Two Types of Approaches

- Direct
  - Obtain an analytical expression
  - Inverse transform
    - Requires inverse of the distribution function
  - Composition & Convolution
    - For special forms of distribution functions
- Indirect
  - Acceptance-rejection

Composition

- Is applied when the distribution F can be expressed as a combination of other distributions $F_1, F_2, \ldots F_n$

$$F(x) = \sum_{j=1}^{\infty} p_j F_j(x),$$

$$\sum_{j=1}^{\infty} p_j = 1, p_j \geq 0$$

- Equivalent to say if X has density function f such that $f(x) = \sum_{j=1}^{\infty} p_j f_j(x),$

That corresponds to decomposing f into its convex combination representation

Composition

- The decomposition can also be seen as dividing the area under $\boldsymbol{f}$ into regions of areas $p_1, p_2, \ldots p_n$ then determine F_j for each j then apply the inverse method on each one.
- The trick is to find Fj's that are easy and fast to gernerate
- Sometimes we can use geometry of distribution to suggest a decomposition

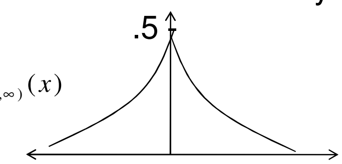
Composition

- The decomposition can also be seen as dividing the area under $\boldsymbol{f}$ into regions of areas $p_1, p_2, \ldots p_n$ then determine F_j for each j then apply the inverse method on each one.
- The trick is to find Fj's that are easy and fast to gernerate
- Sometimes we can use geometry of distribution to suggest a decomposition
- Algorithm
  1. Generate a positive random integer, such that $P(J=j)=p_j$
  2. Return X with distribution F_j

Composition Example1

- For ***0<a<1***, the right trapezoidal distribution has density

$$f(x) = 0.5e^{|x|} \qquad \forall x \quad real$$



Decompose $f(x) = .5e^{x} I_{(-\infty,0)}(x) + .5e^{-x} I_{[0,\infty)}(x)$

$$I_A = \begin{cases} 1 & x \in A \\ 0 & O.w. \end{cases}$$

$$f_1(x) = e^{x} I_{[0,1]}(x)$$

$$f_2(x) = e^{-x} I_{[0,1]}(x)$$

$$U_1 = F_1(x) = e^{x}, \quad U_2 = F_2(x) = e^{-x}$$

Use the inverse method

$$x = \ln U_1 \quad or \quad x = \ln U_2$$

Composition Example1

Algorithm:

1) Generate $U_1 \sim U(0,1), U_2 \sim U(0,1)$

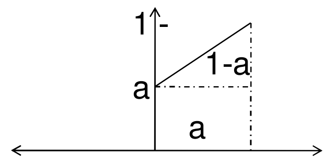
If U_1 <.5 return $x = \ln U_2$

U_1 >=.5 $x = \ln U_2$

Composition Example2

- For ***0<a<1***, the right trapezoidal distribution has density

$$f(x) = \begin{cases} a + 2(1-a)x, 0 \le x \le 1 \\ 0, otherwise \end{cases},$$



- We may divide the area as shown
- ***f(x)*** can be decomposed as

$$f(x) = aI_{[0,1]}(x) + (1-a)2xI_{[0,1]}(x)$$

$f_1(x) = I_{[0,1]}(x)$ is U(0,1) density and

$f_2(x) = 2xI_{[0,1]}(x)$ is a right triangular density

$$p_1 = a,\ p_2 = 1-a,\ p_1 + p_2 = 1$$

Composition Example2

- ***Algorithm: Generate*** $U_1 \sim U(0,1), U_2 \sim U(0,1)$
- ***If***

$U_1 < a\ return\ x = U_1$

$U_2 \ge a \quad f_2 = 2x$

$$U = F_2(x) = x^2$$

$$x = \sqrt{U_2}$$

$$return\ \sqrt{U_2}$$

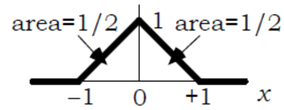
- ***Yet, in some applications, we find computing the square root is expensive So we may use another random number instead U_3 and return x=max(U_2 ,U_3)***

Symmetric Triangle distribution

- The Symmetric triangular distribution on [–1, +1]:

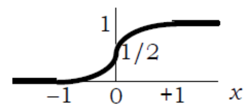
has density

$$f(x)=\begin{cases} x+1 & \text{if } -1\le x\le 0 \\ -x+1 & \text{if } 0\le x\le +1 \\ 0 & \text{otherwise}\end{cases}$$



and CDF

$$F(x)=\begin{cases} 0 & \text{if } x<-1 \\ x^2/2+x+1/2 & \text{if } -1\le x\le 0 \\ -x^2/2+x+1/2 & \text{if } 0<x\le +1 \\ 1 & \text{if } x>+1\end{cases}$$



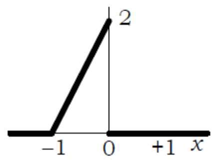
Inverse-transform:

$$U=F(X)=\begin{Bmatrix} X^2/2+X+1/2 & \text{if } U<1/2 \\ -X^2/2+X+1/2 & \text{if } U\ge 1/2\end{Bmatrix}; \text{ solve for}$$

$$X=\begin{Bmatrix} \sqrt{2U}-1 & \text{if } U<1/2 \\ 1-\sqrt{2(1-U)} & \text{if } U\ge 1/2\end{Bmatrix}$$

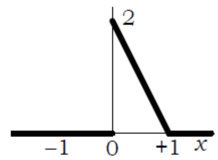
Composition: Define *indicator function* for the set A as $I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$

$$\begin{aligned} f(x) &= (x+1)I_{[-1,0]}(x) + (-x+1)I_{[0,+1]}(x) \\ &= \underbrace{0.5}_{p_1}\underbrace{\{2(x+1)I_{[-1,0]}(x)\}}_{f_1(x)} + \underbrace{0.5}_{p_2}\underbrace{\{2(-x+1)I_{[0,+1]}(x)\}}_{f_2(x)} \end{aligned}$$



$F_1(x) = x^2 + 2x + 1$

$F_1^{-1}(U) = \sqrt{U} - 1$



$F_2(x) = -x^2 + 2x$

$F_2^{-1}(U) = 1 - \sqrt{1-U}$

Composition algorithm:

1. Generate $U_1, U_2 \sim \mathrm{U}(0,1)$ independently

2. If $U_1 < 1/2$, return $X = \sqrt{U_2} - 1$

Otherwise, return $X = 1 - \sqrt{1 - U_2}$

(Can eliminate $\sqrt{\ }$'s)

Convolution

- Assume that $X = Y_1 + Y_2 + ... + Y_m$ (where the ***Y***s are IID with CDF), called m-fold convolution of the distributions of Y's
- Algorithm

1. Generate $Y_1, Y_2, ..., Y_m$ IID each with CDF

2. Return $X = Y_1 + Y_2 + ... + Y_m$

Erlang-*m* Distribution

The Erlang-*m* random variable with mean β, then

$$X = Y_1 + Y_2 + \ldots + Y_m \ ,$$

where the Y_i's are IID exponential random variables, each with mean β/m. If we use the inverse-transform method to generate the exponential Y_i's, then

$$X = \sum_{i=1}^{m} Y_i = \sum \frac{-\beta}{m} \ln U_i = \frac{-\beta}{m} \ln\left(\prod_{i=1}^{m} U_i \right).$$

Erlang-*m* Distribution

Hence we generate random variate x as follows:

1. Generate $u_1, u_2, \ldots, u_m$ as IID $U(0,1)$.
2. Return $x = \frac{-\beta}{m} \ln\left(\prod_{i=1}^{m} u_i \right)$.

Remark:

- Composition: Expressed the distribution function (or density or mass) as a (weighted) sum of other distribution functions (or densities or masses)
- Convolution: Express the random variable itself as the sum of other random variables

Normal Distribution

- Note that if $X \sim N(0,1)$

$$\Rightarrow \mu + \sigma X \sim N(\mu, \sigma)$$

- If we can generate unit normal, then we can generate any normal

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \quad -\infty < x < \infty$$

$$\phi(x) = F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-x^2}{2}} dx$$

- Neither the distribution function nor the density function is invertible
- Use indirect method

Normal: Box-Muller

- Algorithm
  1. Generate independent $U_1, U_2 \sim U(0,1)$
  2. Set $X_1 = \sqrt{-2\ln U_1}\cos 2\pi U_2$,
     $X_2 = \sqrt{-2\ln U_1}\sin 2\pi U_2$
  3. Return X_1
- Each one X_1 or X_2 may be used. Each one of them is an I.I.D N(0,1)
- Technically, independent N(0,1), but serious problem if used with LCGs

Normal: Box-Muller

- For any normal distribution $N(\mu, \sigma^2)$
- Generate $S_i = \sigma X_i + \mu$

What is Monte Carlo Simulation ?

- **Monte Carlo methods** are a widely used class of **computational algorithms** for simulating the behavior of various physical and mathematical systems, and for other computations.
- **Monte Carlo algorithm** is often used to find solutions to mathematical numerical problems (which may have many variables) that cannot easily be solved, (e.g. integral calculus, or other numerical methods)

Monte Carlo Simulation

- A scheme employing random numbers which is used to solve certain stochastic or deterministic problems where the passage of time plays no substantive role.
- Common problem is the estimation of $\int_{\Omega} f(\mathbf{x})d\mathbf{x}$, where f is a function, $\mathbf{x}$ is a vector and Ω is domain of integration.
- Special case: Estimate $\int_a^b f(x)dx$ for scalar x and limits of integration a, b

Monte Carlo Simulation

Let X be a uniform random variable on the interval $[a, b]$ with density

$$p(x) = \frac{1}{b-a}, \quad a \le x \le b$$

and let $x_1, \ldots, x_n$ be a random sample from X. Then

$$\begin{aligned}\int_a^b f(x)dx &= \int_a^b \frac{f(x)}{p(x)} p(x)dx \\ &= (b-a)\int_a^b f(x)p(x)dx \\ &= (b-a)E[f(X)] \\ &\approx \frac{b-a}{n}\sum_{i=1}^{n} f(x_i)\,.\end{aligned}$$

Monte Carlo Simulation

Example: Estimate $\int_0^b \sin(x)dx$.

We approximate this by

$$\frac{b}{n}\sum_{i=1}^{n}\sin(x_i)\,,$$

where $x_1, \ldots, x_n$ are a sample from a uniform $[0, b]$ random variable.

Monte Carlo Simulation

Example: Estimate $\int_0^b \sin(x)dx$.

	n = 10	n = 100	n = 1000	n = 2000
b = 1 (answer = 2)	1.753	2.032	1.994	1.999
b = 2 (answer = 0)	-0.898	-0.013	0.137	0.079

There is considerable variability in the quality of solution; accuracy of numerical integration sensitive to integrand and domain of integration