33. MONTE CARLO TECHNIQUES

Revised July 1995 by S. Youssef (SCRI, Florida State University). Updated February 2000 by R. Cousins (UCLA) in consultation with F. James (CERN); October 2003 by G. Cowan (RHUL) and R. Miquel (LBNL), and September 2005 by G. Cowan (RHUL).

Monte Carlo techniques are often the only practical way to evaluate difficult integrals or to sample random variables governed by complicated probability density functions. Here we describe an assortment of methods for sampling some commonly occurring probability density functions.

33.1. Sampling the uniform distribution

Most Monte Carlo sampling or integration techniques assume a "random number generator" which generates uniform statistically independent values on the half open interval \([0, 1)\). There is a long history of problems with various generators on a finite digital computer, but recently, the RANLUX generator \([1]\) has emerged with a solid theoretical basis in chaos theory. Based on the method of Lüscher, it allows the user to select different quality levels, trading off quality with speed.

Other generators are also available which pass extensive batteries of tests for statistical independence and which have periods which are so long that, for practical purposes, values from these generators can be considered to be uniform and statistically independent. In particular, the lagged-Fibonacci based generator introduced by Marsaglia, Zaman, and Tsang \([2]\) is efficient, has a period of approximately \(10^{43}\), produces identical sequences on a wide variety of computers and, passes the extensive "DIEHARD" battery of tests \([3]\). Many commonly available congruential generators fail these tests and often have sequences (typically with periods less than \(2^{32}\)) which can be easily exhausted on modern computers and should therefore be avoided \([4]\).

33.2. Inverse transform method

If the desired probability density function is \(f(x)\) on the range \(-\infty < x < \infty\), its cumulative distribution function (expressing the probability that \(x \leq a\)) is given by Eq. (33.6). If \(a\) is chosen with probability density \(f(a)\), then the integrated probability up to point \(a\), \(F(a)\), is itself a random variable which will occur with uniform probability density on \([0, 1]\). If \(x\) can take on any value, and ignoring the endpoints, we can then find a unique \(x\) chosen from the p.d.f. \(f(s)\) for a given \(u\) if we set

\[
u = F(x) ,
\]

provided we can find an inverse of \(F\), defined by

\[
x = F^{-1}(u) .
\]

This method is shown in Fig. 33.1a. It is most convenient when one can calculate by hand the inverse function of the indefinite integral of \(f\). This is the case for some common functions \(f(x)\) such as \(\exp(x)\), \((1 - x)^n\), and \(1/(1 + x^2)\) (Cauchy or Breit-Wigner), although it does not necessarily produce the fastest generator. CERNLIB contains routines to implement this method numerically, working from functions or histograms.
2 33. Monte Carlo techniques

Figure 33.1: Use of a random number $u$ chosen from a uniform distribution (0,1) to find a random number $x$ from a distribution with cumulative distribution function $F(x)$.

For a discrete distribution, $F(x)$ will have a discontinuous jump of size $f(x_k)$ at each allowed $x_k$, $k = 1, 2, \cdots$. Choose $u$ from a uniform distribution on (0,1) as before. Find $x_k$ such that

$$F(x_{k-1}) < u \leq F(x_k) \equiv \text{Prob } (x \leq x_k) = \sum_{i=1}^{k} f(x_i) ;$$

then $x_k$ is the value we seek (note: $F(x_0) \equiv 0$). This algorithm is illustrated in Fig. 33.1b.

33.3. Acceptance-rejection method (Von Neumann)

Very commonly an analytic form for $F(x)$ is unknown or too complex to work with, so that obtaining an inverse as in Eq. (33.2) is impractical. We suppose that for any given value of $x$ the probability density function $f(x)$ can be computed and further that enough is known about $f(x)$ that we can enclose it entirely inside a shape which is $C$ times an easily generated distribution $h(x)$ as illustrated in Fig. 33.2.

Frequently $h(x)$ is uniform or is a normalized sum of uniform distributions. Note that both $f(x)$ and $h(x)$ must be normalized to unit area and therefore the proportionality constant $C > 1$. To generate $f(x)$, first generate a candidate $x$ according to $h(x)$. Calculate $f(x)$ and the height of the envelope $Ch(x)$; generate $u$ and test if $uC h(x) \leq f(x)$. If so, accept $x$; if not reject $x$ and try again. If we regard $x$ and $uC h(x)$ as the abscissa and ordinate of a point in a two-dimensional plot, these points
will populate the entire area $Ch(x)$ in a smooth manner; then we accept those which fall under $f(x)$. The efficiency is the ratio of areas, which must equal $1/C$; therefore we must keep $C$ as close as possible to 1.0. Therefore we try to choose $Ch(x)$ to be as close to $f(x)$ as convenience dictates, as in the lower part of Fig. 33.2. This practice is called importance sampling, because we generate more trial values of $x$ in the region where $f(x)$ is most important.

### 33.4. Algorithms

Algorithms for generating random numbers belonging to many different distributions are given by Press [5], Ahrens and Dieter [6], Rubinstein [7], Everett and Cashwell [8], Devroye [9], and Walck [10]. For many distributions alternative algorithms exist, varying in complexity, speed, and accuracy. For time-critical applications, these algorithms may be coded in-line to remove the significant overhead often encountered in making function calls. Variables named “u” are assumed to be independent and uniform on [0,1). (Hence, $u$ must be verified to be non-zero where relevant.)

In the examples given below, we use the notation for the variables and parameters given in Table 31.1.
4 33. Monte Carlo techniques

33.4.1. Exponential decay.

This is a common application of the inverse transform method and uses the fact that if \( u \) is uniformly distributed in \([0, 1]\) then \((1 - u)\) is as well. Consider an exponential p.d.f. \( f(t) = (1/\tau) \exp(-t/\tau) \) that is truncated so as to lie between two values, \( a \) and \( b \), and renormalized to unit area. To generate decay times \( t \) according to this p.d.f., first let \( \alpha = \exp(-a/\tau) \) and \( \beta = \exp(-b/\tau) \); then generate \( u \) and let
\[
    t = -\tau \ln(\beta + u(\alpha - \beta)).
\]
(33.4)

For \((a, b) = (0, \infty)\), we have simply \( t = -\tau \ln u \). (See also Sec. 33.4.6.)

33.4.2. Isotropic direction in 3D:

Isotropy means the density is proportional to solid angle, the differential element of which is \( d\Omega = d(\cos \theta) \phi \). Hence \( \cos \theta \) is uniform \((2u_1 - 1)\) and \( \phi \) is uniform \((2\pi u_2)\). For alternative generation of \( \sin \phi \) and \( \cos \phi \), see the next subsection.

33.4.3. Sine and cosine of random angle in 2D:

Generate \( u_1 \) and \( u_2 \). Then \( v_1 = 2u_1 - 1 \) is uniform \((-1, 1)\), and \( v_2 = u_2 \) is uniform \((0, 1)\). Calculate \( r^2 = v_1^2 + v_2^2 \). If \( r^2 > 1 \), start over. Otherwise, the sine \((S)\) and cosine \((C)\) of a random angle \((i.e., \text{uniformly distributed between zero and } 2\pi)\) are given by
\[
    S = 2v_1v_2/r^2 \quad \text{and} \quad C = (v_1^2 - v_2^2)/r^2.
\]
(33.5)

33.4.4. Gaussian distribution:

If \( u_1 \) and \( u_2 \) are uniform \((0, 1)\), then
\[
    z_1 = \sin 2\pi u_1 \sqrt{-2 \ln u_2} \quad \text{and} \quad z_2 = \cos 2\pi u_1 \sqrt{-2 \ln u_2}
\]
(33.6)

are independent and Gaussian distributed with mean 0 and \(\sigma = 1\).

There are many faster variants of this basic algorithm. For example, construct \( v_1 = 2u_1 - 1 \) and \( v_2 = 2u_2 - 1 \), which are uniform \((-1, 1)\). Calculate \( r^2 = v_1^2 + v_2^2 \), and if \( r^2 > 1 \) start over. If \( r^2 < 1 \), it is uniform \((0, 1)\). Then
\[
    z_1 = v_1 \sqrt{-2 \ln r^2/r^2} \quad \text{and} \quad z_2 = v_2 \sqrt{-2 \ln r^2/r^2}
\]
(33.7)

are independent numbers chosen from a normal distribution with mean 0 and variance 1. \( z_i' = \mu + \sigma z_i \) distributes with mean \(\mu\) and variance \(\sigma^2\). A recent implementation of the fast algorithm of Leva Ref. 11 is in CERNLIB.

For a multivariate Gaussian with an \(n \times n\) covariance matrix \(V\), one can start by generating \(n\) independent Gaussian variables, \(\{\eta_j\}\), with mean 0 and variance 1 as above. Then the new set \(\{x_i\}\) is obtained as \(x_i = \mu_i + \sum_j L_{ij} \eta_j\), where \(\mu_i\) is the mean of \(x_i\), and \(L_{ij}\) are the components of \(L\), the unique lower triangular matrix that fulfils \(V = LL^T\).
The matrix $L$ can be easily computed by the following recursive relation (Cholesky’s method):

\[ L_{jj} = \left( V_{jj} - \sum_{k=1}^{j-1} L_{jk}^2 \right)^{1/2}, \]  
\[ L_{ij} = \frac{V_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk}}{L_{jj}}, \quad j = 1, \ldots, n; \quad i = j + 1, \ldots, n, \]  

(33.8a)  
(33.8b)

where $V_{ij} = \rho_{ij} \sigma_i \sigma_j$ are the components of $V$. For $n = 2$ one has

\[ L = \begin{pmatrix} \sigma_1 & 0 \\ \rho \sigma_2 & \sqrt{1 - \rho^2} \sigma_2 \end{pmatrix}, \]  

(33.9)

and therefore the correlated Gaussian variables are generated as $x_1 = \mu_1 + \sigma_1 \eta_1$, $x_2 = \mu_2 + \rho \sigma_2 \eta_1 + \sqrt{1 - \rho^2} \sigma_2 \eta_2$.

33.4.5. \( \chi^2(n) \) distribution:

For $n$ even, generate $n/2$ uniform numbers $u_i$; then

\[ y = -2 \ln \left( \prod_{i=1}^{n/2} u_i \right) \text{ is } \chi^2(n). \]  

(33.10)

For $n$ odd, generate $\frac{n-1}{2}$ uniform numbers $u_i$ and one Gaussian $z$ as in Sec. 33.4.4; then

\[ y = -2 \ln \left( \prod_{i=1}^{\frac{(n-1)}{2}} u_i \right) + z^2 \text{ is } \chi^2(n). \]  

(33.11)

For $n \geq 30$ the much faster Gaussian approximation for the $\chi^2$ may be preferable: generate $z$ as in Sec. 33.4.4. If $z \geq -\sqrt{2n-1}$ use $y = \left[ z + \sqrt{2n-1} \right]^2 / 2$; otherwise reject.

33.4.6. Gamma distribution:

All of the following algorithms are given for $\lambda = 1$. For $\lambda \neq 1$, divide the resulting random number $x$ by $\lambda$.

- If $k = 1$ (the exponential distribution), accept $x = -\ln u$. (See also Sec. 33.4.1.)

- If $0 < k < 1$, initialize with $v_1 = (e + k)/e$ (with $e = 2.71828\ldots$ being the natural log base). Generate $u_1, u_2$. Define $v_2 = v_1 u_1$. 
Then

\[ x = v_2^{1/k}. \]

If \( u_2 \leq e^{-x} \), accept \( x \) and stop, else restart by generating new \( u_1, u_2 \).

**Case 2:** \( v_2 > 1 \). Define \( x = -\ln([v_1 - v_2]/k) \). If \( u_2 \leq x^{k-1} \), accept \( x \) and stop, else restart by generating new \( u_1, u_2 \). Note that, for \( k < 1 \), the probability density has a pole at \( x = 0 \), so that return values of zero due to underflow must be accepted or otherwise dealt with.

- Otherwise, if \( k > 1 \), initialize with \( c = 3k - 0.75 \). Generate \( u_1 \) and compute \( v_1 = u_1(1 - u_1) \) and \( v_2 = (u_1 - 0.5)\sqrt{c/v_1} \). If \( x = k + v_2 - 1 \leq 0 \), go back and generate new \( u_1 \); otherwise generate \( u_2 \) and compute \( v_3 = 64v_1^3u_2^2 \). If \( v_3 \leq 1 - 2v_2^2/x \) or if \( \ln v_3 \leq 2\{[k - 1] \ln[x/(k - 1)] - v_2 \} \), accept \( x \) and stop; otherwise go back and generate new \( u_1 \).

### 33.4.7. Binomial distribution:

Begin with \( k = 0 \) and generate \( u \) uniform in \([0, 1)\). Compute \( P_k = (1 - p)^n \) and store \( P_k \) into \( B \). If \( u \leq B \) accept \( r_k = k \) and stop. Otherwise, increment \( k \) by one; compute the next \( P_k \); as \( P_k \cdot (p/(1-p)) \cdot (n-k)/(k+1) \); add this to \( B \). Again if \( u \leq B \) accept \( r_k = k \) and stop, otherwise iterate until a value is accepted. If \( p > 1/2 \) it will be more efficient to generate \( r \) from \( f(r; n, q) \), i.e., with \( p \) and \( q \) interchanged, and then set \( r_k = n - r \).

### 33.4.8. Poisson distribution:

Iterate until a successful choice is made: Begin with \( k = 1 \) and set \( A = 1 \) to start. Generate \( u \). Replace \( A \) with \( uA \); if now \( A < \exp(-\mu) \), where \( \mu \) is the Poisson parameter, accept \( n_k = k - 1 \) and stop. Otherwise increment \( k \) by one, generate a new \( u \) and repeat, always starting with the value of \( A \) left from the previous try. For large \( \mu (\geq 10) \) it may be satisfactory (and much faster) to approximate the Poisson distribution by a Gaussian distribution (see our Probability chapter, Sec. 31.4.3). Generate \( z \) from a Gaussian with zero mean and unit standard deviation; then use \( x = \max(0, [\mu + z\sqrt{\mu} + 0.5]) \) where \([\ ]\) signifies the greatest integer \leq \( x \) the expression. [12]

### 33.4.9. Student’s t distribution:

For \( n > 0 \) degrees of freedom (\( n \) not necessarily integer), generate \( x \) from a Gaussian with mean 0 and \( \sigma^2 = 1 \) according to the method of 33.4.4. Next generate \( y \), an independent gamma random variate, according to 33.4.6 with \( \lambda = 1/2 \) and \( k = n/2 \). Then \( z = x/(\sqrt{y}/n) \) is distributed as a \( t \) with \( n \) degrees of freedom.

For the special case \( n = 1 \), the Breit-Wigner distribution, generate \( u_1 \) and \( u_2 \); set \( v_1 = 2u_1 - 1 \) and \( v_2 = 2u_2 - 1 \). If \( v_1^2 + v_2^2 \leq 1 \) accept \( z = v_1/v_2 \) as a Breit-Wigner distribution with unit area, center at 0.0, and FWHM 2.0. Otherwise start over. For center \( M_0 \) and FWHM \( \Gamma \), use \( W = z\Gamma/2 + M_0 \).

### References:

1. F. James, Comp. Phys. Comm. 79 111 (1994), based on M. Lüscher, Comp. Phys. Comm. 79 100 (1994). This generator is available as the CERNLIB routine V115, RANLUX.


11. J.L. Leva, ACM Trans. Math. Softw. 18 449 (1992). This generator has been implemented by F. James in the CERNLIB routine V120, RNORML.

12. This generator has been implemented by D. Drijard and K. Kölblig in the CERNLIB routine V136, RNPSSN.