Chapter 10 Simulation

10.1 Introduction

- How can we determine the probability of our winning a game of solitaire?
- One possible approach is to start with the reasonable hypothesis that all (52)! possible arrangements of the deck of cards are equally likely to occur and then attempt to determine how many of these lead to a win.
- Determination of the win probability for solitaire is mathematically intractable.
- All is not lost, for probability falls not only within the realm of mathematics, but also within the realm of applied science, and as in all applied sciences, experimentation is a valuable technique.
- Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th game results in a win} \\ 0 & \text{otherwise} \end{cases}$$

then $X_i, \quad i = 1, \ldots, n$ will be independent Bernoulli random variables for which

$$E[X_i] = P\{\text{win at solitaire}\}$$

by **the strong law of large numbers**:

$$\sum_{i=1}^{n} \frac{X_i}{n} = \frac{\text{number of games won}}{\text{number of games played}} \longrightarrow P\{\text{win at solitaire}\}$$

- By playing a large number of games we can use the proportion of won games as an estimate of the probability of winning.
- This method of empirically determining probabilities by means of experimentation is known as **simulation**.
- To initiate a simulation study: generate the value of a uniform (0,1) random variable.
- Most random number generators start with an initial value X_0, called the seed.
- Recursively compute values by specifying positive integers $a, c,$ and m, and letting

$$X_{n+1} = (aX_n + c) \text{ modulo } m \quad n \geq 0$$

- the foregoing means that $aX_n + c$ is divided by m and the reminder is taken as the value of X_{n+1}.
- each X_{n+1} is either $0, 1, \ldots, m-1$.
- the quantity X_n/m is taken as an approximation to a uniform (0,1) random variable.

- Suppose that we can simulate from uniform $(0, 1)$ distribution .
- **Random numbers**: Independent random variables from uniform $(0, 1)$.
- In solitaire example we would need to program a computer to play out the game starting with a given ordering of the cards.
  - Begin by randomly choosing one of the elements and then putting it in position n.
  - It then randomly chooses among the remaining elements and put the choice in position $n-1$; and so on.

Example 10.1a. *Generating a random permutation.* Suppose we are interested in generating a permutation of the integers $1, 2, \ldots, n$ that is such that all $n!$ possible orderings are equally likely. Starting with any initial permutation we will accomplish this after $n-1$ steps where at each step we will interchange the positions of two of the numbers of the permutation. Throughput, we keep track of the permutation by letting $X(i)$, $i = 1, \ldots, n$ denote the number currently in position i. The algorithm operates as follows:

1. Consider any arbitrary permutation and let $X(i)$ denote the element in position i, $i = 1, \ldots, n$. [For instance, we could take $X(i) = i$, $i = 1, \ldots, n$]
2. Generate a random variable N_n that is equally likely to equal any of the values $1, 2, \ldots, n$.
3. Interchange the values of $X(N_n)$ and $X(n)$. The value of $X(n)$ will now remain fixed.

[For instance, suppose that $n = 4$ and initially $X(i) = i$, $i = 1, 2, 3, 4$. If $N_4 = 3$, then the now permutation is $X(1) = 1$, $X(2) = 2$, $X(3) = 4$, $X(4) = 3$, and element 3 will remain in position 4 throughout.]

4. Generate a random variable $N_n - 1$ that is equally likely to be either $1, 2, \ldots, n-1$.
5. Interchange the values of $X(N_n - 1)$ and $X(n-1)$. [If $N_3 = 1$, then the new permutation is $X(1) = 4$, $X(2) = 2$, $X(3) = 1$, $X(4) = 3$.]
6. Generate N_{n-2}, which is equally likely to be either $1, 2, \ldots, n-2$.
7. Interchange the values of $X(N_{n-2})$ and $X(2)$. [If $N_2 = 1$, then the new permutation is $X(1) = 2$, $X(2) = 4$, $X(3) = 1$, $X(4) = 3$ and this is the final permutation.]
8. Generate N_{n-3}, and so on. The algorithm continuous until N_2 is generated and after

the next interchange the resulting permutation is the final one.

- To implement this algorithm:
  - Let U denote a random number which is uniformly distributed on (0,1).
  - kU is uniform on $(0, k)$.
  - $$P\{i - 1 < kU < i\} = \frac{1}{k} \quad i = 1, \ldots, k$$
  - Take $N_k = [kU] + 1$, where $[x]$ is the integer part of x.
- Algorithm can be written as follows:

Step 1. Let $X(1), \ldots, X(n)$ be any permutation of $1, 2, \ldots, n$. [For instance, we can set $X(i) = i, i = 1, 2, \ldots, n.$]

Step 2. Let $I = n$.

Step 3. Generate a random number U and set $N = [IU] + 1$.

Step 4. Interchange the values of $X(N)$ and $X(I)$.

Step 5. Reduce the value of I by 1 and if $I > 1$ go to step 3.

Step 6. $X(1), \ldots, X(n)$ is the desired random generated permutation.

10.2 General techniques for simulating continuous random variables

10.2.1 The inverse transformation method

Proposition 2.1: Let U be a uniform (0,1) random variable. For any continuous distribution function F, if we define the random variable Y by

$$Y = F^{-1}(U)$$

then the random variable Y has distribution function F. $[F^{-1}(x)$ is defined to equal that value u for which $F(y) = x$.]

Proof

- $$F_Y(a) = P\{Y \leq a\}$$

$$= P\{F^{-1}(U) \le a\} \qquad (1)$$

- $F(x)$ is a monotone function, it follows that $F^{-1}(U) \le a$ if and only if $U \le F(a)$.
- From Equation (1),

$$F_Y(a) = P\{U \le F(a)\} = F(a).$$

Example 10.2a. *Simulating an exponential random variable.* If $F(x) = 1 - e^{-x}$, then $F^{-1}(u)$ is that value of x such that

$$1 - e^{-x} = u$$

or

$$x = -\log(1-u)$$

- If U is a uniform (0,1) variable, then

  $$F^{-1}(U) = -\log(1-U)$$

  is exponentially distributed with mean 1.
- Since $1-U$ is also uniformly distributed on (0,1), it follows that $-\log U$ is exponential with mean 1.

- Since cX is exponential with mean c when X is exponential with mean 1, it follows that $-c \log U$ is exponential with mean c.

Example 10.2b. *Simulating a gamma (n, λ) random variable.* To simulate from a gamma distribution with parameters (n, λ), when n is an integer, we use the fact that the sum of n independent exponential random variables each having rate λ has this distribution.

- If $U_1, \ldots, U_n$ are independent uniform (0,1) random variables,

$$X = -\sum_{i=1}^{n} \frac{1}{\lambda} \log U_i = -\frac{1}{\lambda} \log \left(\prod_{i=1}^{n} U_i \right)$$

has the desired distribution.

10.2.2 The rejection method

- Suppose that we have a method for simulating a random variable having density function $g(x)$.

- We can use this as the basis for simulating from the continuous distribution having density $f(x)$ by simulating Y from g and then accepting this simulated value with a probability proportional to $f(Y)/g(Y)$.
- Let c be a constant such that

$$\frac{f(y)}{g(y)} \leq c \text{ for all } y$$

- **Rejection Method**

Step 1. Simulate Y having density g and simulate a random number U.

Step 2. If $U \leq f(Y)/cg(Y)$, set $X = Y$. Otherwise return to step 1.

> **Proposition 2.2:** The random variable X generated by the rejection method has density function f.

Proof

- Let X be the value obtained and let N denote the number of necessary iterations.

- $$\begin{aligned} P\{X \le x\} &= P\{Y_N \le x\} \\ &= P\left\{Y \le x | U \le \frac{f(Y)}{cg(Y)}\right\} \\ &= \frac{\left\{Y \le x, U \le \frac{f(Y)}{cg(Y)}\right\}}{K} \end{aligned}$$

  where $K = P\{U \le f(Y)/cg(Y)\}$.

- The joint density function of Y and U:

  $$f(y,u) = g(y) \qquad 0 < u < 1$$

- $$\begin{aligned} P\{X \le x\} &= \frac{1}{K} \int\int_{y \le x; 0 \le u \le f(y)/cg(y)} g(y) du dy \\ &= \frac{1}{K} \int_{-\infty}^{x} \int_{0}^{f(y)/cg(y)} du g(y) dy \\ &= \frac{1}{cK} \int_{-\infty}^{x} f(y) dy \end{aligned}$$

- Letting x approach ∞, and since

  $$1 = \frac{1}{cK} \int_{-\infty}^{\infty} f(y) dy = \frac{1}{cK}$$

- $P\{X \le x\} = \int_{-\infty}^{x} f(y) dy$

Remark:

(a) We ”accept the value Y with probability $f(Y)/cg(Y)$” is by generating a random number U and then accepting Y if $f(Y)/cg(Y)$,

(b) Since each iteration will, independently, result in an accepted value with probability $P\{U \leq f(Y)/cg(Y)\} = K = 1/c$, it follows that the number of iterations has a geometric distribution with mean c.

Example 10.2c. *Simulating a normal random variable.* To simulate a unit normal random variable Z (that is, one with mean 0 and variance 1), note first that the absolute value of has probability density function

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \quad 0 < x < \infty \quad (2.3)$$

We will start by simulating from the preceding density function by using the rejection method with g being the exponential density function with mean 1.

- That is

$$g(x) = e^{-x} \quad 0 < x < \infty$$

Now, note that

$$\begin{aligned}
\frac{f(x)}{g(x)} &= \sqrt{\frac{2}{\pi}} \exp\left\{\frac{-(x^2 - 2x)}{2}\right\} \\
&= \sqrt{\frac{2}{\pi}} \exp\left\{\frac{-(x^2 - 2x + 1)}{2} + \frac{1}{2}\right\} \\
&= \sqrt{\frac{2e}{\pi}} \exp\left\{\frac{-(x-1)^2}{2}\right\} \qquad (2.4) \\
&\le \sqrt{\frac{2e}{\pi}}
\end{aligned}$$

- Hence we can take $c = \sqrt{2e/\pi}$; so, from Equation (2.4),

$$\frac{f(x)}{cg(x)} = \exp\left\{\frac{-(x-1)^2}{2}\right\}$$

- Using the rejection method to simulate the absolute value of a unit normal random variable:

(a) Generate independent random variables Y and U, where $Y \sim \exp(1)$ and $U \sim U(0,1)$.

(b) If $U \leq \exp\{-(Y-1)^2/2\}$, set $X = Y$. Otherwise, return to (a).

- Once we have simulated a random variable X having density function as in Equation (2.3).
- We can generate a unit normal variable Z by letting Z be equally likely to be either X or $-X$.
- In step (b), the value Y is accepted if

$$U \leq \exp\{-(Y-1)^2/2\}$$
$$\Updownarrow$$
$$-\log U \geq (Y-1)^2/2.$$

- $-\log U \sim \exp(1)$ from Example 2a, so steps (a) and (b) are equivalent to

(a') Generate independent exponentials with rate 1, Y_1 and Y_2.

(b') If $Y_2 \geq (Y_1 - 1)^2/2$, set $X = Y_1$. Otherwise, return to (a).

- Suppose that the foregoing results in Y_1 being accepted – and so we know that Y_2 is larger than $(Y_1 - 1)^2/2$.
- How much does the one exceed the other?
- Recall that $Y_2 \sim \exp(1) \Rightarrow Y_2 - (Y_1 - 1)^2/2 \sim \exp(1)$.

- The following algorithm that generates an exponential with rate 1 and an independent unit normal random variable.

Step 1. Generate a random variable Y_1 with $Y_1 \sim \exp(1)$.

Step 2. Generate a random variable Y_2 with $Y_2 \sim \exp(1)$.

Step 3. If $Y_2 - (Y_1 - 1)^2/2 > 0$ set $Y = Y_2 - (Y_1 - 1)^2/2$ and go to step 4. Otherwise, go to step 1.

Step 4. Generate a random number U and set

$$Z = \begin{cases} Y_1 & \text{if } U \leq 1/2 \\ -Y_1 & \text{if } U > 1/2 \end{cases}$$

- The random variables Z and Y generated by the foregoing are independent with $Z \sim N(0,1)$ and $Y \sim \exp(1)$.

Remark:

(a) Since $c = \sqrt{2e/\pi} \approx 1.32$, the foregoing requires a geometrically distributed number of iterations of step 2 with mean 1.32.

(b) If we want to generate a sequence of unit normal random variables, then we can use the exponential random variable Y obtained in step 3 as the initial exponential needed in step 1 for the next normal to be generated.

- On average, we can simulate a unit normal by generating 1.64 $(= 2 \times 1.32 - 1)$ exponentials and computing 1.32 squares.

Example 10.2d. *Simulating normal random variables-the polar method.* It was shown in Example 7b of Chapter 6 that if X and Y are independent unit normal random variables then polar coordinates $R = \sqrt{X^2 + Y^2}$, $\Theta = \tan^{-1}(Y/X)$ are independent, with R^2 being exponentially distributed with mean 2 and Θ being uniformly distributed on $(0, 2\pi)$.

- Hence, if U_1 and U_2 are random numbers then (using the result of Example 10.2a) we can set

$$\begin{aligned} R &= (-2 \log U_1)^{1/2} \\ \Theta &= 2\pi U_2 \end{aligned}$$

which yields that

$$\begin{aligned} X &= R \cos \Theta = (-2 \log U_1)^{1/2} \cos(2\pi U_2) \\ Y &= R \sin \Theta = (-2 \log U_1)^{1/2} \sin(2\pi U_2) \quad (2.5) \end{aligned}$$

are independent unit normals.

- If $U \sim U(0, 1)$, then $2U \sim U(0, 2)$ and $2U - 1 \sim U(-1, 1)$.

- Generate random numbers U_1 and U_2 and set
$$V_1 = 2U_1 - 1$$
$$V_2 = 2U_2 - 1$$
then (V_1, V_2) is uniformly distributed in the square of area 4 centered at (0,0).

- Continually generate such pairs (V_1, V_2) until (V_1, V_2) is such that $V_1^2 + V_2^2 \leq 1$.

- Let $\overline{R}, \overline{\Theta}$ denote the polar coordinates of this pair, then

(a) $\overline{R}$ and $\overline{\Theta}$ are independent.

(b) $\overline{R} \sim U(0, 1)$.

(c) $\overline{\Theta} \sim U(0, 2\pi)$.

- Since
$$\sin\overline{\Theta} = \frac{V_2}{\overline{R}} = \frac{V_2}{\sqrt{V_1^2 + V_2^2}}$$
$$\cos\overline{\Theta} = \frac{V_1}{\overline{R}} = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$

and from Equation (2.5)

$$\begin{aligned} X &= (-2\log U)^{1/2} V_1/\overline{R} \\ Y &= (-2\log U)^{1/2} V_2/\overline{R} \end{aligned}$$

- Since (conditional on $V_1^2 + V_2^2 \le 1$) $\overline{R}^2 \sim U(0,1)$ and independent $\overline{\Theta}$, thus showing that

$$\begin{aligned} X &= (-2\log \overline{R}^2)^{1/2} \frac{V_1}{\overline{R}} = \sqrt{\frac{-2\log S}{S}} V_1 \\ Y &= (-2\log \overline{R}^2)^{1/2} \frac{V_2}{\overline{R}} = \sqrt{\frac{-2\log S}{S}} V_2 \end{aligned}$$

where $S = \overline{R}^2 = V_1^2 + V_2^2$.

- **Polar method:** To generating a pair of independent unit normals:

Step 1. Generate random numbers U_1 and U_2.

Step 2. Set $V_1 = 2U_1 - 1, V_2 = 2U_2 - 1, S = V_1^2 + V_2^2$.

Step 3. If $S > 1$, return to step 1.

Step 4. Return the independent unit normals

$$X = \sqrt{\frac{-2\log S}{S}}V_1, Y = \sqrt{\frac{-2\log S}{S}}V_2$$

- Since the probability that a random point in the square will fall within the circle is equal to $\pi/4$, it follows that, on average, the polar method will require $4/\pi \approx 1.273$ iterations of step 1.

Example 10.2e. *Simulating a chi-squared random variable.* The chi-squared distribution with n degrees of freedom is the distribution of $X_n^2 = Z_1^2 + \cdots + Z_n^2$, where Z_i, $i = 1, \ldots, n$ are independent unit normals.

- Now it was shown in Section 3 of Chapter 6 that $Z_1^2 + Z_2^2$ has an exponential distribution with rate $\frac{1}{2}$.
- Hence, when n is even, say $n = 2k$, X_{2k}^2 has a gamma distribution with parameters $(k, \frac{1}{2})$.

- Hence, $-2\log(\Pi_{i=1}^{k} U_i)$ has a chi-squared distribution with $2k$ degrees of freedom.
- We can simulate a chi-squared random variable with $2k+1$ degrees of freedom by first simulating a unit normal random variable Z and then adding Z^2 to the foregoing.
- That is,
$$X^2_{2k+1} = Z^2 - 2\log(\Pi_{i=1}^{k} U_i)$$
where Z, $U_1, \ldots, U_n$ are independent with Z being a unit normal and the other being uniform (0,1) random variables.

10.3 Simulating from discrete distributions

- To simulate a random variable Z having mass function
$$P\{X = x_j\} = P_j, \quad j = 0, 1, \ldots, \quad \sum_j P_j = 1$$
- Use the following discrete time analog of the inverse transform technique.

- To simulate X for which $P\{X = x_j\} = P_j$
  - Let $U \sim U(0,1)$.
  - Set

$$X = \begin{cases} x_1 & \text{if } U < P_1 \\ x_2 & \text{if } P_1 < U < P_1 + P_2 \\ \vdots & \\ x_j & \text{if } \sum_{i=1}^{j-1} P_i < U < \sum_{i=1}^{j} P_i \\ \vdots & \end{cases}$$

- $P\{X = x_j\} = P\left\{\sum_{i=1}^{j-1} P_i < U < \sum_{i=1}^{j} P_i\right\} = P_j$

Example 10.3a. *The geometric distribution.* Suppose that independent trials each of which results in a "success" with probability p, $0 < p < 1$, are continually performed until a success occurs. Letting X denote the number of necessary trials, then

$$P\{X = i\} = (1-p)^{i-1}p \quad i \geq 1$$

which is seen by noting that $X = i$ if the first

$i - 1$ trials are all failures and the ith is a success.

- The random variable X is said to be a geometric random variable with parameter p.
- Since

$$\begin{aligned}\sum_{i=1}^{j-1} P\{X = i\} &= 1 - P\{X > j - 1\} \\ &= 1 - P\{\text{first } j - 1 \text{ are all failures}\} \\ &= 1 - (1 - p)^{j-1} \quad j \geq 1\end{aligned}$$

  we can simulate such a random variable by generating a random number U and then setting X equal to that value j for which

$$1 - (1 - p)^{j-1} < U < 1 - (1 - p)^j$$

  or, equivalently, for which

$$(1 - p)^j < 1 - U < (1 - p)^{j-1}$$

- Since $1 - U$ has the same distribution as U, we can thus define X by

$$X = \min\{j : (1 - p)^j < U\}$$

$$= \min\{j : j \log(1-p) < \log U\}$$
$$= \min\{j : j > \frac{\log U}{\log(1-p)}\}$$

where the inequality changed sign since $\log(1-p)$ is negative [since $\log(1-p) < \log 1 = 0$].

- Using the notation $[x]$ for the integer part of x (that is , $[x]$ is the largest integer less than or equal to x), we can write

$$X = 1 + \left[\frac{\log U}{\log(1-p)}\right]$$

Example 10.3b. *Simulating a binomial random variable.* A binomial (n, p) random variable can be easily simulated by recalling that it can be expressed as the sum of n independent Bernoulli random variables.

- If $U_1, \ldots, U_n$ are independent uniform (0,1) variables, then letting

$$X_i = \begin{cases} 1 & \text{if } U_i < p \\ 0 & \text{otherwise} \end{cases}$$

- It follows that $X \equiv \sum_{i=1}^{n} X_i$ is a binomial random variable with parameters n and p.

Example 10.3c. *Simulating a Poisson random variable.* To simulate a Poisson random variable with mean λ, generate independent uniform(0,1) random variables $U_1, U_2, \ldots$ stopping at

$$N = \min\{n : \prod_{i=1}^{n} U_i < e^{-\lambda}\}$$

The random variable $X \equiv N - 1$ has the desired distribution.

- That is, if we continue generating random numbers until their product falls below $e^{-\lambda}$, then the number required, minus 1, is Poisson with mean λ.
- That $X \equiv N-1$ is indeed a Poisson random variable having mean λ can perhaps be most easily seen by noting that

$$X + 1 = \min\{n : \prod_{i=1}^{n} U_i < e^{-\lambda}\}$$

is equivalent to

$$X = \max\{n : \prod_{i=1}^{n} U_i \geq e^{-\lambda}\} \quad \text{where } \prod_{i=1}^{0} \equiv 1$$

or, taking logarithms, to

$$X = \max\{n : \sum_{i=1}^{n} \log U_i \geq -\lambda\}$$

or

$$X = \max\{n : \sum_{i=1}^{n} -\log U_i \leq \lambda\}$$

- However, $-\log U_i$ is exponential with rate 1 and X can be thought of as being the maximum number of exponentials having rate 1 that can be summed and still be less that λ.
- But by recalling that the times between successive events of a Poisson process having rate 1 are independent exponentials with rate 1, it follows that X is equal to the number of events by time λ of a Poisson process having rate1; and thus X has a Poisson distribution with mean λ.

10.4 Variance reduction techniques

- Let $X_1, \ldots, X_n$ have a given joint distribution and suppose that we are interested in computing

$$\theta \equiv E[g(X_1, \ldots, X_n)]$$

where g is some specified function.

- It sometimes turns out that it is extremely difficult to analytically compute the foregoing.
- We can attempt to use simulation to estimate θ.

(a) Generate $X_1^{(1)}, \ldots, X_n^{(1)}$ having the same joint distribution as $X_1, \ldots, X_n$ and set

$$Y_1 = g(X_1^{(1)}, \ldots, X_n^{(1)})$$

(b) Simulate a second set of random variables (independent of the first set) $X_1^{(2)}, \ldots, X_n^{(2)}$ having the same joint distribution as $X_1, \ldots, X_n$ and set

$$Y_1 = g(X_1^{(2)}, \ldots, X_n^{(2)})$$

(c) Continue this until you have generated k (some predetermined number) sets and so have also computed $Y_1, Y_2, \ldots, Y_k$.

(d) Now, $Y_1, Y_2, \ldots, Y_k$ i.i.d. $g(X_1, \ldots, X_n)$. Let

$$\overline{Y} = \sum_{i=1}^{k} \frac{Y_i}{k}$$

then

$$\begin{aligned} E[\overline{Y}] &= \theta \\ E[(\overline{Y} - \theta)^2] &= \mathrm{Var}(\overline{Y}) \end{aligned}$$

$\Rightarrow$

– We can use $\overline{Y}$ as an estimate of θ.

– We would like $\mathrm{Var}(\bar{Y})$ to be as small as possible.

– Next present 3 general techniques for reducing the variance of our estimator.

10.4.1 Use of antithetic variables

- Suppose that we have generated Y_1 and Y_2 which are identically distributed random variables having mean θ.

$$\begin{aligned}\mathrm{Var}\left(\frac{Y_1+Y_2}{2}\right) &= \frac{1}{4}[\mathrm{Var}(Y_1)+\mathrm{Var}(Y_2)\\ &\quad +2\mathrm{Cov}(Y_1,Y_2)]\\ &= \frac{\mathrm{Var}(Y_1)}{2}+\frac{\mathrm{Cov}(Y_1,Y_2)}{2}\end{aligned}$$

- It would be advantageous if Y_1 and Y_2 rather than being independent were negatively correlated.
- Suppose that the random variables $X_1, \ldots, X_n$ are independent and each is simulated via the inverse transform technique.
- X_i is simulated from $F_i^{-1}(U_i)$ where U_i is a random number and F_i is the distribution of X_i.

$$Y_1 = g(F_1^{-1}(U_1), \ldots, F_n^{-1}(U_n))$$

- Since $1-U \sim U(0,1)$ whenever U is a random number, it follows that

$$Y_2 = g(F_1^{-1}(1-U_1), \ldots, F_n^{-1}(1-U_n))$$

will have the same distribution as Y_1.

- If Y_1 and Y_2 were negatively correlated, then generating Y_2 by this means would lead to a smaller variance than if it were generated by a new set of random numbers.
- Although we cannot, in general, be certain that Y_1 and Y_2 will be negatively correlated, this often turns out to be the case and indeed it can be proven that it will be so whenever g is a monotonic function.

10.4.2 Variance reduction by conditioning

- $\mathrm{Var}(Y) = E[\mathrm{Var}(Y|Z)] + \mathrm{Var}(E[Y|Z])$.
- We are interested in estimating $E[g(X_1, \ldots, X_n)]$ by simulating $\mathbf{X} = (X_1, \ldots, X_n)$ and computing $Y = g(\mathbf{X})$.
- If for some random variable Z we can compute $E[Y|Z]$ then, as $\mathrm{Var}(Y|Z) \geq 0$, it follows from the conditional variance formula

above that

$$\mathrm{Var}(E[Y|Z]) \leq \mathrm{Var}(Y)$$

since $E[E[Y|Z]] = E[Y]$.

- $E[Y|Z]$ is better estimator of $E[Y]$ than is Y.

Example 10.4a. *Estimation of π.* Let U_1 and U_2 be random numbers and set $V_i = 2U_i - 1, i = 1, 2$. As noted in Example 10.2d, (V_1, V_2) will be uniformly distributed in the square of area 4 centered at (0,0).

- The probability that this point will fall within the inscribed circle of radius 1 centered at (0,0) (see Figure 10.2) is equal to $\pi/4$ (the ratio of the area of the circle to that of the square).
- Hence, upon simulating a large number n of such pairs and setting

$$I_j = \begin{cases} 1 & \text{if the } j\text{th pair fall within the circle} \\ 0 & \text{otherwise} \end{cases}$$

it follows the I_j, $j = 1, \ldots, n$ will be independent and identically distributed random variables having $E[I_j] = \pi/4$.

- Thus, by the strong law of large numbers
$$\frac{I_1 + \cdots + I_n}{n} \to \frac{\pi}{4} \qquad \text{as} \quad n \to \infty$$
- Therefore, it follows that by simulating a large number of pairs (V_1, V_2) and multiplying the proportion of them that fall within the circle by 4, we can accurately approximate π.

- Let I be the indicator variable above for the pair (V_1, V_2).
$$\begin{aligned} E[I|Z] &= P\{V_1^2 + V_2^2 \le 1 | V_1\} \\ &= P\{V_2^2 \le 1 - V_1^2 | V_1\} \end{aligned}$$
- $P\{V_2^2 \le 1 - V_1^2 | V_1 = v\}$
$$\begin{aligned} &= P\{V_2^2 \le 1 - v^2\} \\ &= P\{-\sqrt{1 - v^2} \le V_2 \le \sqrt{1 - v^2}\} \\ &= \sqrt{1 - v^2} \end{aligned}$$

- $E[I|V_1] = E[\sqrt{1 - V_1^2}]$
- An improvement on using the average value of I to estimate $\pi/4$ is to use the average value of $\sqrt{1 - V_1^2}$.
-

$$\begin{aligned} E[\sqrt{1 - V_1^2}] &= \int_{-1}^{1} \frac{1}{2}\sqrt{1 - v^2}dv \\ &= \int_0^1 \sqrt{1 - u^2}du \\ &= E[\sqrt{1 - U^2}] \end{aligned}$$

where $U \sim U(0, 1)$.

- We can generate n random numbers U and use the average value of $\sqrt{1 - U^2}$ as our estimate of $\pi/4$.
- Further improved:
  - $g(u) = \sqrt{1 - u^2}, 0 \leq u \leq 1$ is a monotone decreasing function of u.
  - The method of antithetic variables will reduce the variance of the estimator of $E[\sqrt{1 - U^2}]$.

– Generating only $n/2$ random numbers U and then using one-half the average of $\sqrt{1-U^2} + \sqrt{1-(1-U)^2}$ as the estimator of $\pi/4$.

- The estimates of π resulting from simulations, using $n = 10,000$, based on the three estimators above:

Method	Estimate of π
Proportion of the random points that fall in the circle	3.1612
Average value of $\sqrt{1-U^2}$	3.128448
Average value of $\sqrt{1-U^2} + \sqrt{1-(1-U)^2}$	3.139578

- A further simulation using the final approach and $n = 64,000$ yielded the estimate 3.143288.

10.4.3 Control variates

- To simulation to estimate $E[g(\mathbf{X})]$ where $\mathbf{X} = (X_1, \ldots, X_n)$.
- Suppose that for some function f the expected value of $f(\mathbf{X})$is know – say $E[f(\mathbf{X})] = \mu$.

- For any constant a we can also use
$$W = g(\mathbf{X}) + a[f(\mathbf{X}) - \mu]$$
as an estimator of $E[g(\mathbf{X})]$.
$$\begin{aligned}\mathrm{Var}(W) &= \mathrm{Var}[g(\mathbf{X})] + a^2\mathrm{Var}[f(\mathbf{X})] \\ &\quad +2a\mathrm{Cov}[g(\mathbf{X}), f(\mathbf{X})]\end{aligned}$$

- The foregoing is minimized when
$$a = \frac{-\mathrm{Cov}[g(\mathbf{X}), f(\mathbf{X})]}{\mathrm{Var}[f(\mathbf{X})]}$$
and for this value of a
$$\mathrm{Var}(W) = \mathrm{Var}[g(\mathbf{X})] - \frac{[\mathrm{Cov}[g(\mathbf{X}), f(\mathbf{X})]]^2}{\mathrm{Var}[f(\mathbf{X})]}$$

- Unfortunately, neither $\mathrm{Var}[f(\mathbf{X})]$ nor $\mathrm{Cov}[g(\mathbf{X}), f(\mathbf{X})]$ is usually known, so we cannot usually obtain the foregoing reduction in variance.

- One approach in practice is to use the simulated data to estimate these quantities.

Summary

Inverse transform method:

- Let F be a continuous distribution function and $U \sim U(0,1)$.
- The random variable $F^{-1}(U)$ has distribution function F, where $F^{-1}(u)$ is that value x such that $F(x) = u$.

Rejection method:

- Suppose that we have an efficient procedure for generating a random variable from the density function g, and that we desire to generate a random variable having density function f.
- For accomplishing this starts by determining a constant c such that

$$\max \frac{f(x)}{g(x)} \leq c$$

- The algorithm as follows:
  1. Generate Y having density g.
  2. Generate a random number U.

3. If $U \leq f(Y)/cg(Y)$, set $X = Y$ and stop.
4. Return to step 1.

- The number of passes through step 1 is a geometric random variable with mean c.
- Standard normal random variables can be efficiently simulated by using the rejection method (with $g \sim \exp(1)$) or by using a technique known as the polar algorithm.

Estimate a quantity θ:

- One often generates the value of a partial sequence of random variables whose expected value is θ.
- The efficiency of this approach is increased when these random variables have a small variance.
- Three techniques:

1. The use of antithetic variables.

2. The use of conditional expectations.
3. The use of control variables.