statistics

FOR DATA SCIENCE

UNIT-2

feedback/corrections: vibha@pesu.pes.edu

Vibha Masti

random
NUMBER GENERATION

Inverse transform technique

Direct transformation for the Normal Distribution

Convolution Method

Acceptance / Rejection Method

- All assume that uniformly distributed numbers in [0,1] exists

Inverse transform technique

- cdf is $F(X)$ where X is the random variable
- Set $F(X) = R$ where R is a uniformly distributed random variable in [0,1]
- Solve $F(X) = R$ for X in terms of R (in range of X)
- Generate uniform random numbers $R_1, R_2, R_3 \dots$ and compute X_i by

$$\boxed{X_i = F^{-1}(R_i)}$$

Q25. Generate RVs for given cdf

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

- Range of X : [a, b]

$$R = \frac{X-a}{b-a}$$

$$R(b-a) + a = X \qquad \text{where} \qquad R \in [0,1]$$

$$\boxed{X_i = a + R_i(b-a)}$$

Generation of Bernoulli RVs

$$P(X=1) = p$$
$$P(X=0) = 1-p$$

- A Bernoulli RV can only take up 2 values (0 and 1) and the probability of getting $X=1$ is p
- Generate U from $U(0,1)$
- If $U \le p$, $X=1$; else $X=0$

Generation of Binomial RVs

$$P(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$$

$$\boxed{P(X=i) = {}^nC_i \, p^i (1-p)^{n-i}}$$

for i successes in n independent Bernoulli trials

- Generate n Bernoulli (p) RVs (using the technique shown above) $Y_1, Y_2, Y_3, \ldots, Y_n$
- Set $X = Y_1 + Y_2 + \ldots + Y_n$ (sum of n Bernoulli RVs)
- If $Y_i = 1$, number of successes out of n increases
- Total no. of successes $= \sum_{i=1}^{n} Y_i$

Generation of Poisson RVs

$X \sim \text{Poisson}(\lambda)$

$$\boxed{P(X=i) = \frac{\lambda^i}{i!} e^{-\lambda}}$$

i number of events in a unit time interval

Method 1

- Generate exponential inter-event times $Y_1, Y_2 \ldots$ with mean 1
- Let I be the smallest index such that

$$\sum_{i=1}^{I+1} Y_i > \lambda$$

- Set $X = I$

Method 2

- Generate $U(0,1)$ RVs $U_1, U_2 \ldots$
- Let N be the smallest index such that

$$\prod_{i=1}^{N+1} U_i < e^{-\lambda}$$

Steps

1. Set $i=0$, $P=1$
2. Generate U_{i+1} from $U(0,1)$ and replace P with $P \cdot U_{i+1}$
3. If $P < e^{-\lambda}$, accept $N=i$ and go to step 1.

   Else, reject $N=i$ and increment i and return to step 2.

   Upon completion of step 3, $\boxed{P = \prod_{i=1}^{N+1} U_i}$

- If $N=n$, then $n+1$ RVs are requested. So, the average number is given by $E(N+1) = \lambda + 1$

Q26. Generate 3 Poisson variates with mean $\lambda = 0.2$ for the random numbers R= 0.4357, 0.4146, 0.8353, 0.9952, 0.8004

$$e^{-\lambda} = e^{-0.2} = 0.8187$$

1. For X_1

$i = 0$, $P = 1$, $R = 0.4357$

(a) $P = 0.4357 \times 1 = 0.4357 < 0.8187$ ✓
(b) Accept $N = 0$
(c) $X_1 = 0$

2. For X_2

$i = 0$, $P = 1$, $R = 0.4146$

(a) $P = 1 \times 0.4146 = 0.4146 < 0.8187$ ✓
(b) Accept $N = 0$
(c) $X_2 = 0$

3. For X_3

$i = 0$, $P = 1$, $R_1 = 0.8353$, $R_2 = 0.9952$, $R_3 = 0.8004$

(a) $P = 1 \times 0.8353 = 0.8353 > 0.8187$ ✗ $i = 0$
(b) $P = 0.8353 \times 0.9952 = 0.8313 > 0.8187$ ✗ $i = 1$
(c) $P = 0.8313 \times 0.8004 = 0.6654 < 0.8187$ ✓ $i = 2$
(d) Accept $N = 2$
(e) $X_3 = 2$

Poisson numbers : 0, 0, 2

Generation of Normal RVs

Acceptance / Rejection Technique

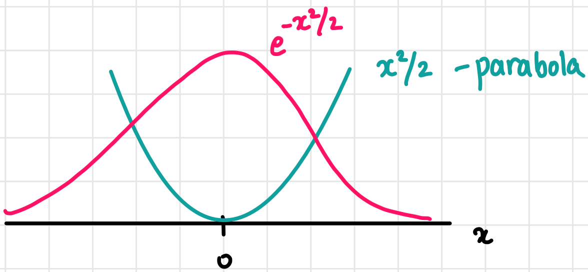
- If $X \sim N(0,1^2)$, the pdf of X is

$$f(x) = \boxed{\frac{1}{\sqrt{2\pi}}} e^{-x^2/2}$$

reason for $1/\sqrt{2\pi}$:

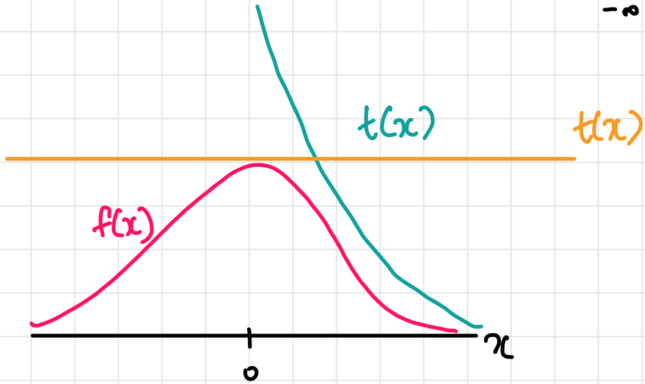
$$\int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}$$

and $\int_{-\infty}^{\infty} f(x)\,dx = 1$



- Let a function $t(x)$ majorise pdf $f(x)$ $\Rightarrow$ t is NOT a density

$$t(x) \geq f(x) \quad \forall x \qquad \left(\int_{-\infty}^{\infty} t(x)\,dx \geq 1 \right)$$



- total area

$$c = \int_{-\infty}^{\infty} t(x)\,dx \geq \int_{-\infty}^{\infty} f(x)\,dx = 1$$

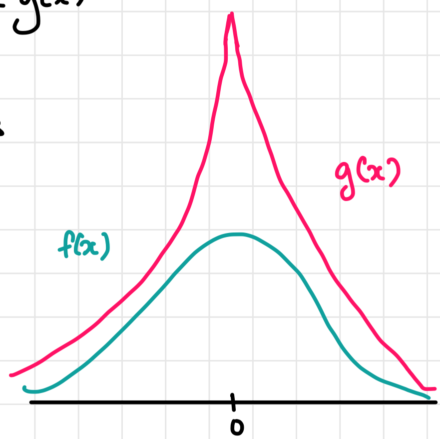
- However, $\boxed{\dfrac{t(x)}{c}}$ is a density

- Let $g(x) = \dfrac{t(x)}{c} \Rightarrow t(x) = c\,g(x)$

- If $X \sim N(0, 1^2)$, the pdf of $|X|$ is

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}$$

- Let $g(x) = \sqrt{\dfrac{2e}{\pi}}\, e^{-x}$



Steps

1. Generate exponential Y with mean 1
2. Generate U from $U(0,1)$
3. If $U \leq e^{-(Y-1)^2/2}$, accept Y $\quad \dfrac{f(Y)}{g(Y)} \leq c$

   Else, go to step 1

4. Return $X = Y$ or $X = -Y$ with probability 0.5

Box-Muller Method

- Box-Muller Transform transforms from a two-dimensional uniform distribution to a two-dimensional bivariate normal distribution (complex normal distribution)

- If U_1 and U_2 are independent RVs from $U(0,1)$

$$\boxed{z_1 = \sqrt{-2\ln U_1}\ \cos(2\pi U_2)}$$

← for single-variable normal distribution

$$z_2 = \sqrt{-2\ln U_1}\ \sin(2\pi U_2)$$

No need to remember this:

- By solving for U_1 and U_2

$$z_1^2 + z_2^2 = -2\ln U_1$$

$$-\ln U_1 = \frac{z_1^2 + z_2^2}{2}$$

$$\boxed{U_1 = e^{-\left(\frac{z_1^2+z_2^2}{2}\right)}}$$

$$\frac{z_2}{z_1} = \tan(2\pi U_2)$$

$$2\pi U_2 = \tan^{-1}\left(\frac{z_2}{z_1}\right)$$

$$\boxed{U_2 = \frac{1}{2\pi}\tan^{-1}\left(\frac{z_2}{z_1}\right)}$$

- Taking the Jacobian $\frac{\partial(U_1, U_2)}{\partial(z_1, z_2)}$

$$\begin{vmatrix} \frac{\partial U_1}{\partial z_1} & \frac{\partial U_1}{\partial z_2} \\ \frac{\partial U_2}{\partial z_1} & \frac{\partial U_2}{\partial z_2} \end{vmatrix} = -\left(\frac{1}{\sqrt{2\pi}} e^{-z_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-z_2^2/2}\right)$$

Q27. Suppose the height of adult males in a certain area is normally distributed with a mean of 168 cm and a standard deviation of 8cm. Simulate the height of 4 adults.

$X \sim N(168, 8^2)$

$\mu = 168$
$\sigma = 8$
$n = 4$

$$z = \frac{X - \mu}{\sigma}$$

- We first generate random z values

- $z = \sqrt{-2\ln(U_1)} \cos(2\pi U_2)$

$\boxed{X_1}$

Let $U_1 = 0.2432$, $U_2 = 0.5214$

$$z_1 = \sqrt{-2\ln(0.2432)} \cos(2\pi \times 0.5214)$$
$$z_1 = -1.666$$

$$X_1 = z_1 \times \sigma + \mu = -1.666 \times 8 + 168$$

$X_1 = 154.67$ cm

$\boxed{X_2}$

Let $U_1 = 0.8921$, $U_2 = 0.6232$

$$z_2 = \sqrt{-2\ln(0.8921)} \cos(2\pi \times 0.6232)$$
$$z_2 = -0.3417$$

$X_2 = 165.27$ cm

$\boxed{X_3}$

Let $U_1 = 0.4421$, $U_2 = 0.0012$

$$z_3 = \sqrt{-2\ln(0.4421)}\cos(2\pi \times 0.0012)$$
$$z_3 = 1.2776$$

$X_3 = 178.22$ cm

$\boxed{X_4}$

Let $U_1 = 0.9921$, $U_2 = 0.7324$

$$z_4 = \sqrt{-2\ln(0.9921)}\cos(2\pi \times 0.7324)$$
$$z_4 = -0.0139$$

$X_4 = 167.89$ cm

Q28. Suppose the no. of shipments, x, on the loading dock of a company is either 0, 1 or 2. Generate RVs given $U = 0.23, 0.52, 0.81, 0.34$

(discrete RVs – inverse)

x	$P(x)$	$F(x)$
0	0.5	0.5
1	0.3	0.8
2	0.2	1.0

Let $U \sim U(0,1)$

$$x = \begin{cases} 0 & U \le 0.5 \\ 1 & 0.5 < U \le 0.8 \\ 2 & U > 0.8 \end{cases}$$

$U_1 = 0.23 \Rightarrow U < 0.5 \Rightarrow X_1 = 0$

$U_2 = 0.52 \Rightarrow 0.5 < U \le 0.8 \Rightarrow X_2 = 1$

$U_3 = 0.81 \Rightarrow U > 0.8 \Rightarrow X_3 = 2$

$U_4 = 0.34 \Rightarrow U < 0.5 \Rightarrow X_4 = 0$