

Hence, if we generate k such random variables $X_1, \dots, X_k$ we can estimate d by

$$d \approx \frac{n \sum_{i=1}^k 1/m(X_i)}{k}$$

Suppose now that each item in the list has a value attached to it— $v(i)$ being the value of the i th element. The sum of the values of the distinct items—call it v —can be expressed as

$$v = \sum_{i=1}^n \frac{v(i)}{m(i)}$$

Now if $X = [nU] + 1$, where U is a random number, then

$$E\left[\frac{v(X)}{m(X)}\right] = \sum_{i=1}^n \frac{v(i)}{m(i)} \frac{1}{n} = \frac{v}{n}$$

Hence, we can estimate v by generating $X_1, \dots, X_k$ and then estimating v by

$$v \approx \frac{n}{k} \sum_{i=1}^k \frac{v(X_i)}{m(X_i)}$$

For an important application of the preceding, let $A_i = \{a_{i,1}, \dots, a_{i,n_i}\}$, $i = 1, \dots, s$ denote events, and suppose we are interested in estimating $P(\bigcup_{i=1}^s A_i)$. Since

$$P\left(\bigcup_{i=1}^s A_i\right) = \sum_{a \in \bigcup A_i} P(a) = \sum_{i=1}^s \sum_{j=1}^{n_i} \frac{P(a_{i,j})}{m(a_{i,j})}$$

where $m(a_{i,j})$ is the number of events to which the point $a_{i,j}$ belongs, the preceding method can be used to estimate $P(\bigcup_1^s A_i)$.

Note that the preceding procedure for estimating v can be effected without prior knowledge of the set of values $\{v_1, \dots, v_n\}$. That is, it suffices that we can determine the value of an element in a specific place and the number of times that element appears on the list. When the set of values is *a priori* known, there is another approach available as will be shown in [Example 11.11](#). ■

11.2 General Techniques for Simulating Continuous Random Variables

In this section we present three methods for simulating continuous random variables.

11.2.1 The Inverse Transformation Method

A general method for simulating a random variable having a continuous distribution—called the *inverse transformation method*—is based on the following proposition.

Proposition 11.1 Let U be a uniform $(0, 1)$ random variable. For any continuous distribution function F if we define the random variable X by

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

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

Proof.

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{F^{-1}(U) \leq a\} \end{aligned} \quad (11.1)$$

Now, since $F(x)$ is a monotone function, it follows that $F^{-1}(U) \leq a$ if and only if $U \leq F(a)$. Hence, from Equation (11.1), we see that

$$\begin{aligned} F_X(a) &= P\{U \leq F(a)\} \\ &= F(a) \end{aligned} \quad \blacksquare$$

Hence, we can simulate a random variable X from the continuous distribution F , when F^{-1} is computable, by simulating a random number U and then setting $X = F^{-1}(U)$.

Example 11.3 (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

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

Hence, 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 . $\blacksquare$

11.2.2 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 $f(Y)/g(Y)$.

Specifically, let c be a constant such that

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

We then have the following technique for simulating a random variable having density f .

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 11.2 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. Then

$$\begin{aligned}
 P\{X \leq x\} &= P\{Y_N \leq x\} \\
 &= P\{Y \leq x | U \leq f(Y)/cg(Y)\} \\
 &= \frac{P\{Y \leq x, U \leq f(Y)/cg(Y)\}}{K} \\
 &= \frac{\int P\{Y \leq x, U \leq f(Y)/cg(Y) | Y = y\} g(y) dy}{K} \\
 &= \frac{\int_{-\infty}^x (f(y)/cg(y)) g(y) dy}{K} \\
 &= \frac{\int_{-\infty}^x f(y) dy}{Kc}
 \end{aligned}$$

where $K = P\{U \leq f(Y)/cg(Y)\}$. Letting $x \rightarrow \infty$ shows that $K = 1/c$ and the proof is complete. ■

Remarks

- (i) The preceding method was originally presented by Von Neumann in the special case where g was positive only in some finite interval (a, b) , and Y was chosen to be uniform over (a, b) (that is, $Y = a + (b - a)U$).
- (ii) Note that the way in which we “accept the value Y with probability $f(Y)/cg(Y)$ ” is by generating a uniform $(0, 1)$ random variable U and then accepting Y if $U \leq f(Y)/cg(Y)$.
- (iii) Since each iteration of the method will, independently, result in an accepted value with probability $P\{U \leq f(Y)/cg(Y)\} = 1/c$ it follows that the number of iterations is geometric with mean c .
- (iv) Actually, it is not necessary to generate a new uniform random number when deciding whether or not to accept, since at a cost of some additional computation, a single random number, suitably modified at each iteration, can be used throughout. To see how, note that the actual value of U is not used—only whether or not $U < f(Y)/cg(Y)$. Hence, if Y is rejected—that is, if $U > f(Y)/cg(Y)$ —we can use the fact that, given Y ,

$$\frac{U - f(Y)/cg(Y)}{1 - f(Y)/cg(Y)} = \frac{cUg(Y) - f(Y)}{cg(Y) - f(Y)}$$

is uniform on $(0, 1)$. Hence, this may be used as a uniform random number in the next iteration. As this saves the generation of a random number at the cost of the preceding computation, whether it is a net savings depends greatly upon the method being used to generate random numbers. ■

Example 11.4 Let us use the rejection method to generate a random variable having density function

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1$$

Since this random variable (which is beta with parameters 2, 4) is concentrated in the interval $(0, 1)$, let us consider the rejection method with

$$g(x) = 1, \quad 0 < x < 1$$

To determine the constant c such that $f(x)/g(x) \leq c$, we use calculus to determine the maximum value of

$$\frac{f(x)}{g(x)} = 20x(1-x)^3$$

Differentiation of this quantity yields

$$\frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] = 20 \left[(1-x)^3 - 3x(1-x)^2 \right]$$

Setting this equal to 0 shows that the maximal value is attained when $x = \frac{1}{4}$, and thus

$$\frac{f(x)}{g(x)} \leq 20 \left(\frac{1}{4} \right) \left(\frac{3}{4} \right)^3 = \frac{135}{64} \equiv c$$

Hence,

$$\frac{f(x)}{cg(x)} = \frac{256}{27} x(1-x)^3$$

and thus the rejection procedure is as follows:

Step 1: Generate random numbers U_1 and U_2 .

Step 2: If $U_2 \leq \frac{256}{27} U_1(1-U_1)^3$, stop and set $X = U_1$. Otherwise return to step 1.

The average number of times that step 1 will be performed is $c = \frac{135}{64}$. ■

Example 11.5 (Simulating a Normal Random Variable) To simulate a standard normal random variable Z (that is, one with mean 0 and variance 1) note first that the absolute value of Z has density function

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad 0 < x < \infty \quad (11.2)$$

We will start by simulating from the preceding density by using the rejection method with

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

Now, note that

$$\frac{f(x)}{g(x)} = \sqrt{2e/\pi} \exp\{-(x-1)^2/2\} \leq \sqrt{2e/\pi}$$

Hence, using the rejection method we can simulate from [Equation \(11.2\)](#) as follows:

- (a) Generate independent random variables Y and U , Y being exponential with rate 1 and U being uniform on $(0, 1)$.
- (b) If $U \leq \exp\{-(Y-1)^2/2\}$, or equivalently, if

$$-\log U \geq (Y-1)^2/2$$

set $X = Y$. Otherwise return to step (a).

Once we have simulated a random variable X having Density Function (11.2) we can then generate a standard normal random variable Z by letting Z be equally likely to be either X or $-X$.

To improve upon the foregoing, note first that from [Example 11.3](#) it follows that $-\log U$ will also be exponential with rate 1. Hence, steps (a) and (b) are equivalent to the following:

- (a') Generate independent exponentials with rate 1, Y_1 , and Y_2 .
- (b') Set $X = Y_1$ if $Y_2 \geq (Y_1-1)^2/2$. Otherwise return to step (a').

Now suppose that we accept step (b'). It then follows by the lack of memory property of the exponential that the amount by which Y_2 exceeds $(Y_1-1)^2/2$ will also be exponential with rate 1.

Hence, summing up, we have the following algorithm which generates an exponential with rate 1 and an independent standard normal random variable:

Step 1: Generate Y_1 , an exponential random variable with rate 1.

Step 2: Generate Y_2 , an exponential with rate 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

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

The random variables Z and Y generated by the preceding are independent with Z being normal with mean 0 and variance 1 and Y being exponential with rate 1. (If we want the normal random variable to have mean μ and variance σ^2 , just take $\mu + \sigma Z$.)



Remarks

- (i) Since $c = \sqrt{2e/\pi} \approx 1.32$, the preceding requires a geometric distributed number of iterations of step 2 with mean 1.32.
- (ii) The final random number of step 4 need not be separately simulated but rather can be obtained from the first digit of any random number used earlier. That is, suppose we generate a random number to simulate an exponential; then we can strip off the initial digit of this random number and just use the remaining digits (with the decimal point moved one step to the right) as the random number. If this initial digit is 0, 1, 2, 3, or 4 (or 0 if the computer is generating binary digits), then we take the sign of Z to be positive and take it to be negative otherwise.
- (iii) If we are generating a sequence of standard normal random variables, then we can use the exponential obtained in step 4 as the initial exponential needed in step 1 for the next normal to be generated. Hence, on the average, we can simulate a unit normal by generating 1.64 exponentials and computing 1.32 squares.

11.2.3 The Hazard Rate Method

Let F be a continuous distribution function with $\bar{F}(0) = 1$. Recall that $\lambda(t)$, the hazard rate function of F , is defined by

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}, \quad t \geq 0$$

(where $f(t) = F'(t)$ is the density function). Recall also that $\lambda(t)$ represents the instantaneous probability intensity that an item having life distribution F will fail at time t given it has survived to that time.

Suppose now that we are given a bounded function $\lambda(t)$, such that $\int_0^\infty \lambda(t) dt = \infty$, and we desire to simulate a random variable S having $\lambda(t)$ as its hazard rate function.

To do so let λ be such that

$$\lambda(t) \leq \lambda \quad \text{for all } t \geq 0$$

To simulate from $\lambda(t)$, $t \geq 0$, we will

- (a) simulate a Poisson process having rate λ . We will then only “accept” or “count” certain of these Poisson events. Specifically we will
- (b) count an event that occurs at time t , independently of all else, with probability $\lambda(t)/\lambda$.

We now have the following proposition.

Proposition 11.3 The time of the first counted event—call it S —is a random variable whose distribution has hazard rate function $\lambda(t)$, $t \geq 0$.

Proof.

$$\begin{aligned}
P\{t < S < t + dt | S > t\} \\
&= P\{\text{first counted event in } (t, t + dt) | \text{no counted events prior to } t\} \\
&= P\{\text{Poisson event in } (t, t + dt), \text{ it is counted} | \text{no counted events prior to } t\} \\
&= P\{\text{Poisson event in } (t, t + dt), \text{ it is counted}\} \\
&= [\Lambda dt + o(dt)] \frac{\lambda(t)}{\lambda} = \lambda(t) dt + o(dt)
\end{aligned}$$

which completes the proof. Note that the next to last equality follows from the independent increment property of Poisson processes. ■

Because the interarrival times of a Poisson process having rate λ are exponential with rate λ , it thus follows from [Example 11.3](#) and the previous proposition that the following algorithm will generate a random variable having hazard rate function $\lambda(t)$, $t \geq 0$.

Hazard Rate Method for Generating S : $\lambda_s(t) = \lambda(t)$

Let λ be such that $\lambda(t) \leq \lambda$ for all $t \geq 0$. Generate pairs of random variables U_i, X_i , $i \geq 1$, with X_i being exponential with rate λ and U_i being uniform (0, 1), stopping at

$$N = \min \left\{ n : U_n \leq \lambda \left(\sum_{i=1}^n X_i \right) / \lambda \right\}$$

Set

$$S = \sum_{i=1}^N X_i$$

■

To compute $E[N]$ we need the result, known as Wald's equation, which states that if $X_1, X_2, \dots$ are independent and identically distributed random variables that are observed in sequence up to some random time N then

$$E \left[\sum_{i=1}^N X_i \right] = E[N]E[X]$$

More precisely let $X_1, X_2, \dots$ denote a sequence of independent random variables and consider the following definition.

Definition 11.1 An integer-valued random variable N is said to be a *stopping time* for the sequence $X_1, X_2, \dots$ if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \dots$ for all $n = 1, 2, \dots$.

Intuitively, we observe the X_n s in sequential order and N denotes the number observed before stopping. If $N = n$, then we have stopped after observing $X_1, \dots, X_n$ and before observing $X_{n+1}, X_{n+2}, \dots$ for all $n = 1, 2, \dots$.

Example 11.6 Let $X_n, n = 1, 2, \dots$, be independent and such that

$$P\{X_n = 0\} = P\{X_n = 1\} = \frac{1}{2}, \quad n = 1, 2, \dots$$

If we let

$$N = \min\{n: X_1 + \dots + X_n = 10\}$$

then N is a stopping time. We may regard N as being the stopping time of an experiment that successively flips a fair coin and then stops when the number of heads reaches 10. ■

Proposition 11.4 (Wald's Equation) If $X_1, X_2, \dots$ are independent and identically distributed random variables having finite expectations, and if N is a stopping time for $X_1, X_2, \dots$ such that $E[N] < \infty$, then

$$E\left[\sum_{n=1}^N X_n\right] = E[N]E[X]$$

Proof. Letting

$$I_n = \begin{cases} 1, & \text{if } N \geq n \\ 0, & \text{if } N < n \end{cases}$$

we have

$$\sum_{n=1}^N X_n = \sum_{n=1}^{\infty} X_n I_n$$

Hence,

$$E\left[\sum_{n=1}^N X_n\right] = E\left[\sum_{n=1}^{\infty} X_n I_n\right] = \sum_{n=1}^{\infty} E[X_n I_n] \quad (11.3)$$

However, $I_n = 1$ if and only if we have not stopped after successively observing $X_1, \dots, X_{n-1}$. Therefore, I_n is determined by $X_1, \dots, X_{n-1}$ and is thus independent of X_n . From Equation (11.3) we thus obtain

$$\begin{aligned} E\left[\sum_{n=1}^N X_n\right] &= \sum_{n=1}^{\infty} E[X_n]E[I_n] \\ &= E[X] \sum_{n=1}^{\infty} E[I_n] \\ &= E[X]E\left[\sum_{n=1}^{\infty} I_n\right] \\ &= E[X]E[N] \end{aligned}$$

■

Returning to the hazard rate method, we have

$$S = \sum_{i=1}^N X_i$$

As $N = \min\{n: U_n \leq \lambda (\sum_{i=1}^n X_i) / \lambda\}$ it follows that the event that $N = n$ is independent of $X_{n+1}, X_{n+2}, \dots$. Hence, by Wald's equation,

$$\begin{aligned} E[S] &= E[N]E[X_i] \\ &= \frac{E[N]}{\lambda} \end{aligned}$$

or

$$E[N] = \lambda E[S]$$

where $E[S]$ is the mean of the desired random variable.

11.3 Special Techniques for Simulating Continuous Random Variables

Special techniques have been devised to simulate from most of the common continuous distributions. We now present certain of these.

11.3.1 The Normal Distribution

Let X and Y denote independent standard normal random variables and thus have the joint density function

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \quad -\infty < x < \infty, -\infty < y < \infty$$

Consider now the polar coordinates of the point (X, Y) . As shown in [Figure 11.1](#),

$$\begin{aligned} R^2 &= X^2 + Y^2, \\ \Theta &= \tan^{-1} Y/X \end{aligned}$$

To obtain the joint density of R^2 and Θ , consider the transformation

$$d = x^2 + y^2, \quad \theta = \tan^{-1} y/x$$

The Jacobian of this transformation is

$$\begin{aligned} J &= \begin{vmatrix} \frac{\partial d}{\partial x} & \frac{\partial d}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix} = \begin{vmatrix} 2x & 2y \\ \frac{1}{1+y^2/x^2} \left(\frac{-y}{x^2} \right) & \frac{1}{1+y^2/x^2} \left(\frac{1}{x} \right) \end{vmatrix} \\ &= 2 \begin{vmatrix} x & y \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix} = 2 \end{aligned}$$

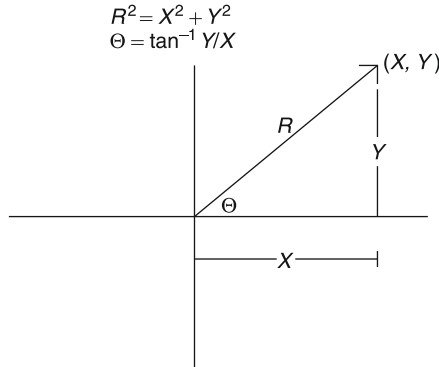


Figure 11.1

Hence, from [Section 2.5.3](#) the joint density of R^2 and Θ is given by

$$\begin{aligned} f_{R^2, \Theta}(d, \theta) &= \frac{1}{2\pi} e^{-d/2} \frac{1}{2} \\ &= \frac{1}{2} e^{-d/2} \frac{1}{2\pi}, \quad 0 < d < \infty, 0 < \theta < 2\pi \end{aligned}$$

Thus, we can conclude that R^2 and Θ are independent with R^2 having an exponential distribution with rate $\frac{1}{2}$ and Θ being uniform on $(0, 2\pi)$.

Let us now go in reverse from the polar to the rectangular coordinates. From the preceding if we start with W , an exponential random variable with rate $\frac{1}{2}$ (W plays the role of R^2) and with V , independent of W and uniformly distributed over $(0, 2\pi)$ (V plays the role of Θ) then $X = \sqrt{W} \cos V$, $Y = \sqrt{W} \sin V$ will be independent standard normals. Hence, using the results of [Example 11.3](#) we see that if U_1 and U_2 are independent uniform $(0, 1)$ random numbers, then

$$\begin{aligned} X &= (-2 \log U_1)^{1/2} \cos(2\pi U_2), \\ Y &= (-2 \log U_1)^{1/2} \sin(2\pi U_2) \end{aligned} \tag{11.4}$$

are independent standard normal random variables.

Remark The fact that $X^2 + Y^2$ has an exponential distribution with rate $\frac{1}{2}$ is quite interesting for, by the definition of the chi-square distribution, $X^2 + Y^2$ has a chi-squared distribution with two degrees of freedom. Hence, these two distributions are identical.

The preceding approach to generating standard normal random variables is called the *Box–Muller approach*. Its efficiency suffers somewhat from its need to compute the preceding sine and cosine values. There is, however, a way to get around this potentially time-consuming difficulty. To begin, note that if U is uniform on $(0, 1)$, then $2U$ is uniform on $(0, 2)$, and so $2U - 1$ is uniform on $(-1, 1)$. Thus, if we generate random

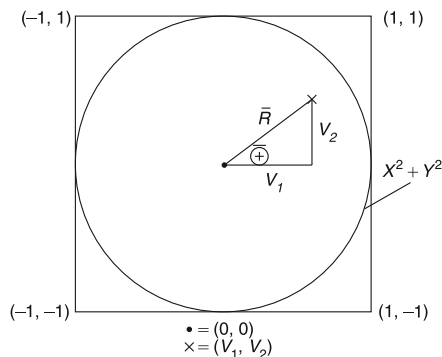


Figure 11.2

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 $(0, 0)$ (see Figure 11.2).

Suppose now that we continually generate such pairs (V_1, V_2) until we obtain one that is contained in the circle of radius 1 centered at $(0, 0)$ —that is, until (V_1, V_2) is such that $V_1^2 + V_2^2 \leq 1$. It now follows that such a pair (V_1, V_2) is uniformly distributed in the circle. If we let $\bar{R}$, $\bar{\Theta}$ denote the polar coordinates of this pair, then it is easy to verify that $\bar{R}$ and $\bar{\Theta}$ are independent, with $\bar{R}^2$ being uniformly distributed on $(0, 1)$, and $\bar{\Theta}$ uniformly distributed on $(0, 2\pi)$.

Since

$$\sin \bar{\Theta} = V_2 / \bar{R} = \frac{V_2}{\sqrt{V_1^2 + V_2^2}},$$

$$\cos \bar{\Theta} = V_1 / \bar{R} = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}$$

it follows from Equation (11.4) that we can generate independent standard normals X and Y by generating another random number U and setting

$$X = (-2 \log U)^{1/2} V_1 / \bar{R},$$

$$Y = (-2 \log U)^{1/2} V_2 / \bar{R}$$

In fact, since (conditional on $V_1^2 + V_2^2 \leq 1$) $\bar{R}^2$ is uniform on $(0, 1)$ and is independent of $\bar{\Theta}$, we can use it instead of generating a new random number U ; thus showing that

$$X = (-2 \log \bar{R}^2)^{1/2} V_1 / \bar{R} = \sqrt{\frac{-2 \log S}{S}} V_1,$$

$$Y = (-2 \log \bar{R}^2)^{1/2} V_2 / \bar{R} = \sqrt{\frac{-2 \log S}{S}} V_2$$

are independent standard normals, where

$$S = \bar{R}^2 = V_1^2 + V_2^2$$

Summing up, we thus have the following approach to generating a pair of independent standard 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

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

The preceding is called the *polar method*. Since the probability that a random point in the square will fall within the circle is equal to $\pi/4$ (the area of the circle divided by the area of the square), it follows that, on average, the polar method will require $4/\pi = 1.273$ iterations of step 1. Hence, it will, on average, require 2.546 random numbers, 1 logarithm, 1 square root, 1 division, and 4.546 multiplications to generate 2 independent standard normals.

11.3.2 The Gamma Distribution

To simulate from a gamma distribution with parameters (n, λ) , where n is an integer, we use the fact that the sum of n independent exponential random variables each having rate λ has this distribution. Hence, if $U_1, \dots, U_n$ are independent uniform $(0, 1)$ random variables,

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

has the desired distribution.

When n is large, there are other techniques available that do not require so many random numbers. One possibility is to use the rejection procedure with $g(x)$ being taken as the density of an exponential random variable with mean n/λ (as this is the mean of the gamma). It can be shown that for large n the average number of iterations needed by the rejection algorithm is $e[(n-1)/2\pi]^{1/2}$. In addition, if we wanted to generate a series of gammas, then, just as in [Example 11.4](#), we can arrange things so that upon acceptance we obtain not only a gamma random variable but also, for free, an exponential random variable that can then be used in obtaining the next gamma (see [Exercise 8](#)).

11.3.3 The Chi-Squared Distribution

The chi-squared distribution with n degrees of freedom is the distribution of $\chi_n^2 = Z_1^2 + \dots + Z_n^2$ where $Z_i, i = 1, \dots, n$ are independent standard normals. Using

the fact noted in the remark at the end of [Section 3.1](#) we see that $Z_1^2 + Z_2^2$ has an exponential distribution with rate $\frac{1}{2}$. Hence, when n is even—say, $n = 2k$ — χ_{2k}^2 has a gamma distribution with parameters $(k, \frac{1}{2})$. Hence, $-2 \log (\prod_{i=1}^k U_i)$ 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 standard normal random variable Z and then adding Z^2 to the preceding. That is,

$$\chi_{2k+1}^2 = Z^2 - 2 \log \left(\prod_{i=1}^k U_i \right)$$

where $Z, U_1, \dots, U_n$ are independent with Z being a standard normal and the others being uniform $(0, 1)$ random variables.

11.3.4 The Beta (n, m) Distribution

The random variable X is said to have a beta distribution with parameters n, m if its density is given by

$$f(x) = \frac{(n+m-1)!}{(n-1)!(m-1)!} x^{n-1} (1-x)^{m-1}, \quad 0 < x < 1$$

One approach to simulating from the preceding distribution is to let $U_1, \dots, U_{n+m-1}$ be independent uniform $(0, 1)$ random variables and consider the n th smallest value of this set—call it $U_{(n)}$. Now $U_{(n)}$ will equal x if, of the $n + m - 1$ variables,

- (i) $n - 1$ are smaller than x ,
- (ii) one equals x ,
- (iii) $m - 1$ are greater than x .

Hence, if the $n + m - 1$ uniform random variables are partitioned into three subsets of sizes $n - 1, 1$, and $m - 1$ the probability (density) that each of the variables in the first set is less than x , the variable in the second set equals x , and all the variables in the third set are greater than x is given by

$$(P\{U < x\})^{n-1} f_U(x) (P\{U > x\})^{m-1} = x^{n-1} (1-x)^{m-1}$$

Hence, as there are $(n+m-1)!/(n-1)!(m-1)!$ possible partitions, it follows that $U_{(n)}$ is beta with parameters (n, m) .

Thus, one way to simulate from the beta distribution is to find the n th smallest of a set of $n + m - 1$ random numbers. However, when n and m are large, this procedure is not particularly efficient.

For another approach consider a Poisson process with rate 1, and recall that given S_{n+m} , the time of the $(n+m)$ th event, the set of the first $n+m-1$ event times is distributed independently and uniformly on $(0, S_{n+m})$. Hence, given S_{n+m} , the n th smallest of the first $n+m-1$ event times—that is, S_n —is distributed as the n th smallest of a set of $n+m-1$ uniform $(0, S_{n+m})$ random variables. But from the

preceding we can thus conclude that S_n/S_{n+m} has a beta distribution with parameters (n, m) . Therefore, if $U_1, \dots, U_{n+m}$ are random numbers,

$$\frac{-\log \prod_{i=1}^n U_i}{-\log \prod_{i=1}^{m+n} U_i} \text{ is beta with parameters } (n, m)$$

By writing the preceding as

$$\frac{-\log \prod_{i=1}^n U_i}{-\log \prod_{i=1}^n U_i - \log \prod_{i=n+1}^{n+m} U_i}$$

we see that it has the same distribution as $X/(X+Y)$ where X and Y are independent gamma random variables with respective parameters $(n, 1)$ and $(m, 1)$. Hence, when n and m are large, we can efficiently simulate a beta by first simulating two gamma random variables.

11.3.5 The Exponential Distribution—The Von Neumann Algorithm

As we have seen, an exponential random variable with rate 1 can be simulated by computing the negative of the logarithm of a random number. Most computer programs for computing a logarithm, however, involve a power series expansion, and so it might be useful to have at hand a second method that is computationally easier. We now present such a method due to Von Neumann.

To begin let $U_1, U_2, \dots$ be independent uniform $(0, 1)$ random variables and define $N, N \geq 2$, by

$$N = \min\{n: U_1 \geq U_2 \geq \dots \geq U_{n-1} < U_n\}$$

That is, N is the index of the first random number that is greater than its predecessor. Let us now compute the joint distribution of N and U_1 .

$$\begin{aligned} P\{N > n, U_1 \leq y\} &= \int_0^1 P\{N > n, U_1 \leq y | U_1 = x\} dx \\ &= \int_0^y P\{N > n | U_1 = x\} dx \end{aligned}$$

Now, given that $U_1 = x$, N will be greater than n if $x \geq U_2 \geq \dots \geq U_n$ or, equivalently, if

$$(a) \quad U_i \leq x, \quad i = 2, \dots, n$$

and

$$(b) \quad U_2 \geq \dots \geq U_n$$

Now, (a) has probability x^{n-1} of occurring and given (a), since all of the $(n-1)!$ possible rankings of $U_2, \dots, U_n$ are equally likely, (b) has probability $1/(n-1)!$ of occurring. Hence,

$$P\{N > n | U_1 = x\} = \frac{x^{n-1}}{(n-1)!}$$

and so

$$P\{N > n, U_1 \leq y\} = \int_0^y \frac{x^{n-1}}{(n-1)!} dx = \frac{y^n}{n!}$$

which yields

$$\begin{aligned} P\{N = n, U_1 \leq y\} &= P\{N > n-1, U_1 \leq y\} - P\{N > n, U_1 \leq y\} \\ &= \frac{y^{n-1}}{(n-1)!} - \frac{y^n}{n!} \end{aligned}$$

Upon summing over all the even integers, we see that

$$\begin{aligned} P\{N \text{ is even}, U_1 \leq y\} &= y - \frac{y^2}{2!} + \frac{y^3}{3!} - \frac{y^4}{4!} + \dots \\ &= 1 - e^{-y} \end{aligned} \tag{11.5}$$

We are now ready for the following algorithm for generating an exponential random variable with rate 1.

- Step 1:** Generate uniform random numbers $U_1, U_2, \dots$ stopping at $N = \min\{n: U_1 \geq \dots \geq U_{n-1} < U_n\}$.
- Step 2:** If N is even accept that run, and go to step 3. If N is odd reject the run, and return to step 1.
- Step 3:** Set X equal to the number of failed runs plus the first random number in the successful run.

To show that X is exponential with rate 1, first note that the probability of a successful run is, from Equation (11.5) with $y = 1$,

$$P\{N \text{ is even}\} = 1 - e^{-1}$$

Now, in order for X to exceed x , the first $[x]$ runs must all be unsuccessful and the next run must either be unsuccessful or be successful but have $U_1 > x - [x]$ (where $[x]$ is the largest integer not exceeding x). As

$$\begin{aligned} P\{N \text{ even}, U_1 > y\} &= P\{N \text{ even}\} - P\{N \text{ even}, U_1 \leq y\} \\ &= 1 - e^{-1} - (1 - e^{-y}) = e^{-y} - e^{-1} \end{aligned}$$

we see that

$$P\{X > x\} = e^{-[x]}[e^{-1} + e^{-(x-[x])} - e^{-1}] = e^{-x}$$

which yields the result.

Let T denote the number of trials needed to generate a successful run. As each trial is a success with probability $1 - e^{-1}$ it follows that T is geometric with mean $1/(1 - e^{-1})$. If we let N_i denote the number of uniform random variables used on the i th run, $i \geq 1$, then T (being the first run i for which N_i is even) is a stopping time for this sequence.