STA 5106
Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 18
November 5, 2019

Continuous Random Variables

- **Continuous random variables** take all real values (not discrete, countable values) according to a pre-defined probability density function.

- There are two commonly used techniques to simulate a given continuous random variable:

  1. Inverse Transform Method

  2. Acceptance/Rejection Method

Inverse Transform Method

- **Proposition 3** Let U be uniform in [0, 1]. If for any random variable with continuous cumulative distribution function F we define a random variable Y according to

$$Y = F^{-1}(U),$$

then the random variable Y has distribution function F.

Proof: The probability distribution function of Y is given by

$$F_Y(a) = P\{Y \le a\} = P\{F^{-1}(U) \le a\} \ .$$

Since F is a monotonically increasing function

$$F^{-1}(U) \le a \Leftrightarrow U \le F(a)$$

and, therefore, $F_Y(a) = P\{U \le F(a)\} = F(a)$ since U is uniform in [0, 1].

Exponential Random Variable

- The density and distribution functions of an **exponential random variable** are given by

$$f(x) = \exp(-x),\ x \ge 0, \quad F(x) = 1 - \exp(-x),\ x \ge 0\ .$$

- To find $F^{-1}(U)$, let $U = F(x)$ and solving for x provides

$$x = F^{-1}(U) = -\log(1 - U)\ .$$

- Note that if U is uniform on [0, 1], then $1 - U$ is also uniform on [0, 1]. Hence, $x = -\log(U)$ is an exponential random variable with mean 1.

- To generate an exponential random variable with mean λ utilize $x = -\lambda \log(U)$.

Acceptance/Rejection Method

- In this case we assume that we have a method for simulating from some density function g and our task is to utilize samples from g to simulate from a given density function f.
- The basic idea is to simulate from g and accept the samples with probability proportional to the ratio $f/[Cg]$, where C is a constant such that $f(Y)/g(Y) \leq C$, for all Y.
- Simulation procedure:

  (i) Simulate Y from the density g and simulate U from U[0, 1].

  (ii) If $U \leq f(Y)/[Cg(Y)]$ then $X = Y$, else go to step (i).
- Note: If C is large, then the simulation process will be slow.

Mathematical Proof

- **Proposition 4** X is a random variable with density f.

Proof: Let X be the value obtained. Then,

$$P(X \le x) = P(Y \le x \,|\, U \le \frac{f(Y)}{Cg(Y)})$$

$$= P(Y \le x, U \le \frac{f(Y)}{Cg(Y)}) / K$$

where $K = P(U \le f(Y)/[Cg(Y)])$.

Since Y and U are independent random variables and U is uniform in [0, 1], their joint density function is the product of the marginal

$$g(Y) \times 1.$$

Mathematical Proof

Therefore,

$$P(X \le x) = \frac{1}{K} \int_{Y \le x} \int_{U \le \frac{f(Y)}{Cg(Y)}} g(Y) dU dY$$

$$= \frac{1}{K} \int_{-\infty}^{x} \left(\int_{0}^{\frac{f(Y)}{Cg(Y)}} dU \right) g(Y) dY$$

$$= \frac{1}{K} \int_{-\infty}^{x} \frac{f(Y)}{Cg(Y)} g(Y) dY = \frac{1}{KC} \int_{-\infty}^{x} f(Y) dY$$

For x $\rightarrow \infty$, the left side goes to 1 and the integral on the right side also goes to 1. Therefore, $CK = 1$ and

$$P(X \le x) = \int_{-\infty}^{x} f(Y) dY$$

Hence, X is random with probability density f.

Standard Normal Density Function

- We illustrate the use of **acceptance/rejection method** by generating sample from **standard normal density function**.

- As a first step we will simulate from the density function given by
$$f(x) = 2 \cdot \frac{1}{\sqrt{2\pi}} \exp(\frac{-x^2}{2}), \quad x \geq 0.$$

- We will also assume that we have tools to sample from the standard exponential density function which becomes our g for the above discussion ($g(x) = \exp(-x)$).

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

Algorithm

- And

$$\frac{f(x)}{Cg(x)} = \exp\left(-\frac{(x-1)^2}{2}\right)$$

- To generate a random variable with density *f* the following algorithm is used:

  (i) Generate Y, an exponential random variable with mean 1, and U, a uniform [0, 1] random variable.

  (ii) If

$$U \le \exp\left(-\frac{(Y-1)^2}{2}\right),$$

  set $X = Y$, otherwise return to (i).

Algorithm

- Having generated a random variable which is the absolute value of a standard normal, we can generate sample from standard normal according to the following algorithm.

  (i) Generate U a uniform random variable between [0, 1] and generate X according to the algorithm described above.

  (ii) If $U \in (0, 1/2]$ set $Z = X$, else set $Z = -X$.

- To generate Y, a normal random variable with mean μ and standard deviation σ, generate X a standard normal random variable and set

$$Y = \sigma X + \mu.$$

Polar Method

- There is another method used popularly to generate samples from standard normal density. It is also called **Box-Muller** method.

- If X and Y are independent and standard normal random variables then for

$$\theta = \tan^{-1}(Y/X), \quad R = \sqrt{X^2 + Y^2}$$

θ is uniform in $[0, 2\pi]$ and R^2 is exponential with mean 2.

- To reverse this result, if U_1 and U_2 are uniform in [0, 1] then for

$$R = (-2\log(U_1))^{1/2}, \theta = 2\pi U_2,$$

and

$$X = R\cos(\theta), Y = R\sin(\theta)$$

X and Y are independent samples from standard normal density.

Drawback

- **Drawback**: the method involves computing trigonometric functions which is always computationally expensive.

- This method can be modified in the following way to avoid computing sines and cosines.

- If U_1 and U_2 are uniform in [0, 1] then
$$V_1 = 2U_1 - 1,\ V_2 = 2U_2 - 1,$$
are uniform in [−1, 1].

- They may or may not lie in the circle of radius 1 and centered at (0, 0). Generate the pair (V_1, V_2) until it lies in this circle and let $(\overline{R}, \overline{\theta})$ be the polar coordinates of this pair.

Independent Sample

- It can be shown that $\bar{R}^2$ is uniform in [0, 1] and $\bar{\theta}$ is uniform in [0, 2π].

- For these quantities

$$\sin(\bar{\theta}) = \frac{V_2}{\bar{R}}, \quad \cos(\bar{\theta}) = \frac{V_1}{\bar{R}}.$$

- Therefore, if U is a uniform [0, 1] random variable independent of $\bar{\theta}$, then

$$X = (-2\log(U))^{1/2}\frac{V_1}{\bar{R}}, \quad Y = (-2\log(U))^{1/2}\frac{V_2}{\bar{R}}$$

are independent samples from standard normal density.

Independent Sample

- Since $\overline{R}^2$ is uniform in [0, 1] and independent of $\overline{\theta}$, it can be used in place of U.

- Hence X and Y can be generated according to the equations,

$$X = (-2\log(\overline{R}^2))^{1/2}\frac{V_1}{\overline{R}}$$

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

Algorithm

- Generate independent samples from standard normal:

  (i) Generate independent, uniform [0, 1] random variables.

  (ii) Set $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, and $S = V_1^2 + V_2^2$.

  (iii) If $S > 1$, return to (i).

  Else

$$T = \sqrt{\frac{-2\log(S)}{S}}$$

  and $X = TV_1$ and $Y = TV_2$.