Generating Random Variables

Saravanan Vijayakumaran
sarva@ee.iitb.ac.in

Department of Electrical Engineering
Indian Institute of Technology Bombay

March 21, 2025

Generating Random Variables

- Applications where random variables need to be generated
  - Simulations
  - Lotteries
  - Computer Games
- General strategy for generating an arbitrary random variable
  - Generate uniform random variables in the unit interval
  - Transform the uniform random variables to obtain the desired random variables

Generating Uniform Random Variables

- $X \sim \mathcal{U}[a,b]$ has density function

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- The distribution function is

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

- $Y \sim \mathcal{U}[0,1]$ has distribution function

$$F_Y(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

- Given Y, can we generate X?
- $(b-a)Y + a$ has the same distribution as $\mathcal{U}[a,b]$

Generating $\mathcal{U}[0, 1]$

- Computers can represent reals upto a finite precision
- Generate a random integer X from 0 to some positive integer m
- Generate the uniform random variable in $[0, 1]$ as

$$U = \frac{X}{m}$$

- The linear congruential method for generating integers from 0 to m

$$X_{n+1} = (aX_n + c) \bmod m, \quad n \geq 0$$

  where m, a, c are integers called the modulus, multiplier and increment respectively. X_0 is called the starting value.
- For $m = 10$ and $X_0 = a = c = 7$, the sequence generated is

$$7, 6, 9, 0, 7, 6, 9, 0, \cdots$$

- The linear congruential method is eventually periodic

Maximal Period Linear Congruential Generators

$$X_{n+1} = (aX_n + c) \bmod m, \quad n \geq 0$$

Theorem

The linear congruential sequence has period m if and only if

- *c is relatively prime to m*
- $b = a - 1$ *is a multiple of p, for every prime p dividing m*
- *b is a multiple of 4, if m is a multiple of 4.*

Remarks

- Having maximal period is not a guarantee of randomness
- For $a = c = 1$, we have $X_{n+1} = (X_n + 1) \bmod m$
- Additional tests are needed (see reference on last slide)

Generating a Bernoulli Random Variable

- The probability mass function is given by

$$P[X = x] = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

  where $0 \leq p \leq 1$
- Generate a uniform random variable $U \sim \mathcal{U}[0, 1]$
- Generate the Bernoulli random variable by the following rule

$$X = \begin{cases} 1 & \text{if } U \leq p \\ 0 & \text{if } U > p \end{cases}$$

- How can we generate a binomial random variable?

The Inverse Transform Method

- Suppose we want to generate a random variable with distribution function F. Assume F is one-to-one.
- Generate a uniform random variable $U \sim \mathcal{U}[0, 1]$
- $X = F^{-1}(U)$ has the distribution function F

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

Example (Generating Exponential RVs)

X is an exponential RV with parameter $\lambda > 0$ if it has distribution function

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

How can it be generated?

Generating Discrete Random Variables

- Suppose we want to generate a discrete random variable X with distribution function F. F is usually not one-to-one.
- Let $x_1 \leq x_2 \leq x_3 \leq \cdots$ be the values taken by X
- Generate a uniform random variable $U \sim \mathcal{U}[0, 1]$
- Generate X according to the rule

$$X = \begin{cases} x_1 & \text{if } 0 \leq U \leq F(x_1) \\ x_k & \text{if } F(x_{k-1}) < U \leq F(x_k) \text{ for } k \geq 2 \end{cases}$$

Example (Generating Binomial RVs)

The probability mass function of a Binomial RV X with parameters n and p is

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{if } 0 \leq k \leq n$$

How can it be generated?

Box-Muller Method for Generating Gaussian RVs

1. Generate two independent uniform RVs U_1 and U_2 between 0 and 1
2. Let $V_1 = 2U_1 - 1$ and $V_2 = 2U_2 - 1$
3. Let $S = V_1^2 + V_2^2$.
4. If $S \geq 1$, go to Step 1
5. If $S < 1$, let
$$X_1 = V_1\sqrt{\frac{-2\ln S}{S}}, \quad X_2 = V_2\sqrt{\frac{-2\ln S}{S}}$$
6. X_1 and X_2 are independent standard Gaussian random variables

Proof

- (V_1, V_2) represents a random point in the unit circle
- Let $V_1 = R\cos\Theta$ and $V_2 = R\sin\Theta$
- $\Theta \sim \mathcal{U}[0, 2\pi]$ and $R^2 = S \sim \mathcal{U}[0, 1]$. Θ and S are independent
- $X_1 = \sqrt{-2\ln S}\cos\Theta$ and $X_2 = \sqrt{-2\ln S}\sin\Theta$
- X_1, X_2 also are in polar coordinates with radius $R' = \sqrt{-2\ln S}$ and angle Θ

Proof Continued

- The probability density function of R' is $f_R(r) = re^{-r^2/2}$

$$\Pr\left[R' \le r\right] = \Pr\left[\sqrt{-2\ln S} \le r\right] = \Pr\left[S \ge e^{-r^2/2}\right] = 1 - e^{-r^2/2}$$

- The joint probability distribution of X_1 and X_2 is given by

$$\begin{aligned} P(X_1 \le x_1, X_2 \le x_2) &= \int_{\{(r,\theta)|r\cos\theta\le x_1, r\sin\theta\le x_2\}} \frac{1}{2\pi} e^{-\frac{r^2}{2}}\, r\, dr\, d\theta \\ &= \frac{1}{2\pi}\int_{\{x\le x_1, y\le x_2\}} e^{-\frac{x^2+y^2}{2}}\, dx\, dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_1} e^{-\frac{x^2}{2}}\, dx \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x_2} e^{-\frac{y^2}{2}}\, dy \end{aligned}$$

- This proves that X_1 and X_2 are independent and have standard Gaussian distribution

Acceptance-Rejection Method

- Suppose we want to generate a random variable X having density f
- Suppose X is difficult to generate using the inversion method
- Suppose there is a random variable Y with density g which is easy to generate
- For some $c \in \mathbb{R}$, suppose f and g satisfy

$$\frac{f(y)}{cg(y)} \leq 1 \text{ for all } y.$$

- Generate a uniform random variable $U \sim \mathcal{U}[0, 1]$
- Generate the random variable Y
- If $U \leq \frac{f(Y)}{cg(Y)}$, set $X = Y$. Otherwise, generate another pair (U, Y) and keep trying until the inequality is satisfied
- To show that the method is correct, we have to show that

$$P\left(Y \leq x \middle| U \leq \frac{f(Y)}{cg(Y)}\right) = F(x)$$

where $F(x) = \int_{-\infty}^{x} f(t)\, dt$

Example of Acceptance-Rejection Method

- Suppose we want to generate a random variable X with probability density function

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1$$

- We need a pdf $g(x)$ such that $\frac{f(x)}{g(x)} \leq c$ for some $c \in \mathbb{R}$
- Consider $g(x) = 1$ for $0 < x < 1$

$$\frac{f(x)}{g(x)} = 20x(1-x)^3 \leq 20 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^3 = \frac{135}{64}$$

- Let $c = \frac{135}{64} \implies \frac{f(x)}{cg(x)} = \frac{256}{27}x(1-x)^3$
- X can now be generated as follows

  1. Generate $U \sim \mathcal{U}[0,1]$ and $Y \sim \mathcal{U}[0,1]$
  2. If $U \leq \frac{256}{27}Y(1-Y)^3$, set $X = Y$
  3. Otherwise, return to step 1

Reference

- Chapter 3, *The Art of Computer Programming, Seminumerical Algorithms (Volume 2)*, Third Edition, Pearson Education, 1998.
- Karl Sigman, Acceptance-Rejection Method, 2007. `http://www.columbia.edu/~ks20/4703-Sigman/4703-07-Notes-ARM.pdf`