2 Random Variable Generation

Most Monte Carlo computations require, as a starting point, a sequence of i.i.d. random variables with given marginal distribution. We describe here some of the basic methods that are available for sampling such a sequence.

Notation reminder: For a random variable (RV) X, the cdf (cumulative distribution function) is given by:

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$$

$F(x)$ is a non-decreasing function, continuous from the right, with $\lim_{x\to-\infty} = 0$ and $\lim_{x\to\infty} = 1$. The pdf (probability density function) f is defined by $f(x) = F'(x)$, when the derivative exists. For discrete RVs we use the probability mass function $p(x) = \mathbb{P}(X = x)$. To simplify notation we sometime use $f(x)$ for $p(x)$ as well.

Some common probability distributions are listed at the end of this chapter.

2.1 Random Number Generators

Uniformly distributed samples form the basis for most other sampling distributions. In general, the sampled sequences are *pseudo-random*, namely are generated by a deterministic algorithm but "appear" random. Most software packages have effective random number generators, so the programmer does not have to worry about that. Here we only describe briefly some common schemes.

Linear congruential generators (LCG):

$$X_t = (aX_{t-1} + c) \bmod m; \quad U_t = \frac{X_t}{m}$$

Here a, c, m are positive integers, and X_0 is the *seed*.

Note that (U_i) is periodic, with period m at most.

A baseline choice of parameters is $a = 7^5$, $c = 0$, $m = 2^{31} - 1$. It gives good statistical properties, but its period is too short for most modern purposes.

Monte Carlo Methods – Lecture Notes, N. Shimkin, Spring 2015

Multiple Recursive Generators (MRG):

$$X_t = (a_1 X_{t-1} + \cdots + a_k X_{t-k}) \bmod m, \quad t \geq k$$
$$U_t = \frac{X_t}{m}$$

Here the state is $S_t = (X_t, \ldots, X_{t-k+1})$, and the maximum period is m^k.

Combined Generators: Superior performance is obtained by combining several different MRGs. For example, the MRG32k3a algorithm which is implemented in several software packages including Matlab employs two MRGs of order 3:

$$X_t = (1403580 X_{t-2} - 810728 X_{t-3}) \bmod m_1, \quad m_1 = 2^{32} - 209$$
$$Y_t = (527612 Y_{t-1} - 1370589 Y_{t-3}) \bmod m_2, \quad m_2 = 2^{32} - 22853$$

and combines their output using

$$U_t = \frac{X_t - Y_t + m_1}{m_1 + 1} 1_{\{X_t \leq Y_t\}} + \frac{X_t - Y_t}{m_1 + 1} 1_{\{X_t > Y_t\}}$$

The period length here is approximately 3×10^{57}, and the resulting sequences passes a comprehensive set of statistical tests.

2.2 Inverse-Transform Method

Let

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}, \quad y \in [0, 1]$$

denote the inverse of the cdf F. It is easily verified that, if $U \sim U[0, 1]$, then

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

is an RV with cdf F. Indeed, by definition of F^{-1},

$$F^{-1}(y) \leq x \quad \Leftrightarrow \quad F(x) \geq y$$

so that

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)\,.$$

Therefore, to generate an i.i.d. sequence (X_i) with marginal cdf F, we first generate a sequence (U_i) of (pseudo) random numbers, and apply the inverse transform to each U_i.

Example 1: Exponential Distribution

Here $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$, and $F^{-1}(u) = -\lambda^{-1}\ln(1-u)$, $u \in [0,1)$. Therefore $X = -\lambda^{-1}\ln(1-U)$ will have an exponential distribution. (Note that $1-U$ can be replaced by U since they have the same distribution.)

Example 2: Discrete Distributions

Let X be a discrete RV with $\mathbb{P}(X = x_i) = p_i$, $i = 1, \ldots, n$, and $x_i < x_{i+1}$. The transform method then yields the following scheme:

– generate $U \sim U(0,1)$
– Find the smallest integer k such that $U \leq F(x_k)$
– return x_k

Example 3: Order statistics

Let $X_1, \ldots, X_n$ be iid random variables with cdf F. Let

$$X_{(1)} = \min(X_1, \ldots, X_n), \quad X_{(n)} = \max(X_1, \ldots, X_n)$$

denote the minimal and maximal elements. We wish to generate RVs distributed as $X_{(1)}$ and $X_{(n)}$.

One option is to generate $X_1, \ldots, X_n$, and take the extremal elements. When n is large, a more efficient scheme is the following. Observe that the cdf of $X_{(n)}$ is $F_{(n)}(x) = F(x)^n$. Therefore, for a uniformly distributed RV $U \sim U(0,1)$,

$$X_{(n)} = F_{(n)}^{-1}(U) = F^{-1}(U^{1/n})$$

will have the required distribution. Similarly, the cdf of $X_{(1)}$ is $F_{(1)}(x) = 1-(1-(F(x))^n$, so that

$$X_{(1)} = F^{-1}(1-(1-U)^{1/n})$$

will have the required distribution. Again, we may replace $1-U$ by U in the last formula.

2.3 Acceptance-Rejection Methods

Suppose that we wish to sample from a pdf $f(x)$, the target pdf. Let $g(x)$ be another pdf, the *proposal pdf*, such that $f(x) \leq Cg(x)$ for some $C > 1$.

Consider the following *Acceptance-Rejection Algorithm*:

1. Sample Z from $g(x)$.
2. Sample $U \sim U(0,1)$.
3. If $Cg(Z)U \leq f(Z)$, return $X = Z$. Otherwise, return to step 1.

It is easily verified that the RV X generated according to this algorithm has the required pdf $f(x)$.

Assuming that sampling from $g(x)$ is easy, we obtain a procedure for sampling from $f(x)$. The *efficiency* of the algorithm is defined as

$$\mathbb{P}\{(Z,U) \text{ is accepted}\} = \frac{1}{C}$$

Since the trials are independent, the number of trials till acceptance has a geometric distribution with parameter C^{-1}, and mean C.

Example 4: Gaussian distribution

To generate a $N(0,1)$ RV, we may first generate positive RV X with distribution

$$f(x) = \sqrt{2/\pi}e^{-x^2/2}, \quad x \geq 0$$

and then assign a random sign to X. To sample from the target pdf $f(x)$, we can use as proposal pdf the exponential distribution : $g(x) = e^{-x}$, $x \geq 0$. It may be seen that $f(x) \leq Cg(x)$ for $C = \sqrt{2e/\pi}$.

We therefore sample independent RVs $Z \sim \text{Exp}(1)$, $U \sim U(0,1)$, and accept Z if

$$U \leq \frac{f(Z)}{Cg(Z)} = \exp(-(Z-1)^2/2)$$

2.4 Some Specific Formulae

There are numerous specific formulae and methods that apply to specific distributions. Below are some examples.

Gaussian: Box-Müller approach. If U_1, U_2 are independent $U(0,1)$ RVs, then

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

are two independent $N(0,1)$ RVs.

Binomial: A binomial RV $X \sim \text{Bin}(p,n)$ can be written at the sum $X_1 + \cdots + X_n$ of independent Bern(p) RVs. We can therefore write

$$X = \sum_{i=1}^{n} 1_{\{U_i \le p\}}$$

When n is large, $\frac{1}{n}X - p$ converges to a Normal $N(0, p(1-p)$ RV (by the CLT). Hence X is close to a $N(np, np(1-p))$ RV. We can therefore approximate a binomial RV by generating such a Normal RV, and rounding the result to the nearest non-negative integer. This normal approximation is reasonably accurate starting from $n \max\{p, 1-p\} > 10$.

Geometric: The geometric distribution $f(x) = p(1-p)^{x-1}$, $x = 1, 2 \dots$ may be interpreted as the number of Bernoulli(p) trials till the first success. It may be seen that if $Y \sim \text{Exp}(\lambda)$ with $\lambda = -\ln(1-p)$, then $X = 1 + \lfloor Y \rfloor$ has the required geometric distribution.

2.5 Random Vectors

Suppose we wish to generate a random vector $X = (X_1, \dots, X_n)$ from a given n-dimensional distribution with pdf $F(x)$ or cdf $f(x)$.

In the *sequential* approach, we observe that the joint distribution can be represented as

$$f(x_1, \dots, x_n) = f_1(x_1) f_2(x_2|x_1) \dots f_n(x_n|x_1, \dots, x_{n-1})$$

We can therefore sample sequentially $X_1 \sim f_1$, $X_2 \sim f_2(\cdot|X_1)$, etc.

Feasibility of this approach depends on the ability to calculate the conditional distributions. For Markov models, for example, this is naturally available.

The *Acceptance-Rejection* approach is directly applicable in the vector case as well. An important special case is the following:

Example: Uniform Samples. Suppose we wish to sample uniformly from a set S in $\mathbb{R}^n$, of positive volume. If S is contained in a box B, we can sample X from B and accept it if $X \in S$. The efficiency of this scheme evidently depends on the ratio $\text{vol}(S)/\text{vol}(B)$.

For *Gaussian* random vectors, we can sample from $N(\mu, \Sigma)$ by noting that the required distribution can be obtained by $X = BZ + \mu$, where B satisfies the Cholesky decomposition $\Sigma = BB^T$, and Z is a vector of independent $N(0, 1)$ RVs.

Some specific methods for generating uniform samples from useful sets are listed below.

1. **Unit Sphere:** Let $X_1, \ldots X_n$ be iid Gaussian RVs from $N(0, 1)$. Then the vector
$$Y = \frac{1}{||X||}(X_1, \ldots, X_n)$$
where $||X|| = (\sum_i X_i^2)^{1/2}$, is uniformly distributed over the unit sphere $\{y \in \mathbb{R}^N : ||y|| = 1\}$.

2. **Unit Ball:** To obtain a uniform sample from the unit ball $\{||z|| \leq 1\}$, we can multiply the above uniform sample Y from the unit sphere by $U^{1/n}$, where $U \sim U(0, 1)$.

3. **Unit Simplex:** The unit corner simplex
$$\Delta_c^n = \{y \in \mathbb{R}^n : y_i \geq 0, \sum_i y_i \leq 1\}$$
is the convex hull of the points $0, e_1, \ldots, e_n$. It can of course be sampled uniformly by the rejection method relative to the unit box, but the efficiency decreases quickly in the dimension n. A more efficient method is therefore required for large n.

   Let S be another n-dimensional simplex,
$$S = \{x \in \mathbb{R}^n : x_i \geq 0, x_1 \leq x_2, \ldots, x_n \leq 1\}$$
which is the convex hull of the points $0, e_n, e_n + e_{n-1}, \ldots, \mathbf{1}$, where $\mathbf{1} = e_1 + \cdots + e_n$. The simplex Δ_c^n can be obtained from the simplex S by the linear transformation
$$y_1 = x_1,\ y_2 = x_2 - x_1, \ldots, y_n = x_n - x_{n-1}\,.$$
Drawing uniform samples from S is easily done by sampling n independent $U(0, 1)$ RVs, $U_1, \ldots, U_n$, and reordering them in increasing size, $U_{(1)}, \ldots, U_{(n)}$.

   To obtain a uniform sample from Δ_c^n we now apply the above transformation to $x = (U_{(1)}, \ldots, U_{(n)})$ (note that a *linear transformation preserves uniformity*).

   To obtain uniform samples from the unit n-simplex
$$\Delta^n = \{y \in \mathbb{R}^{n+1} : y_i \geq 0, \sum_i y_i = 1\},$$
we can sample uniformly $Y \in R^n$ from the corner simplex Δ_c^n, and add the $n + 1$ coordinate $y_{n+1} = 1 - \sum_{i=1}^n Y_i$.

4. **General n-Simplex:** Consider a n-dimensional simplex defined by arbitrary $n+1$ vertices, namely S is the convex hull of points $z_0, z_1, \ldots, z_n$. Sampling uniformly from S can be done by sampling Y uniformly from Δ_c^n, and applying the linear transformation

$$Z = MY + z_0,$$

where M is the matrix with columns $z_1 - z_0, \ldots, z_n - z_0$.

5. **Random Permutations:** Recall that $\{1, 2, \ldots, n\}$ has $n!$ different permutations. To sample a random permutation uniformly we may proceed in two ways:

   a. Sample $U(0,1)$ independent RVs, $U_1, \ldots, U_N$, and sort them in increasing order. The resulting index sequence forms a random permutation.

   b. Sample n times sequentially, uniformly and without replacement from $\{1, 2, \ldots, n\}$.

2.6 Appendix: Some Common Probability Distributions

Continuous:

- Uniform: $X \sim U[a,b]$, $b > a$. $f(x) = \frac{1}{b-a}, \quad a \le x \le b$
- Normal: $X \sim N(m, \sigma^2)$, $\sigma > 0$. $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2}), \quad x \in \mathbb{R}$
- Exponential: $X \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$. $f(x) = \lambda e^{-\lambda x}, \quad x \ge 0$
- Gamma: $X \sim \mathrm{Gamma}(a, \lambda)$, $a, \lambda > 0$. $f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x \ge 0$
  $\Gamma(a) \triangleq \int_0^\infty e^{-x} x^{a-1} dx$, $E(X) = a/\lambda$, $\mathrm{Var}(X) = a/\lambda^2$
- Beta: $X \sim \mathrm{Beta}(a,b)$, $a, b > 0$. $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad x \in [0,1]$
- Weibull: $X \sim \mathrm{Weib}(a, \lambda)$, $a, \lambda > 0$. $f(x) = a\lambda(\lambda x)^{a-1} \exp(-(\lambda x)^a), \quad x \ge 0$
- Pareto: $X \sim \mathrm{Pareto}(a, \lambda)$, $a, \lambda > 0$. $f(x) = a\lambda(1+\lambda x)^{-(a+1)}, \quad x \ge 0$

Discrete:

- Bernoulli: $X \sim \mathrm{Ber}(p)$, $0 \le p \le 1$. $p(x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}$
- Binomial: $X \sim \mathrm{Bin}(n,p)$, $0 \le p \le 1$. $p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, \dots, n\}$
  $E(X) = np$, $\mathrm{Var}(X) = np(1-p)$
- Geometric: $X \sim G(p)$,$0 \le p \le 1$. $p(x) = p(1-p)^{x-1}, \quad x = 1, 2, \dots$
  $E(X) = p^{-1}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$
- Poisson: $X \sim \mathrm{Pois}(\lambda)$, $\lambda > 0$. $p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, \dots$
  $E(X) = \lambda$, $\mathrm{Var}(X) = \lambda$