Lecture 2

Random number generation

To generate a random variable from any distribution it is typically enough to be able to generate a uniform distributed random variable. How can one generate a sequence of numbers $u_1, u_2, \ldots$, such that they behave as independent realisations of $U \sim U[0,1]$ (a random variable that is distributed uniformly over $[0,1]$)?

One of the first generators of uniform (pseudo) random variables are linear congruent generators. They were used e.g., in Visual Basic (up to 6) and Rand48 in Unix. Set $m > 0$ (modulus), $a > 0$ (multiplier), $r > 0$ (increment) and z_0 (seed, start value). Then, set

$$z_{n+1} = (a \cdot z_n + r) \bmod m, \quad n = 0, 1, 2, 3, \ldots$$

and define

$$u_n = \frac{z_n}{m}, \quad n = 0, 1, 2, \ldots.$$

Values $u_0, u_1, \ldots$ are in $\{0/m, 1/m, \ldots, (m-1)/m\} \subset [0,1)$. Typically one chooses $m = 2^b$. As soon as a number from $\{0, \ldots, m-1\}$ appears for the second time, that is, as soon as $z_j = z_i$, $i < j$, one would get periodic numbers $z_i, \ldots, z_{j-1}$.

For example, set $m = 2^6 = 64$, $a = 4$, $r = 1$. Then, if $z_0 = 21$, then $z_1 = (4 \cdot 21 + 1) \bmod 64 = 85 \bmod 64 = 21 = z_0$, so that $z_0 = z_1 = \ldots = 21$ and the period is 1.

One can show that a linear congruent generator has a full period $m = 2^b$, $b \geq 2$ if and only if $r \in (0, m)$ is odd and $a \bmod 4 = 1$. However, even if m is large enough and the generator has a full period, it has a drawback to generate random variables that lay in hyperplanes in higher dimensions.

Another similar generator is a multiplicative congruent generator, which is defined for $m > 0$ (modulus), $a > 0$ (multiplier) and z_0 (seed) via $z_{n+1} = a \cdot z_n \bmod m$, $n = 0, 1, 2, \ldots$. This generator has similar drawback as a linear one.

In R, SAS, Matlab, Mathematica, Maple more modern generators are used. For example, the default in R is the Mersenne-Twister generator, see help to `Random` for details and other possible random variables generators.

Once one can generate a sequence of (pseudo) uniform random variables, there are several approaches to random variables generation from other distributions.

The first approach is the **inversion method**.
Let F be a distribution function on $\mathbb{R}$. Function F^{-1} defined by

$$F^{-1}(u) = \inf\{x : F(x) \geq u,\ 0 < u < 1\}$$

is called the quantile function. Note that if F is strictly increasing, then F^{-1} is just a regular inverse.
Theorem
Let F be a distribution function. If $U \sim U[0,1]$, then $F^{-1}(U)$ has a distribution function F.
Proof
We have to show that $P(F^{-1}(U) \leq x) = F(x)$, $x \in \mathbb{R}$. First, let F be continuous and we show equality of two events $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$, so that taking probabilities yields the result: $P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$.
Since F is continuous, $F(F^{-1}(u)) = u$ and therefore, if $F^{-1}(U) \leq x$, then $U \leq F(x)$. Similarly, if $U \leq F(x)$, then $F^{-1}(U) \leq x$. In general case it is easily shown that

$$\{U < F(x)\} \subseteq \{F^{-1}(U) \leq x\} \subseteq \{U \leq F(x)\}.$$

Since U is a continuous random variable with $P(U = F(x)) = 0$, taking probabilities in the above equations yields the result. $\square$

Several examples of continuous distributions:

1. Exponential $F(x) = 1 - \exp(-\lambda x)$, $x > 0$, $F^{-1}(u) = \lambda^{-1}\log(1-u)$, $u \in (0,1)$
2. Pareto $F(x) = 1 - (1+x)^{-\alpha}$, $x > 0$, $F^{-1}(u) = (1-u)^{-1/\alpha} - 1$, $u \in (0,1)$
3. Standard Cauchy $F(x) = 1/2 + \pi^{-1}\arctan(x)$, $F^{-1}(u) = \tan\{\pi(u - 1/2)\}$, $u \in (0,1)$

Simulate U_i and set $X_i = F^{-1}(U_i)$. If F^{-1} can not be inverted analytically, appropriate numerical methods can be applied.

Let X be a discrete random variable with possible values $\{x_1, x_2, \ldots\}$, so that $F(x) = \sum_{i:x_i \leq x} P(X = x_i)$ and

$$F^{-1} = \min\left\{k \in \mathbb{N} : \sum_{j=1}^{k} P(X = x_j) = \sum_{j=1}^{k} p_j \geq r\right\}.$$

Then, the inverse method becomes: set $X = x_1$ if and only if $U_i \in [0, p_1)$ and $X = x_k$ if and only if $U_i \in \left[\sum_{j=1}^{k-1} p_j, \sum_{j=1}^{k} p_j\right)$, $k = 2, 3, \ldots$. Note that

$$P(X_i = x_k) = P\left(\sum_{j=1}^{k-1} p_j \le U_i < \sum_{j=1}^{k} p_j\right) = \sum_{j=1}^{k} p_j - \sum_{j=1}^{k-1} p_j = p_k.$$

For example, to simulate a Bernoulli random variable $Ber(p)$, generate $U \in U[0,1]$ and set $X = 0$, if $U \le 1 - p$ and $X = 1$ if $U > 1 - p$.

Another general approach to random variables generation is the **rejection method** (Accept-Reject). Assume we would like to simulate random variables from a density f. Rejection method assumes the existence of a density g and the knowledge of a constant $c \ge 1$ (in practice we want to have c as close to 1 as possible), such that $f(x) \le cg(x)$, for all x. In contrast to f, we are able to generate random variables from density g (e.g., with the inversion method).

Algorithm: Repeat the following steps:

1. Generate a random variable X from density g
2. Generate a random variable $U \sim U[0,1]$ (independent from X)
3. If $Ucg(X) \le f(X)$, accept X, otherwise reject X.

Remarks

1. $f(X)/\{cg(X)\}$ is a random variable independent on U.
2. $f(X)/\{cg(X)\} \in (0,1]$
3. The number of iterations needed to successfully generate X is itself a random variable, which is geometrically distributed with the success probability $p = P(Ucg(X) \le f(X))$ (=acceptance probability), that is $P(N = n) = (1-p)^{n-1}p$, $n = 1, 2, \ldots$. Hence, the expected number of iterations is $E(N) = 1/p$.

Calculating p we get

$$P\left(U \le \frac{f(X)}{cg(X)}\right) = \int_{-\infty}^{\infty} P\left(U \le \frac{f(X)}{cg(X)} \Big| X = x\right) g(x)dx = \frac{1}{c}\int_{-\infty}^{\infty} f(x)dx = \frac{1}{c}$$

Here we used that $P(U \le x) = x$. Hence, $E(N) = c$ and it makes sense to choose $c = \sup_x\{f(x)/g(x)\}$.

Theoretical justification of the accept-reject method is given in the following theorems.

Theorem 1

Let $c > 0$ be an arbitrary constant, X be a random variable on $\mathbb{R}^d$, $d \geq 1$ with the c.d.f. $G(x)$ and p.d.f. $g(x)$ and $U \sim U[0,1]$, independent on X. Then, $(X, cUg(X)) \sim U[A]$, where $A = \{(x,u) : x \in \mathbb{R}^d, 0 \leq u \leq cg(x)\}$. Vice versa, If $(X,U) \sim U[A]$, then $X \sim G$.

Proof

First note that since g is a density, the area under the graph of cg, that is, the Lebesgue measure of A equals to c. From this follows that the uniform distribution on A has the density $c^{-1}\mathbb{1}_{\{\cdot \in A\}}$. For the first statement, let B be a Borel set $B \subseteq A$ and denote $B_x = \{u : (x,u) \in B\}$. Since X and U are independent, by Tonelli's theorem,

$$P(\{X, cUg(X)\} \in B) = \int \int_{B_x} \frac{1}{cg(x)} du g(x) dx = \frac{1}{c} \int_B du dx.$$

Hence, $(X, cUg(X)) \sim U[A]$.

Now assume that $(X,U) \sim U[A]$. Then, the density of the marginal distribution of X is obtained by integrating out u

$$\int \frac{1}{c} \mathbb{1}_{\{(x,u) \in A)\}} du = \frac{1}{c} \int_0^{cg(x)} du = g(x).$$

Hence, $X \sim G$. $\square$

Theorem 2

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables on $\mathbb{R}^d$, $d \geq 1$ and let $A \subseteq \mathbb{R}^d$ be a Borel set such that $P(X_1 \in A) = p > 0$. Let Y be the first X_i taking values in A. Then

$$P(Y \in B) = \frac{P(X_1 \in A \cap B)}{p}$$

for all Borel sets $B \subset \mathbb{R}^d$. In particular, if X_1 is uniformly distributed in A_0 for $A_0 \supseteq A$, then Y is uniformly distributed in A.

Proof

We have

$$\begin{aligned} P(Y \in B) &= \sum_{i=1}^{\infty} P(X_1 \notin A, \ldots, X_{i-1} \notin A, X_i \in B \cap A) \\ &= \sum_{i=1}^{\infty} (1-p)^{i-1} P(X_1 \in A \cap B) = \frac{1}{1-(1-p)} P(X_1 \in A \cap B). \end{aligned}$$

We have shown that

$$P(Y \in B) = \frac{P(X_1 \in A \cap B)}{P(X_1 \in A)}$$

Now, if X is uniformly distributed in some D, then by definition for any Borel set $C \supseteq D$

$$P(X \in C) = \frac{\int_{C \cap D} dx}{\int_D dx}.$$

Hence, since X_1 is uniformly distributed on A_0, we get

$$P(Y \in B) = \frac{P(X_1 \in A \cap B)}{P(X_1 \in A)} = \frac{\int_{A_0 \cap A \cap B} dx}{\int_{A_0} dx} \frac{\int_{A_0} dx}{\int_{A \cap A_0} dx} = \frac{\int_{A \cap B} dx}{\int_A dx},$$

since $A \subseteq A_0$. That is, Y is uniformly distributed on A. $\square$

From the first statement of Theorem 1 we have that $(X, Ucg(X))$ is uniformly distributed under the curve cg. By Theorem 2 we conclude that random variable $(X, Ucg(X))$ generated by the accept-reject algorithm (at the exit with X accepted) is uniformly distributed under the curve f. By the second statement of Theorem 1 we conclude that X must have density f.