STATISTICS FOR DATA SCIENCE

Generation of Random Variates

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Generation of Random Variates

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

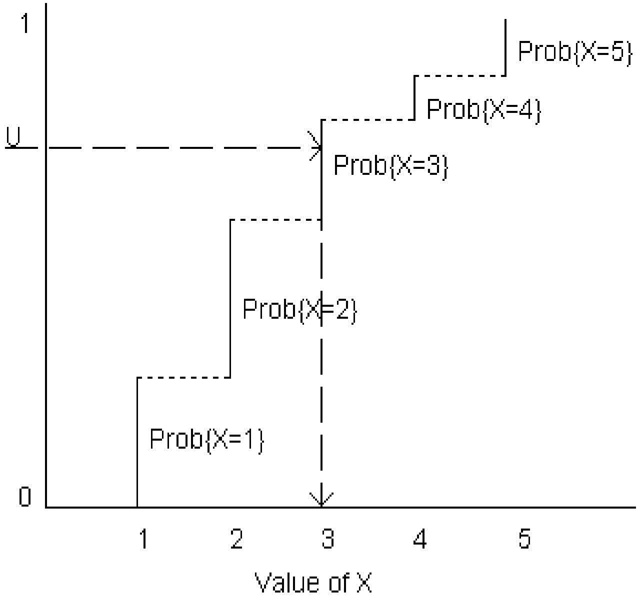
Topics to be covered...

- Random Numbers
- Random Variate Generator
- Random Variates
- Techniques for Generating Random Variates

Random Numbers

- Random numbers are very important for a simulation.

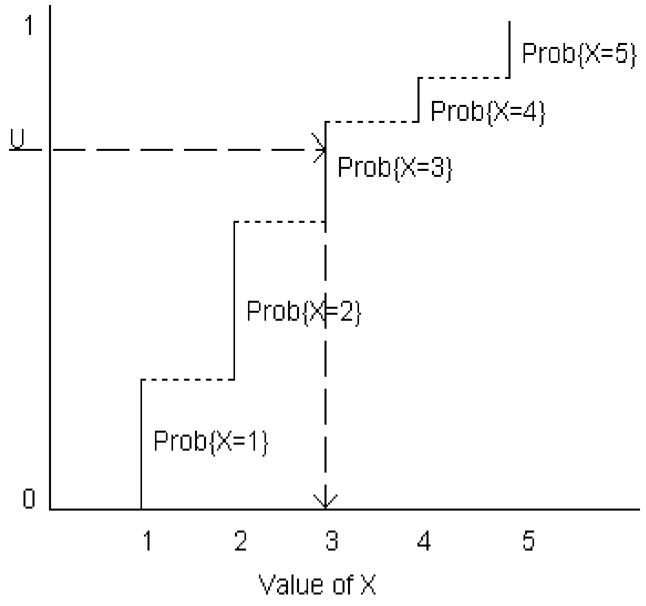
- Random numbers are at the foundations of computer simulation methods, not only to the probabilistic methods. One needs them to generate configurations or states of a system, as well as for the decision process to accept or reject a configuration or state.



Random Numbers

- Random numbers are very important for a simulation.

- Since all the randomness required by the model is simulated by a random number generator,
  - Whose output is assumed to be a sequence of independent and identically (uniformly) distributed random numbers between 0 and 1.
  - Then these random numbers are transformed into required probability distributions.

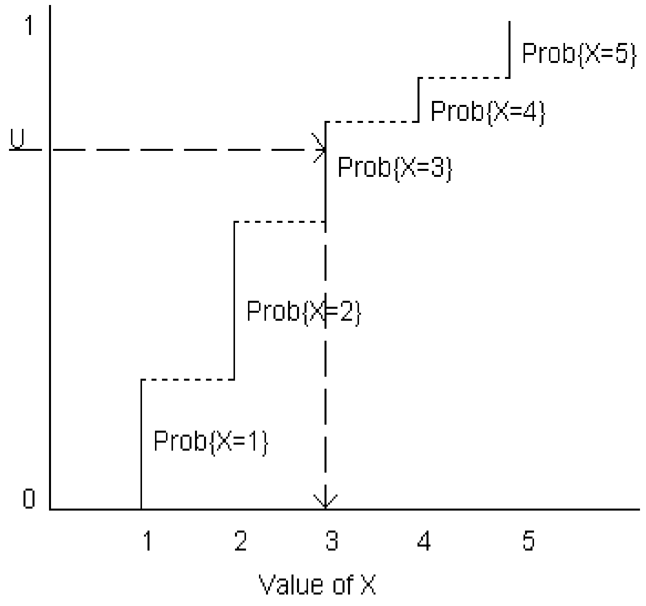


Random Numbers

- Random numbers are very important for a simulation.

- Since all the randomness required by the model is simulated by a random number generator,
  - Whose output is assumed to be a sequence of independent and identically (uniformly) distributed random numbers between 0 and 1.

  - Then these random numbers are transformed into required probability distributions.

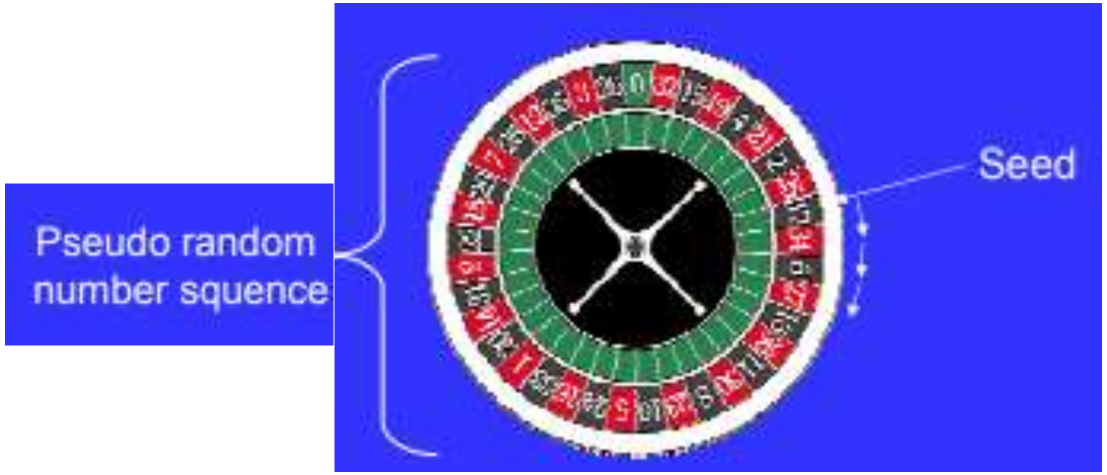


Source: http://www.cs.bilkent.edu.tr/~cagatay/cs503/_M&S_04_Random_Variate_Generation.pdf,Sigma

Random Number Seed

- Computer-based generators use random number seeds for setting the starting point of the random number sequence.

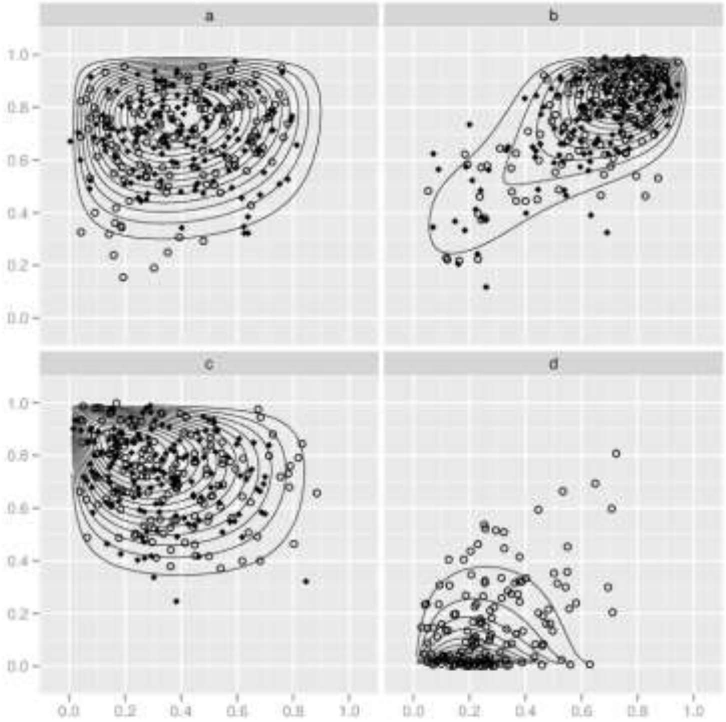
- These seeds are often initialized using a computer's real time clock in order to have some external noise.



Source: http://www.cs.bilkent.edu.tr/~cagatay/cs503/_M&S_04_Random_Variate_Generation.pdf

Random Variate Generation

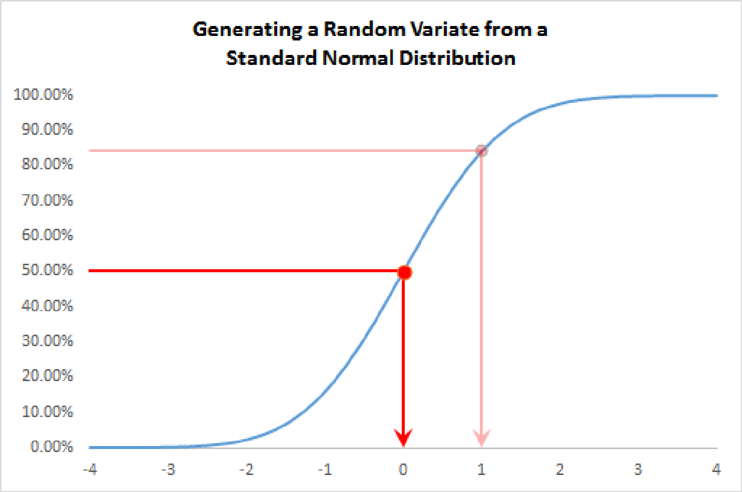
It is assumed that a distribution is completely specified and we wish to generate samples from this distribution as input to a simulation model.



Source: mathworld.wolfram.com , www.slideshare.net,ResearchGate, Lambert

Random Variate Generation

- Random variate generation is the process of producing observations that have the distribution of the given random variables.
- This is to develop simulation models for the purpose of analysis and decision making.
- The process of generating random variates for the distribution rely on generating uniformly distributed random number on the interval (0,1).
- Random variate generators use as starting point random numbers distributed U[0,1].



Random Number Generators

- A computational or physical device designed to generate a sequence of numbers that lack any pattern (i.e. appear random).

- Computer-based generators are simple deterministic programs trying to fool the user by producing a deterministic sequence that looks random (pseudo random numbers).

- Therefore, they should meet some statistical tests for randomness intended to ensure that they do not have any easily discernible patterns.

Source: http://www.cs.bilkent.edu.tr/~cagatay/cs503/_M&S_04_Random_Variate_Generation.pdf

Random Variate

A **random variate** is a variable generated from uniformly distributed pseudorandom numbers. Depending on how they are generated, a **random variate** can be uniformly or non-uniformly distributed.

Random variates are frequently used as the input to simulation models

Examples: Inter-arrival time and service time.

Source: mathworld.wolfram.com , www.slideshare.net

Random Variate

RV Generators – Techniques used to generate random variates.

- Inverse transform technique
- Direct transformation for the Normal Distribution
- Convolution Method
- Acceptance and Rejection Technique

Note: All the techniques assume that a source of uniform [0,1] random numbers R1,R2,…(uniformly distributed random numbers) is readily available.

Source: mathworld.wolfram.com , www.slideshare.net

Techniques in Generating Random Variates

- Inverse transform technique
- Direct transformation for the normal distribution
- Convolution method
- Acceptance and rejection technique

All the techniques assume that a source of uniform [0,1] random numbers R1,R2,…(uniformly distributed random numbers) is readily available.

Source:www.slideshare.net

Inverse Transform Technique

- This technique is used to sample from discrete continuous or uniform type of distribution.

- So, our empirical kind of distributions, and the goal is to develop a procedure for generating X1, X 2, X 3 all that which have a particular kind of distribution function.

- So, you have for a particular type say any anything like Weibull exponential or gamma or so.

- The technique is useful when CDF that is F (x) is of simple form. So that F^{-1} can be computed easily.

Steps in Inverse Transform technique

Steps:

1. Compute CDF of the desired random variable X.
2. Set F(X)=R on the range of X.
3. Solve the equation F(X)=R for X in terms of R.
4. Generate uniform random numbers R1,R2,R3… and compute the desired random variate by

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

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Steps in Inverse Transform Technique

Inverse transform method – Uniform Distibution Example:

Step 1 – compute *cdf* of the desired random variable X

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

Step 2 – Set F(X) = R where R is a random number ~U[0,1)

$$F(x) = R = \frac{x-a}{b-a}$$

Step 3 – Solve F(X) = R for X in terms of R. $X = F^{-1}(R)$.

$$R(b-a) = X - a, \quad X = R(b-a) + a$$

Step 4 – Generate random numbers R_i and compute desired random variates:

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

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Generation of Bernoulli Random Variates

Bernoulli Two possible outcomes of X (success or failure):

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

Algorithm:

- Generate U from U(0, 1);
- If $U \le p$, then $X = 1$; else $X = 0$.

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Binomial Random Variates

A random variable X has a binomial distribution with parameters n and p if

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \qquad i = 0, 1, \ldots, n$$

X is the number of successes in n independent Bernoulli trials, each with success probability p.

Algorithm:

- Generate n Bernoulli(p) random variables Y1, . . . , Yn;
- Set X = Y1 + Y2 + · · · + Yn.

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Binomial Random Variates(Alternate Algorithm)

Let Y1, Y2, . . . be independent geometric(p) random variables, and I the smallest index such that

$$\sum_{i=1}^{I+1}(Y_i + 1) > n.$$

Then the index I has a binomial distribution with parameters n and p.

Let Y1, Y2, . . . be independent exponential random variables with mean 1, and I the smallest index such that

$$\sum_{i=1}^{I+1}\frac{Y_i}{n-i+1} > -\ln(1-p).$$

Then the index I has a binomial distribution with parameters n and p.

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Poisson Random Variate

A random variable X has a Poisson distribution with parameter λ if

$$P(X=i)=\frac{\lambda^i}{i!}e^{-\lambda}, \qquad i=0,1,2,\ldots$$

X is the number of events in a time interval of length 1 if the inter-event times are independent and exponentially distributed with parameter λ.

Algorithm:

• Generate exponential inter-event times Y1, Y2, . . . with mean 1; let I be the smallest index such that

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

• Set X = I.

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Poisson Random Variate(Alternate)

Algorithm:

- Generate U(0,1) random variables U1, U2, . . .let I be the smallest index such that

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

- Set X = I

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Normal Random Variate

Acceptance-Rejection method If X is N(0, 1), then the density of |X| is given by

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x > 0.$$

Now the function

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

majorizes f .

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Normal Random Variate

This leads to the following algorithm:

1. Generate an exponential Y with mean 1;
2. Generate U from U(0, 1), independent of Y ;
3. If $g(x) = \sqrt{2e/\pi}e^{-x}$ then accept Y ; else reject Y and return to step 1.
4. Return X = Y or X = −Y , both with probability 1/2.

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Generation of Normal Random Variate

Box-Muller method

If U1 and U2 are independent U(0, 1) random variables, then

$$X_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2)$$
$$X_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

are independent standard normal random variables

Source:https://www.win.tue.nl/~marko/2WB05/lecture8.pdf

Inverse-transform Technique: Other Distributions

Examples of other distributions for which inverse CDF works are:

- Uniform distribution
- Weibull distribution
- Triangular distribution

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Inverse-transform Technique: Discrete Distribution

All discrete distributions can be generated via inverse transform technique.

Method:
Numerically, table-lookup procedure, algebraically, or a formula

Examples of application:
- Empirical
- Discrete uniform
- Geometric

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Inverse-transform Technique: Discrete Distribution

Example: Suppose the number of shipments, x, on the loading dock of a company is either 0, 1, or 2

- Data - Probability distribution:

x	$P(x)$	$F(x)$
0	0.50	0.50
1	0.30	0.80
2	0.20	1.00

- The inverse-transform technique as table-lookup procedure.

$$F(x_{i-1}) = r_{i-1} < R \le r_i = F(x_i)$$

- Set X = xi

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

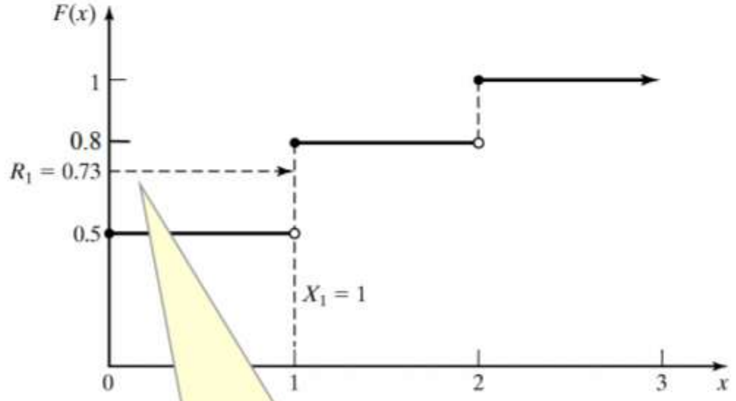
Inverse-transform Technique: Discrete Distribution

Method - **Given R, the generation scheme becomes:**

$$x = \begin{cases} 0, & R \le 0.5 \\ 1, & 0.5 < R \le 0.8 \\ 2, & 0.8 < R \le 1.0 \end{cases}$$

Table for generating the discrete variate X

i	Input r_i	Output x_i
1	0.5	0
2	0.8	1
3	1.0	2



Consider $R_1 = 0.73$:
$F(x_{i-1}) < R \le F(x_i)$
$F(x_0) < 0.73 \le F(x_1)$
Hence, $X_1 = 1$

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Inverse-transform Technique: Continuous Distributions

A number of continuous distributions do not have a closed form expression for their CDF, e.g. Normal, Gamma and Beta.

The presented method does not work for these distributions.

Solution

- Approximate the CDF or numerically integrate the CDF

Problem

- Computationally slow

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Acceptance and Rejection Technique

- Useful particularly when inverse CDF does not exist in closed form
  -Thinning
- **Illustration:** To generate random variates, X ~ U(1/4,1)

Procedure:

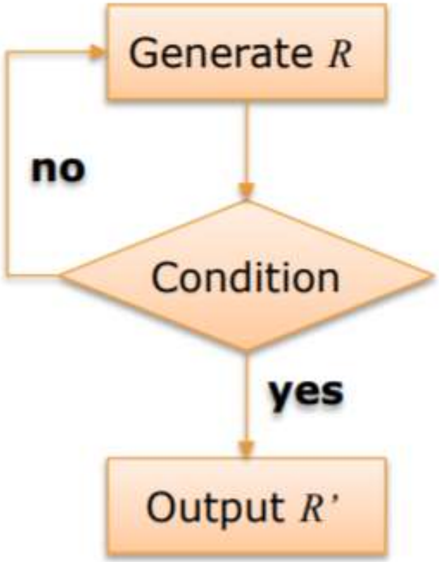
Step 1. Generate $R \sim U(0,1)$

Step 2. If $R \geq ¼$, accept $X=R$.

Step 3. If $R < ¼$, reject R, return to Step 1

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Acceptance and Rejection Technique



R does not have the desired distribution, but R conditioned (R’) on the event {R ≥ ¼} does.

- Efficiency: Depends heavily on the ability to minimize the number of rejections.

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Acceptance and Rejection Technique – Poisson Distribution

Procedure of generating a Poisson random variate N is as follows

1. Set $n=0$, $P=1$
2. Generate a random number R_{n+1}, and replace P by $P \times R_{n+1}$
3. If $P < \exp(-\alpha)$, then accept $N=n$
   - Otherwise, reject the current n, increase n by one, and return to step 2.

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Acceptance and Rejection Technique – Poisson Distribution

- Example: Generate three Poisson variates with mean α=0.2
  - $\exp(-0.2) = 0.8187$
- Variate 1
  - Step 1: Set $n = 0, P = 1$
  - Step 2: $R1 = 0.4357, P = 1 \text{ x } 0.4357$
  - Step 3: Since $P = 0.4357 < \exp(-0.2)$, accept $N = 0$
- Variate 2
  - Step 1: Set $n = 0, P = 1$
  - Step 2: $R1 = 0.4146, P = 1 \text{ x } 0.4146$
  - Step 3: Since $P = 0.4146 < \exp(-0.2)$, accept $N = 0$
- Variate 3
  - Step 1: Set $n = 0, P = 1$
  - Step 2: $R1 = 0.8353, P = 1 \text{ x } 0.8353$
  - Step 3: Since $P = 0.8353 > \exp(-0.2)$, reject $n = 0$ and return to Step 2 with $n = 1$
  - Step 2: $R2 = 0.9952, P = 0.8353 \text{ x } 0.9952 = 0.8313$
  - Step 3: Since $P = 0.8313 > \exp(-0.2)$, reject $n = 1$ and return to Step 2 with $n = 2$
  - Step 2: $R3 = 0.8004, P = 0.8313 \text{ x } 0.8004 = 0.6654$
  - Step 3: Since $P = 0.6654 < \exp(-0.2)$, accept $N = 2$

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Acceptance and Rejection Technique – Poisson Distribution

- It took five random numbers to generate three Poisson variates
- In long run, the generation of Poisson variates requires some overhead!

N	R_{n+1}	P	Accept/Reject		Result
0	0.4357	0.4357	P < exp(- α)	Accept	N=0
0	0.4146	0.4146	P < exp(- α)	Accept	N=0
0	0.8353	0.8353	P ≥ exp(- α)	Reject	
1	0.9952	0.8313	P ≥ exp(- α)	Reject	
2	0.8004	0.6654	P < exp(- α)	Accept	N=2

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Direct Transformation

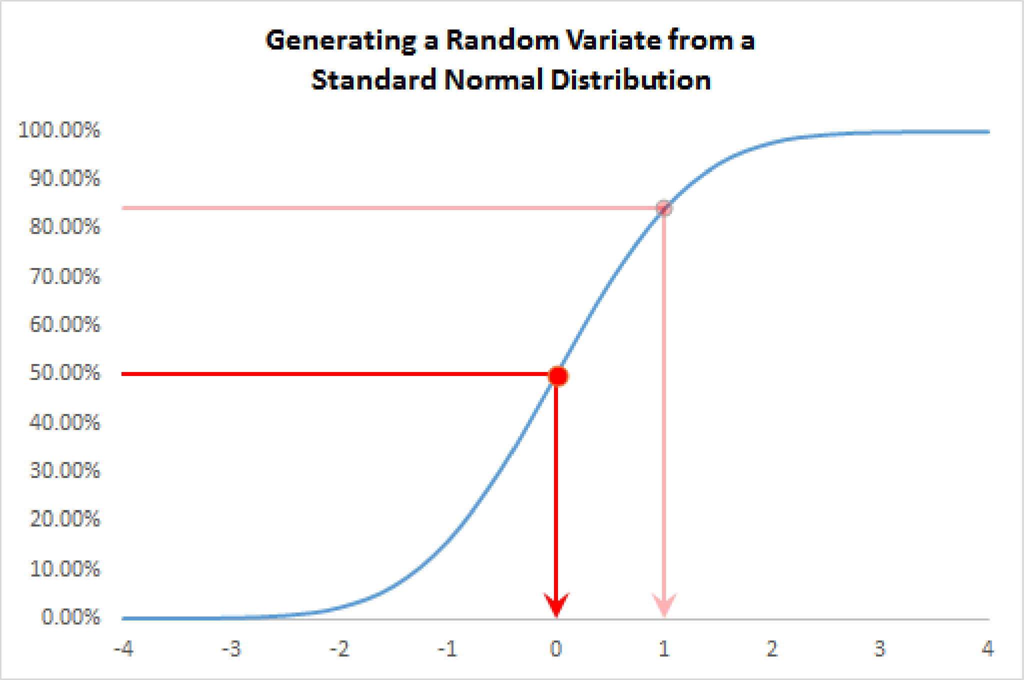
Approach for $N(0,1)$

- PDF

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

- CDF, No closed form available

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



Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf, tvmcalcs.com

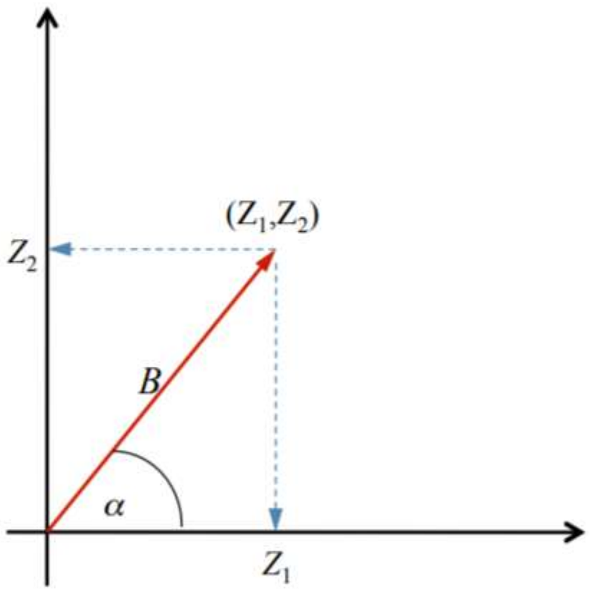
Direct Transformation

Approach for N(0,1)

• Consider two standard normal random variables, Z1 and Z2, plotted as a point in the plane:

• In polar coordinates:

- Z1 = B cos(α)
- Z2 = B sin(α)



Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Direct Transformation

- Approach for $N(\mu, \sigma^2)$:
  - Generate $Z_i \sim N(0,1)$

$$\boxed{X_i = \mu + \sigma Z_i}$$

- Approach for Lognormal(μ,σ^2):
  - Generate $X \sim N(\mu,\sigma^2)$

$$\boxed{Y_i = e^{X_i}}$$

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

Direct Transformation-Example

Let R1 = 0.1758 and R2=0.1489

- Two standard normal random variates are generated as follows:

$$Z_1 = \sqrt{-2\ln(0.1758)}\cos(2\pi 0.1489) = 1.11$$

$$Z_2 = \sqrt{-2\ln(0.1758)}\sin(2\pi 0.1489) = 1.50$$

- To obtain normal variates Xi with mean μ=10 and variance σ^2 = 4

$$X_1 = 10 + 2 \cdot 1.11 = 12.22$$

$$X_2 = 10 + 2 \cdot 1.50 = 13.00$$

Source:https://www.mi.fu-berlin.de/inf/groups/ag-tech/teaching/2012_SS/L_19540_Modeling_and_Performance_Analysis_with_Simulation/07.pdf

THANK YOU

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering