

Generating random variables II

Anatoli Iambartsev

IME-USP

Distribution of Random Variables and Simulation of Random Variables

- Random variable uniquely determined by the cumulative distribution function (cdf) $F(x)$
- There exists pseudo-random number generation for uniformly distributed on $[0, 1]$ r.v.
- Inverse method, if it is possible to calculate explicitly inverse or generalized inverse F^{-1}
- Accept-Reject method

Superposition method (Variant of mixture representation).

Suppose that the cdf of a r.v. that we are interested in can be represented as a composition

$$F(x) = \sum_{k=1}^m c_k F_k(x),$$

where all F_k 's are cdf's, and $c_k > 0$. Obviously,

$$c_1 + \cdots + c_m = 1.$$

Let η be a discrete r.v. with probability distribution

$$\mathbb{P}(\eta = k) = c_k.$$

Superposition method (Compare with mixture representation).

Lemma 4. Let $U, V \sim U[0, 1]$. If using V we generate a value $\eta = k$ of r.v. η , and using U , find ξ s.t. $F_k(\xi) = U$, then such generated ξ has cdf $F(x)$.

Proof.

$$\begin{aligned}\mathbb{P}(\xi \leq x) &= \sum_{k=1}^m \mathbb{P}(\xi \leq x \mid \eta = k) \mathbb{P}(\eta = k) \\ &= \sum_{k=1}^m F_k(x) c_k = F(x)\end{aligned}$$

□

Superposition method.

The generalization to infinite case is obvious.

Example. R.v. ξ is defined on the interval $[0, 1]$ and has cdf

$$F(x) = \sum_{k=1}^{\infty} c_k x^k, \quad c_k > 0.$$

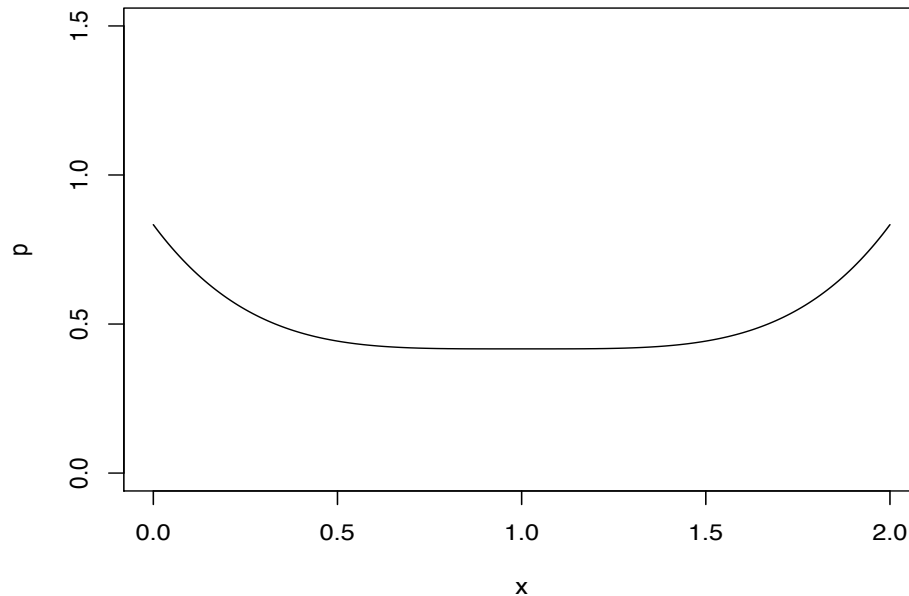
Here we can consider $F_k(x) = x^k, x \in [0, 1]$, and by the method of superposition we obtain:

$$\text{if } \sum_{i=1}^{k-1} c_i \leq V < \sum_{i=1}^k c_i, \quad \text{then } \xi = (U)^{1/k}.$$

Superposition method.

Example. Let $\xi \in [0, 2]$ with density

$$p(x) = \frac{5}{12}(1 + (x - 1)^4), \quad x \in [0, 2].$$



Superposition method.

Example. Let $\xi \in [0, 2]$ with density

$$p(x) = \frac{5}{12}(1 + (x - 1)^4), \quad x \in [0, 2].$$

The inverse method gives the equation

$$(\xi - 1)^5 + 5\xi = 12u - 1,$$

and we should resolve the equation of 5th order.....
it is difficult

Superposition method.

Example. The density $p(x)$ can be represented

$$p(x) = \frac{5}{12}(1 + (x - 1)^4) = \frac{5}{6}p_1(x) + \frac{1}{6}p_2(x), \quad x \in [0, 2],$$

where

$$p_1(x) = \frac{1}{2}, \quad x \in [0, 2], \quad p_2(x) = \frac{5}{2}(x - 1)^4.$$

Then, with $U, V \sim U[0, 1]$

$$\xi = \begin{cases} 2U, & \text{if } V < 5/6, \\ 1 + (2U - 1)^{1/5}, & \text{if } V \geq 5/6. \end{cases}$$

Superposition method.

Lemma 4 utilizes two uniform random variables in order to simulate r.v. with distribution $F(x) = \sum_{i=1}^m c_i F_i(x)$. The next lemma shows that we can use only one uniform random variable.

Lemma 5.* Using $U \sim U[0, 1]$ generate value $\eta = k$ of r.v. η , and after that define ξ from the equation $F(\xi) = \theta$, where

$$\theta = \frac{1}{c_k} \left(U - \sum_{i=1}^{k-1} c_i \right),$$

then the cdf of generated ξ is $F(x) = \sum_{i=1}^m c_i F_i(x)$.

Proof. It is enough to prove that θ is uniformly distributed on the interval $[0, 1]$:

$$(\theta < y \mid \eta = k) = y).$$

□

*Mikhailov, G.A. *On the question of efficient algorithms for modeling of random variables.* (Russian) USSR Computational Mathematics and Mathematical Physics, 1966, **6**:6, 26973.

Superposition method.

Example. In the previous example with $U, V \sim U[0, 1]$ the random variable

$$\xi = \begin{cases} 2U, & \text{if } V < 5/6, \\ 1 + (2U - 1)^{1/5}, & \text{if } V \geq 5/6, \end{cases}$$

can be represented by the last lemma as

$$\xi = \begin{cases} \frac{12}{5}V, & \text{if } V < 5/6, \\ 1 + (12V - 11)^{1/5}, & \text{if } V \geq 5/6. \end{cases}$$

Generation multi-dimensional r.v.s.

Let $Q = (X_1, \dots, X_n)$ be a vector with independent components, then

$$F_Q(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n),$$

where F_i is cdf of component X_i . Here we generate vector Q simulating X_i independently.

Generation multi-dimensional r.v.s.

Example. Simulate (X, Y) uniform on the disc

$$\{(x, y) : x^2 + y^2 \leq 1\}.$$

Generation multi-dimensional r.v.s.

Example. Simulate (X, Y) uniform on the disc

$$\{(x, y) : x^2 + y^2 \leq 1\}.$$

Use polar coordinates (R, Θ) . These are independent and s.t. $\Theta \sim U[0, 2\pi]$ and R has pdf $f_R(r) = 2r, r \in [0, 1]$. It gives cdf $F_R(r) = r^2, r \in [0, 1]$ and inverse $F^{-1}(r) = \sqrt{r}$. Thus, if $U, V \sim U[0, 1]$ are uniform, then

$$\Theta = 2\pi U, \quad R = \sqrt{V}.$$

Generation multi-dimensional r.v.s.

$$p_Q(x_1, \dots, x_n) = p_1(x_1)p_2(x_2 | x_1)p_3(x_3 | x_1, x_2) \dots p_n(x_n | x_1, \dots, x_{n-1})$$

$$p_1(x_1) = \int \dots \int p_Q dx_2 \dots dx_n$$

$$p_2(x_2 | x_1) = [p_1(x_1)]^{-1} \int \dots \int p_Q dx_3 \dots dx_n$$

$$p_3(x_3 | x_1, x_2) = [p_1(x_1)p_2(x_2 | x_1)]^{-1} \int \dots \int p_Q dx_4 \dots dx_n$$

...

$$p_{n-1}(x_{n-1} | x_1, \dots, x_{n-2}) = [p_1(x_1) \dots p_{n-2}(x_{n-2} | x_1, \dots, x_{n-3})]^{-1} \int p_Q dx_n$$

$$p_n(x_n | x_1, \dots, x_{n-1}) = [p_1(x_1) \dots p_{n-1}(x_{n-1} | x_1, \dots, x_{n-2})]^{-1} p_Q$$

Generation multi-dimensional r.v.s.

Let

$$F_i(x_i | x_1, \dots, x_{i-1}) = \int_{-\infty}^{x_i} p_i(x | x_1, \dots, x_{i-1}) dx$$

Lemma 3. Let U_1, \dots, U_n be i.i.d. $U[0, 1]$. Then $Q = (X_1, \dots, X_n)$ s.t.

$$\begin{aligned} X_1 &= F_1^{-1}(U_1), \quad X_2 = F_2^{-1}(U_2 | X_1), \dots, \\ X_n &= F_n^{-1}(U_n | X_1, \dots, X_{n-1}) \end{aligned}$$

has the density $p_Q(x_1, \dots, x_n)$.

Generation multi-dimensional r.v.s.

Example. Simulate (X, Y) uniform on the triangle T with corners in $(0, 1)$, $(0, 0)$ and $(1, 0)$, i.e.

$$T = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1 - x\}.$$

The joint density

$$f(x, y) = \begin{cases} 2, & \text{if } (x, y) \in T, \\ 0, & \text{if } (x, y) \notin T. \end{cases}$$

gives us the marginal

$$f_X(x) = \int_0^{1-x} f(x, y) dy = 2 \int_0^{1-x} dy = 2(1 - x).$$

Generation multi-dimensional r.v.s.

Example. The marginal density

$$f_X(x) = \int_0^{1-x} f(x, y) dy = 2 \int_0^{1-x} dy = 2(1-x).$$

gives us the marginal cdf

$$F_X(x) = \int_0^x 2(1-z) dz = 2x - x^2, \quad x \in [0, 1]$$

and inverse

$$F^{-1}(t) = 1 - \sqrt{1-t}, \quad t \in [0, 1]$$

(note that when you solve the quadratic equation you choose only one correct solution with “-” within interval $[0, 1]$)

Generation multi-dimensional r.v.s.

Example. Further, the conditional pdf of Y given $X = x$ is

$$f_Y(y | x) = \frac{f(x, y)}{f_X(x)} = \frac{1}{1 - x},$$

i.e. $Y | X = x \sim U[0, 1 - x]$.

The inverse transformation method thus gives that if U and V are independent uniform $[0, 1]$, then

$$\begin{aligned} X &= 1 - \sqrt{1 - U} \\ Y &= V(1 - X) \end{aligned}$$

gives a pair (X, Y) which is uniform on the triangle T .

Change of variables method.

[S, p.58] Sometimes we can simplify formulas of modelling of multi-dimensional random variables by choosing new coordinate system.

The rule of density transformation under changing of variables: let $y_i = g_i(x_1, \dots, x_n), i = 1, \dots, n$, one-to-one differentiable transformation of an area B in the space x_1, \dots, x_n into an area B' em the space y_1, \dots, y_n . Let $p_Q(x_1, \dots, x_n)$ be the density of a random vector $Q = (\xi_1, \dots, \xi_n)$ in the area B , then the density of the vector $Q' = (\eta_1, \dots, \eta_n)$ in B' , where $\eta_i = g_i(\xi_1, \dots, \xi_n)$, is

$$p_{Q'}(y_1, \dots, y_n) = p_Q(x_1, \dots, x_n) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|$$

in the right-hand side x_i should be expressed by $(y_i)_{i=1, \dots, n}$.

Change of variables method. Box-Muller.

Example. (RC, Example 2.3) Generate standard normal r.v.s $X, Y \sim N(0, 1)$. Consider a transformation to polar coordinates: $(x, y) \rightarrow (d, \theta)$

$$\begin{cases} d &= x^2 + y^2 \\ \theta &= \tan^{-1}\left(\frac{y}{x}\right). \end{cases}$$

To get the joint distribution of d and θ need Jacobian of the transformation

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

Change of variables method. Box-Muller.

Example. (RC, Example 2.3) Since $f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ then

$$f_{d,\theta}(d, \theta) = \frac{1}{2\pi} e^{-d/2} \cdot \frac{1}{2} = \frac{e^{-d/2}}{2} \cdot \frac{1}{2\pi}$$

for $0 < d < \infty$ and $0 < \theta < 2\pi$. It means that d and θ are independent. Furthermore

$$d \sim \text{Exp}(1/2), \quad \theta \sim U[0, 2\pi].$$

Thus, if $U, V \sim U[0, 1]$ are independent, then the variables defined by

$$X = \sqrt{-2 \ln(U)} \cos(2\pi V), \quad Y = \sqrt{-2 \ln(U)} \sin(2\pi V),$$

are independent and have standard normal distribution.

Change of variables method. Box-Muller.

[RC, p 47] “ ... the Box-Muller algorithm is exact, producing two normal random variables from two uniform r.v.s, the only drawback (in speed) being the necessity of calculating transcendental functions s.t. \ln , \cos and \sin .”

References:

[RC] Cristian P. Robert and George Casella. *Introducing Monte Carlo Methods with R*. Series "Use R!". Springer

[S] Sobol, I.M. *Monte-Carlo numerical methods*. Nauka, Moscow, 1973. (In Russian)