton sequence is one of a number of *low discrepancy sequences*, with D_N^* of the order $N^{-1}(lnN)^d$.
- The expected error for a simple *Monte Carlo* integration is of the order $N^{-1/2}$.
- Relatively, we have

$$\frac{D_N^*(Halton)}{D_N^*(MC)} = \frac{N^{-1}(lnN)^d}{N^{-1/2}} \tag{55}$$

- Geweke (1996) argues that low discrepancy methods outperform standard Monte Carlo simulation methods for $d < 8$ or 9.

Halton Sequences

- Halton sequences are among the available low discrepancy sequences.
- Sequences are based off of
  - a starting integer, m, usually chosen to be prime
  - a starting sequence, usually $s_0 = \{0\}$.
  - The sequence is built up iteratively using

$$s_{t+1} = \left\{ s_t, s_t + \frac{1}{m}, s_t + \frac{2}{m}, \ldots, s_t + \frac{m-1}{m} \right\} \tag{56}$$

Halton Sequence Example: m=3

starting with $m = 3$ and $s_0 = \{0\}$, we have
$s_{t+1} = \left\{s_t, s_t + \frac{1}{m}, s_t + \frac{2}{m}, \ldots, s_t + \frac{m-1}{m}\right\}$ so that

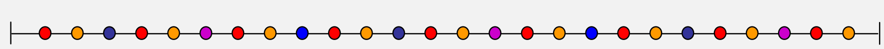
$$s_1 = \left\{0, 0 + \frac{1}{3}, 0 + \frac{2}{3}\right\} = \left\{0, \frac{1}{3}, \frac{2}{3}\right\} \tag{57}$$

$$\begin{aligned} s_2 &= \left\{s_1, s_1 + \frac{1}{3^2}, s_1 + \frac{2}{3^2}\right\} = \left\{ \begin{array}{l} \left\{0, \frac{1}{3}, \frac{2}{3}\right\} \\ \left\{0, \frac{1}{3}, \frac{2}{3}\right\} + \frac{1}{9} \\ \left\{0, \frac{1}{3}, \frac{2}{3}\right\} + \frac{2}{9} \end{array} \right. \\ &= \left\{ \begin{array}{l} \left\{0, \frac{1}{3}, \frac{2}{3}\right\} \\ \left\{\frac{1}{9}, \frac{4}{9}, \frac{7}{9}\right\} \\ \left\{\frac{2}{9}, \frac{5}{9}, \frac{8}{9}\right\} \end{array} \right. = \left\{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}\right\} \end{aligned}$$

$$s_3 = \left\{s_2, s_2 + \frac{1}{3^3}, s_1 + \frac{2}{3^3}\right\} = \left\{0, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{4}{9}, \frac{7}{9}, \frac{2}{9}, \frac{5}{9}, \frac{8}{9}, \frac{1}{27}, \frac{10}{27}, \frac{19}{27}, \ldots\right\} \tag{58}$$

Halton Sequence Example: m=3

Visually, the Halton sequences are systematically filling up the unit interval



Halton Sequences

Notice that

- The initial sequence, s_0, is dropped
- Each cycle of the sequence covers areas not covered by previous cycles
- Draws from one observation tend to be negatively correlated with previous draws
    - but this property diminishes with the number of draws for an observation
    - creates a tradeoff between coverage and antithetics