Chapter 8
Random-Variate Generation

Banks, Carson, Nelson & Nicol

Discrete-Event System Simulation

Purpose & Overview

- Develop understanding of generating samples from a specified distribution as input to a simulation model.
- Illustrate some widely-used techniques for generating random variates.
  - Inverse-transform technique
  - Acceptance-rejection technique
  - Special properties

Random Number Generator

- All the techniques assume that a source of uniform [0,1] random numbers $R_1, R_2, \ldots$ is readily available, where each R_i has pdf

$$f_R(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

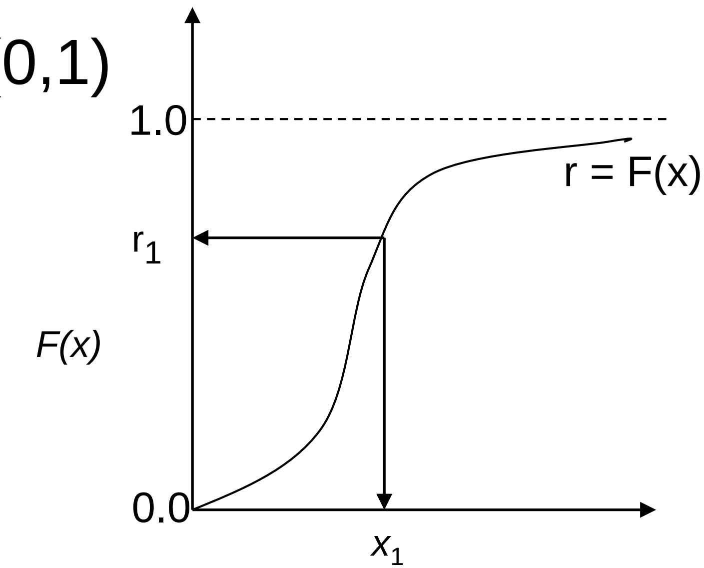
and cdf

$$F_R(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

Inverse-transform Technique

- The concept:
  - For cdf function: $r = F(x)$
  - Generate r from uniform (0,1)
  - Find x:

$$\boxed{x = F^{-1}(r)}$$



Note $0 \le F(x) \le 1$

Steps to follow

1. Compute the 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 (as needed) uniform Random Numbers R_1, R_2, R_3, …. and compute the desired random variates by $X_i=F^{-1}(R_i)$

Exponential Distribution [Inverse-transform]

- Exponential Distribution:
  - Exponential cdf:

Density $f(x)=\lambda e^{-\lambda x}$

$$r = F(x)$$
$$= 1 - e^{-\lambda x} \text{ for } x \geq 0$$

  - To generate $X_1, X_2, X_3 \ldots$

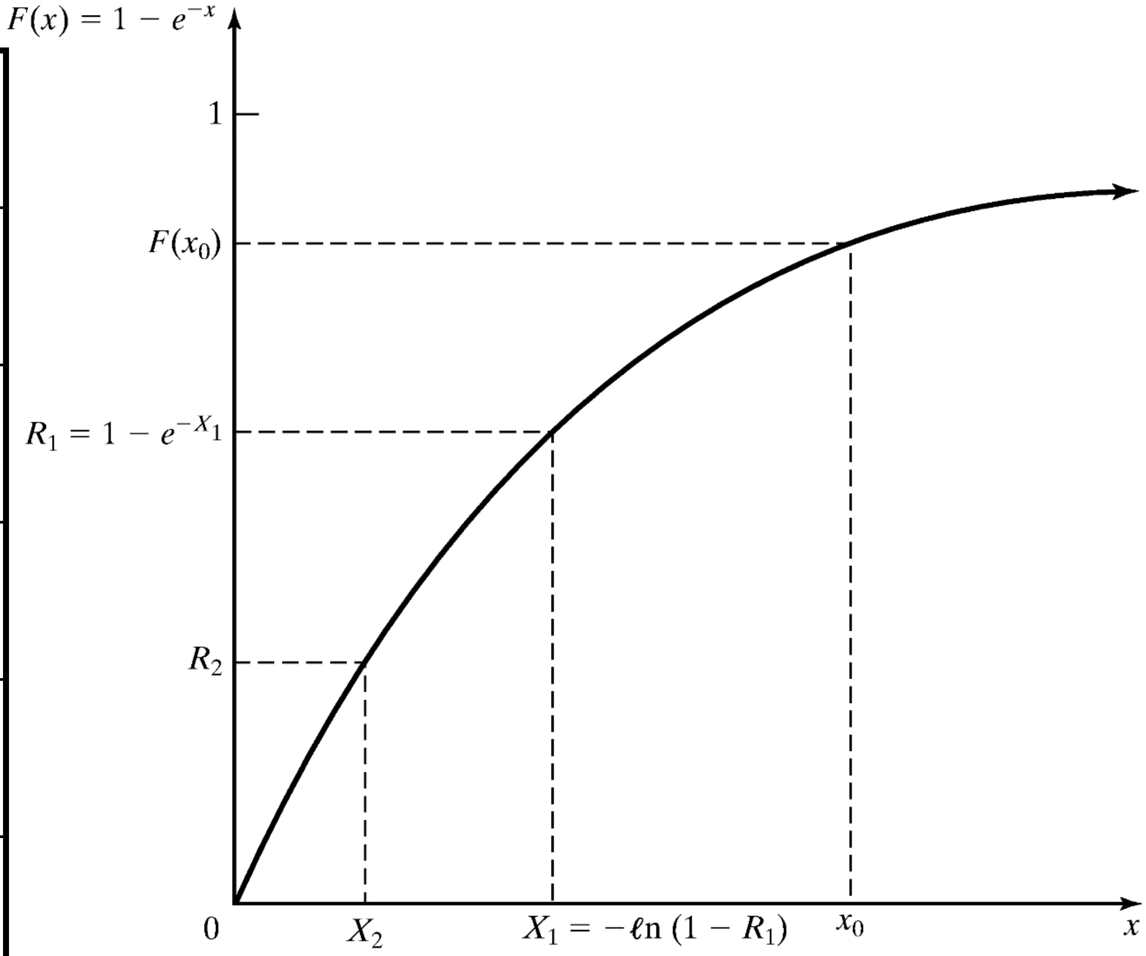
$$X_i = F^{-1}(R_i)$$
$$= -(1/\lambda) \ln(1-R_i) \quad \text{[Eq'n 8.3]}$$

- One simplification is use to Replace $(1-R_i)$ with R_i as both R_i and $(1-R_i)$ are uniformly distributed on [0,1]

$$X_i = -(1/\lambda) \ln(R_i)$$

Exponential Distribution (λ=1)

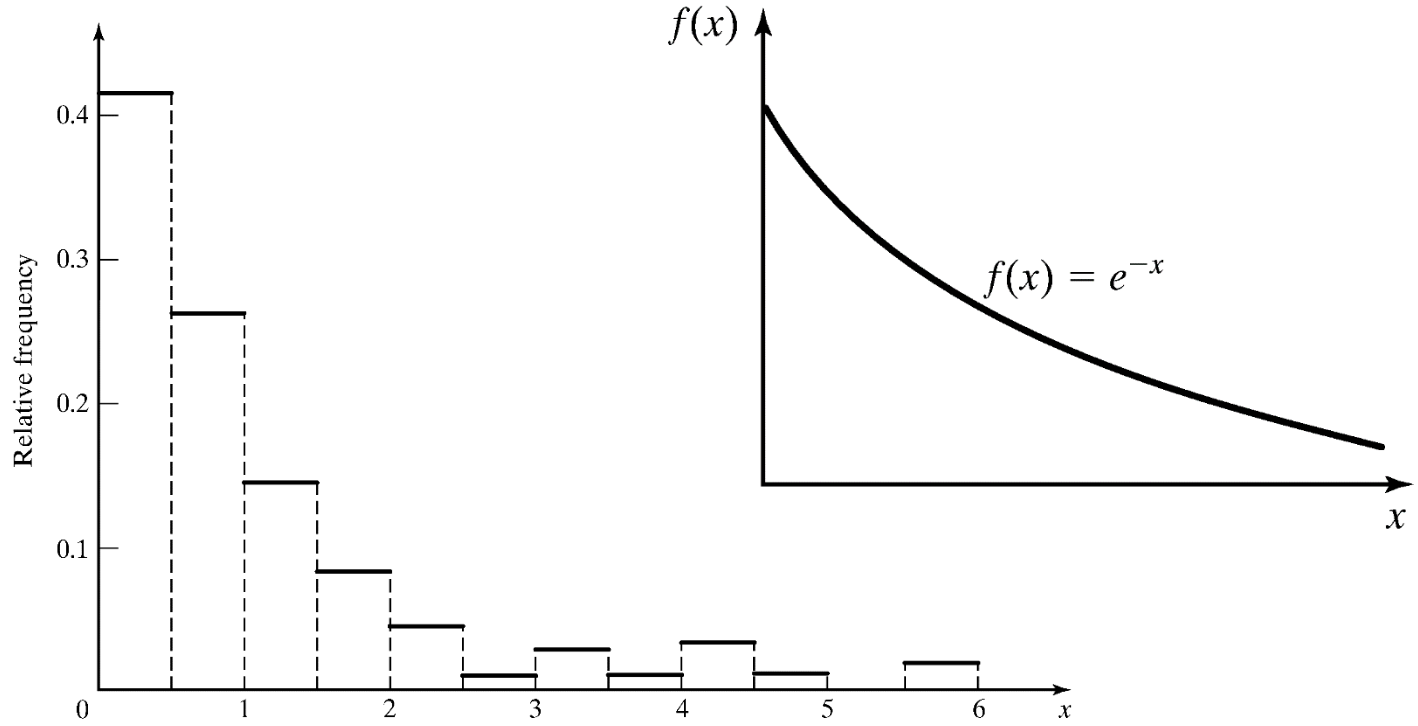
i	R_i	X_i
1	0.1306	0.1400
2	0.0422	0.0431
3	0.6597	1.078
4	0.7965	1.592
5	0.7696	1.468



Exponential Distribution

[Inverse-transform]

- Example: Generate *200* variates X_i with distribution *exp(λ= 1)*
  - Generate *200* R_i with *U(0,1)* and utilize eq'n 8.3, the histogram of X_i becomes:



Does X have the desired distribution??

- Check: Does the random variable X_1 generated through transformation have the desired distribution?

$$\because R_1 \text{ is Uniformly distributed on } [0,1]$$

$$P(X_1 \le x_0) = P(R_1 \le F(x_0)) = F(x_0)$$

Remember for Uniform Distribution U(0,1)
$P(Y \le y) = y$ for $0 \le y \le 1$

Java Code: Exponential Distribution

```
public static double exponential(Random rng, double mean)
{
        return -mean*Math.log( rng.nextDouble() );
}
```

Other Distributions

- Examples of other distributions for which inverse cdf works are:
  - Uniform distribution
  - Weibull distribution
  - Triangular distribution

Uniform Distribution [Inverse-transform]

- Consider a random variable X uniformly distributed on the interval [a,b].
- The pdf of X is: $f(x)=\begin{cases} 1/b-a, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$
- The cdf is given by:

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

- Set *F(X)=(X-a)/(b-a)=R*
- Solving for X yields *X=a+(b-a)R*
- Therefore the random variate is *X=a+(b-a)R*

Weibull Distribution

[Inverse-transform]

- Weibull is used to model for time to failures for machines or components
- The pdf of X is:

$$f(x) = \begin{cases} \dfrac{\beta}{\alpha^{\beta}} x^{\beta-1} e^{-(x/\alpha)^{\beta}}, & x \geq 0; \\ 0, & \text{otherwise}; \end{cases}$$

$\alpha > 0$ and $\beta > 0$ are shape parameters

- The cdf is given by:

$$F(x) = 1 - e^{-(x/\alpha)^{\beta}}, x \geq 0$$

- Set $F(x) = 1 - e^{-(x/\alpha)^{\beta}} = R$
- Solving for X yields $X = \alpha\left[-\ln(1-R)\right]^{1/\beta}$

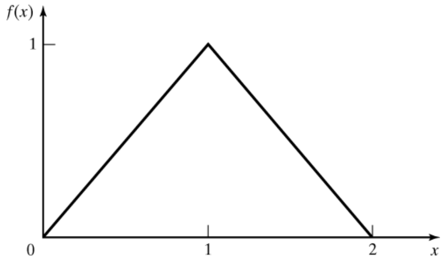
Compare this with exponential variate

- If X is a Weibull variate, then X^{β} is an exponential variate with mean α^{β}
- Conversely, if Y is an exponential variate with mean μ, then $Y^{1/\beta}$ is a Weibull variate with shape parameter β and scale parameter $\alpha=\mu^{1/\beta}$

Triangular Distribution [Inverse-transform]

- Consider a random variable X that has the pdf as:

$$f(x)=\begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 < x \le 2 \\ 0, & \text{otherwise} \end{cases}$$



- The cdf is given by:

$$F(x)=\begin{cases} 0, & x \le 0 \\ x^2/2, & 0 < x \le 1 \\ 1-\left((2-x)^2/2\right), & 1 < x \le 2 \\ 1, & x > 2 \end{cases}$$

- For $0 \le X \le 1$, $R=X^2/2$ and for $1 \le X \le 2$, $R=1-((2-X)^2/2)$
- Thus X is generated by:

$$X=\begin{cases} \sqrt{2R}, & 0 \le R \le 1/2 \\ 2-\sqrt{2(1-R)}, & 1/2 < R \le 1 \end{cases}$$

Empirical Continuous Dist'n [Inverse-transform]

- When theoretical distribution is not applicable that provides a good model, then it is necessary to use empirical distribution of data
- To collect empirical data:
  - One possibility is to resample the observed data itself
    - This is known as *using the empirical distribution*
    - It makes sense if the input process takes a finite number of values
  - If the data is drawn from a continuous valued input process, then we can interpolate between the observed data points to fill in the gaps
- This section looks into defining and generating data from a continuous empirical distribution

Empirical Continuous Dist'n [Inverse-transform]

- Given a small sample set (size *n*):
  - Arrange the data from smallest to largest $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$
  - Define $x_{(0)} = 0$
  - Assign the probability $1/n$ to each interval $x_{(i-1)} \le x \le x_{(i)}$
  - The resulting empirical cdf has a i^{th} line segment slope as

$$a_i = \frac{x_{(i)} - x_{(i-1)}}{1/n - (i-1)/n} = \frac{x_{(i)} - x_{(i-1)}}{1/n}$$

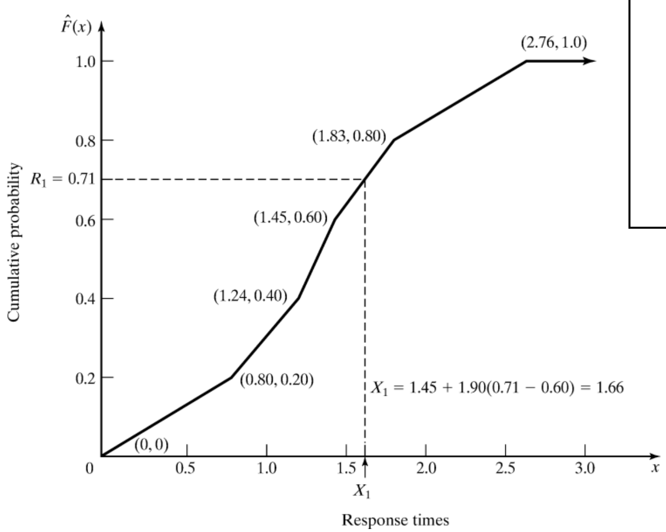
  - The inverse cdf is calculated by

$$\boxed{X = \hat{F}^{-1}(R) = x_{(i-1)} + a_i\left(R - \frac{(i-1)}{n}\right)}$$

where $(i-1)/n < R \le i/n$; *R is the random number generated*

Example Empirical Distribution

- Five observations of fire crew response times (in minutes) to incoming alarms are collected to be used in a simulation investigating possible alternative staffing and crew-scheduling policies. The data are: 2.76, 1.83, 0.80, 1.45, 1.24
- Arranging in ascending order: 0.80, 1.24, 1.45, 1.83, 2.76



i	*Interval (minutes)*	*Probability 1/n*	*Cumulative Probability, i/n*	*Slope, a_i*
1	$0 \le x \le 0.80$	0.2	0.20	4.00
2	$0.80 \le x \le 1.24$	0.2	0.40	2.2
3	$1.24 \le x \le 1.45$	0.2	0.60	1.1
4	$1.45 \le x \le 1.83$	0.2	0.80	1.9
5	$1.83 \le x \le 2.76$	0.2	1.00	4.65

If a random number R1=0.71 is generated, then R1 lies in the fourth interval (between 3/5 and 4/5. Therefore,
$X1 = x_{(4-1)} + a_4(R1-(4-1)/n) = 1.66$

Empirical Distribution

- In general, given c_i is the cumulative probability of first i intervals, $x_{(i-1)} \le x \le x_{(i)}$ is the i^{th} interval, then the slope of the i^{th} line segment is:

$$a_i = \frac{x_{(i)} - x_{(i-1)}}{c_i - c_{i-1}}$$

- The inverse cdf is calculated as:

$$\boxed{\begin{array}{l} X = \hat{F}^{-1}(R) = x_{(i-1)} + a_i\left(R - c_{i-1}\right) \\ \text{when } c_{i-1} < R \le c_i \end{array}}$$

Empirical Continuous Dist'n [Inverse-transform]

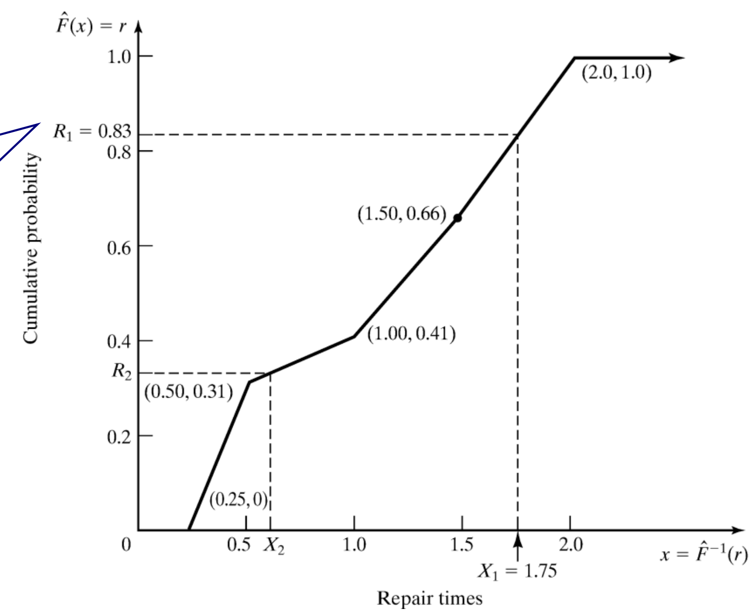
- Example: Suppose the data collected for 100 broken-widget repair times are:

i	*Interval (Hours)*	*Frequency*	*Relative Frequency*	*Cumulative Frequency, c_i*	*Slope, a_i*
1	$0.25 \le x \le 0.5$	31	0.31	0.31	0.81
2	$0.5 \le x \le 1.0$	10	0.10	0.41	5.0
3	$1.0 \le x \le 1.5$	25	0.25	0.66	2.0
4	$1.5 \le x \le 2.0$	34	0.34	1.00	1.47

Consider $R_1 = 0.83$:

$c_3 = 0.66 < R_1 < c_4 = 1.00$

$$X_1 = x_{(4-1)} + a_4(R_1 - c_{(4-1)})$$
$$= 1.5 + 1.47(0.83-0.66)$$
$$= 1.75$$



Continuous Distributions with no closed-form inverse

- A number of useful continuous distributions do not have a closed form expression for their cdf or inverse
- Examples are: Normal, Gamma, Beta
- Approximations are possible to inverse cdf
- A simple approximation to inverse cdf of normal is given by

$$\boxed{X = F^{-1}(R) \approx \frac{R^{0.135} - (1-R)^{0.135}}{0.1975}}$$

- The problems with approximate inverse cdf is that some of them can be computationally intensive

Popular Normal Variate

- Method is due to Box and Muller (1958)
- Generate two identical and independent (IID) uniformly distributed variables, U1 and U2 as IID U(0,1)
- Then set

$$X_1 = \sqrt{-2\ln U_1}\cos 2\pi U_2$$

$$X_2 = \sqrt{-2\ln U_1}\sin 2\pi U_2$$

- *X1* and *X2* are IIDs distributed Normally *N(0,1)*, with zero mean
- To generate $X' \sim N(\mu,\sigma^2)$, generate $X \sim N(0,1)$ and then set $X' = \mu + \sigma X$

Java code: Normal Distrbution

```
public static double SaveNormal;
public static int  NumNormals = 0;
public static final double  PI = 3.1415927 ;
public static double normal(Random rng, double mean, double sigma) {
        double ReturnNormal;
        // should we generate two normals?
        if(NumNormals == 0 ) {
          double r1 = rng.nextDouble();
          double r2 = rng.nextDouble();
          ReturnNormal = Math.sqrt(-2*Math.log(r1))*Math.cos(2*PI*r2);
          SaveNormal   = Math.sqrt(-2*Math.log(r1))*Math.sin(2*PI*r2);
          NumNormals = 1;
        } else {
          NumNormals = 0;
          ReturnNormal = SaveNormal;
        }
        return ReturnNormal*sigma + mean ;
 }
}
```

Discrete Distribution

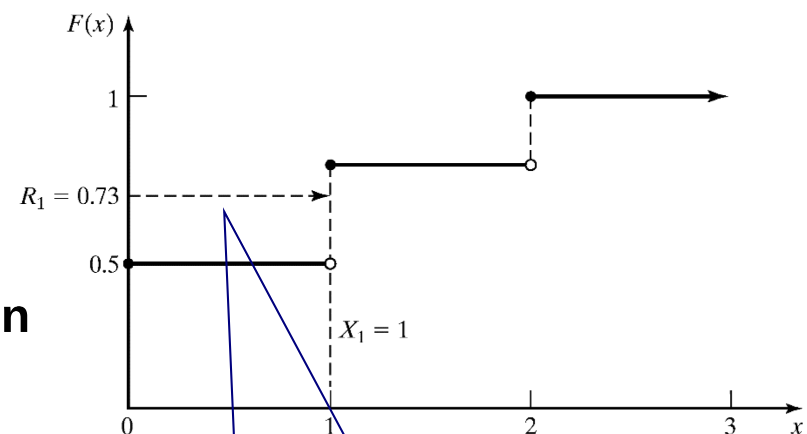
[Inverse-transform]

- All discrete distributions can be generated via inverse-transform technique, either numerically through a table-lookup procedure, or algebraically using a formula
- Examples of application:
  - Empirical
  - Discrete uniform
  - Geometric
  - Gamma

Empirical Discrete Distribution [Inverse-transform]

- 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



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

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

In general, generate R
if $F(x_i\text{-}1) < R \le F(x_i)$
set $X = x_i$

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

Another Empirical Example

- A computer on a network generates the following packet sizes →

Size	Probability
64	0.4
512	0.2
1500	0.4

- Therefore the distribution function is:

$$F(x) = \begin{cases} 0.0, & 0 \le x < 64 \\ 0.4, & 64 \le x < 512 \\ 0.6, & 512 \le x < 1500 \\ 1.0, & 1500 \le x \end{cases}$$

- The inverse function is →

$$F^{-1}(u) = \begin{cases} 64, & 0 \le u < 0.4 \\ 512, & 0.4 \le u < 0.6 \\ 1500, & 0.6 \le u < 1.0 \end{cases}$$

- Generate *u~U(0,1)*
  - *if u<0.4 Size = 64*
  - *if 0.4≤u<0.6, Size = 512*
  - *if 0.6≤u, Size = 1500*

Discrete Uniform Distribution

- Consider Uniform distribution on {1,2,…,k} with pmf and cdf given by

$$p(x) = 1/k, \;\; x = 1,2,\cdots,k$$

$$F(x) = \begin{cases} 0, & x < 1 \\ 1/k, & 1 \le x < 2 \\ 2/k, & 2 \le x < 3 \\ \vdots & \\ (k-1)/k, & k-1 \le x < k \\ 1, & k \le x \end{cases}$$

- If the generated random number R satisfies:

$$\boldsymbol{r_{i-1}} = \frac{\boldsymbol{i}-1}{\boldsymbol{k}} < \boldsymbol{R} \le \boldsymbol{r_i} = \frac{\boldsymbol{i}}{\boldsymbol{k}}, \textbf{ then } \boldsymbol{X} \textbf{ is generated by setting } \boldsymbol{X} = \boldsymbol{i}$$

$$\textbf{Solving the inequality we get } \boldsymbol{R.k} \le \boldsymbol{i} < \boldsymbol{R.k} + 1$$

$$\Rightarrow \boldsymbol{X} = \lceil \boldsymbol{Rk} \rceil$$

Example: *R1 = 0.78 then X1=⌈7.8⌉=8; R2 = 0.03 then X2=⌈0.03⌉=1*;

Geometric Distribution

- Consider the Geometric distribution with pmf

$$p(x) = p(1-p)^x, \;\; x = 0,1,2,\ldots.$$

- It's cdf is given by

$$F(x) = \sum_{j=0}^{x} p(1-p)^j = \frac{p\left\{1-(1-p)^{x+1}\right\}}{1-(1-p)} = 1-(1-p)^{x+1}$$

- Using the inverse transform technique, Geometric RV assume the value *x* whenever,

$$F(x-1) = 1-(1-p)^x < R \le (1-p)^{x+1} = F(x)$$

$$\Rightarrow (1-p)^{x+1} \le 1-R < (1-p)^x$$

$$\Rightarrow (x+1)\ln(1-p) \le \ln(1-R) < x\ln(1-p)$$

$$\Rightarrow \frac{\ln(1-R)}{\ln(1-p)} - 1 \le x < \frac{\ln(1-R)}{\ln(1-p)}$$

$$\therefore \text{using the round-up function, } X = \left\lceil \frac{\ln(1-R)}{\ln(1-p)} - 1 \right\rceil$$

Geometric Distribution (continued ..)

- Since p is a fixed parameter, let $\beta=-1/\ln(1-p)$.
- Then $X=\lceil -\beta \ln(1-R)-1 \rceil$. This is nothing but exponentially distributed RV with mean β (see Slide 6).
- Therefore, one way to generate a geometric variate with parameter p is to generate exponential variate with parameter $\beta^{-1}=-\ln(1-p)$, *subtract 1 and Round-up!!*
- Occasionally, there is a need to generate Geometric variate X that can assume values $\{q, q+1, q+2, \ldots\}$
- This can be generated by

$$X = q + \left\lceil \frac{\ln(1-R)}{\ln(1-p)} - 1 \right\rceil$$

Most common case is $q=1$

Example

- Generate three values from a Geometric distribution on the range *{X ≥1}* with mean 2.
- This has pmf as

$$p(x) = p(1-p)^x,\ \ x = 0,1,2,....;\ \text{with mean } 1/p = 2 \text{ or } p = 1/2$$

- Thus *X* can be generated with *q=1, p=1/2* and *1/ln(1-p)=-1.443*
- R1=0.932, X1=1+$\lceil$-1.443ln(1-0.932)-1$\rceil$=4
- R2=0.105, X1=1+$\lceil$-1.443ln(1-0.105)-1$\rceil$=1
- R3=0.687, X1=1+$\lceil$-1.443ln(1-0.687)-1$\rceil$=2

Acceptance-Rejection technique

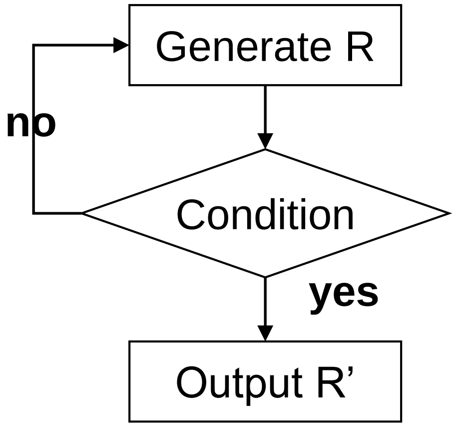
- Useful particularly when inverse cdf does not exist in closed form, a.k.a. thinning
- Illustration: To generate random variates, $X \sim U(1/4, 1)$

Procedures:

Step 1. Generate R ~ U[0,1]

Step 2a. If R >= ¼, accept X=R.

Step 2b. If R < ¼, reject R, return to Step 1



- R does not have the desired distribution, but R conditioned (R') on the event $\{R \geq 1/4\}$ does.
- Efficiency: Depends heavily on the ability to minimize the number of rejections.

Poisson Distribution

- Recall from Chapter 5, a Poisson RV, *N* with mean $\alpha>0$ has pmf

$$p(n)=P(N=n)=\frac{e^{-\alpha}\alpha^{n}}{n!},\quad n=0,1,\ldots$$

- Recall that inter-arrival times *A1, A2, …* of successive customers are exponentially distributed with rate α (i.e., α is the mean arrivals per unit time)
- Then,

$$N=n \text{ iff } A_1+A_2+\cdots+A_n\le 1<A_1+A_2+\cdots+A_n+A_{n+1}$$

Relation N=n states that there were exactly n arrivals during one unit of time
Second relation states that the n^{th} arrival occurred before time 1 while the *(n+1)* arrival occurred after time 1. Clearly these two are equivalent

Poisson Distribution (continued)

- Recall A_i can be generated by $A_i = (-1/\alpha)\ln R_i$
- Therefore,

$$A_1 + A_2 + \cdots + A_n \le 1 < A_1 + A_2 + \cdots + A_n + A_{n+1}$$

$$\sum_{i=1}^{n} -\frac{1}{\alpha} \ln R_i \le 1 < \sum_{i=1}^{n+1} -\frac{1}{\alpha} \ln R_i$$

Multiply by $-\alpha$ and use the fact that sum of logarithms is the logarithm of a product to get

$$\ln \prod_{i=1}^{n} R_i = \sum_{i=1}^{n} \ln R_i \ge -\alpha > \sum_{i=1}^{n+1} \ln R_i = \ln \prod_{i=1}^{n+1} R_i$$

Use the relation $e^{\ln x} = x$ for any number x to get

$$\prod_{i=1}^{n} R_i \ge e^{-\alpha} > \prod_{i=1}^{n+1} R_i$$

Poisson Distribution (continued)

- Therefore, the procedure for generating a Poisson random variate N is:
  - Step 1: Set $n=0; P=1$
  - Step 2: Generate a random number R_{n+1} and replace P by $P.R_{n+1}$
  - Step 3: If $P < e^{-\alpha}$, then accept $N=n$. Otherwise, reject the current n, increase n by one and return to step 2.
- How many random numbers are required to generate one Poisson variate N?
  - Answer: If $N=n$, then $n+1$ random numbers are required. So the average # is: $E(N+1)=\alpha+1$.

Example

- Generate three Poisson variates with mean α=0.2
- First compute $e^{-\alpha}$=0.8187
- Step 1: Set $n=0$, $P=1$
- Step 2: $R1=0.4357$, $P=1.R1=0.4357$
- Step 3: Since $P=0.4357< e^{-\alpha}=0.8187$, accept $N=0$
- Step 1-3. ($R1=0.4146$ leads to $N=0$)
- Step 1. Set $n=0$, $P=1$
- Step 2. $R1=0.8352$, $P=1.R1=0.8353$
- Step 3 Since P$\geq$ $e^{-\alpha}$ reject $n=0$ and return to Step 2 with $n=1$
- Step 2. $R2=0.9952$, $P=R1.R2=0.8313$
- Step 3 Since $P\geq e^{-\alpha}$ reject $n=1$ and return to Step 2 with $n=2$
- Step 2 $R3=0.8004$, $P=R1.R2.R3=0.6654$
- Step 3. Since $P < e^{-\alpha}$, accept N=2

Example (continued)

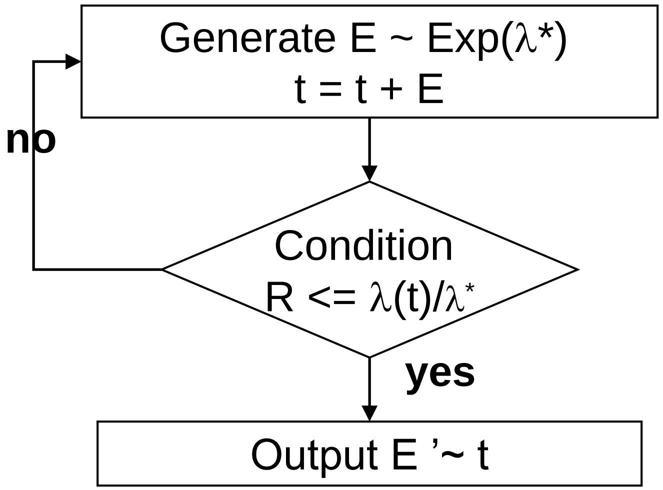
- The calculations for the generation of these Poisson random variates is summarized as:

n	R_{n+1}	P	**Accept/Reject**	Result
0	0.4357	0.4357	Accept	N=0
0	0.4146	0.4146	Accept	N=0
0	0.8353	0.8353	Reject	
1	0.9952	0.8313	Reject	
2	0.8004	0.6654	Accept	N=2

NSPP

[Acceptance-Rejection]

- Non-stationary Poisson Process (NSPP): a Possion arrival process with an arrival rate that varies with time
- Idea behind thinning:
  - Generate a stationary Poisson arrival process at the fastest rate, $\lambda^* = \max \lambda(t)$
  - But “accept” only a portion of arrivals, thinning out just enough to get the desired time-varying rate



NSPP

[Acceptance-Rejection]

- Generic algorithm that generates T_i as the time of i^{th} arrival
  - Step 1: Let $\lambda^* = \max \lambda(t)$ be the maximum arrival rate function and set $t=0$ and $i=1$
  - Step 2: Generate E from the exponential distribution with rate λ^* and let $t=t+E$ (this is the arrival time of the stationary Poisson process)
  - Step 3: Generate random number R from the $U(0,1)$ distribution. If $R \le \lambda(t)/\lambda^*$, then $T_i = t$ and $i=i+1$

NSPP

[Acceptance-Rejection]

- Example: Generate a random variate for a NSPP

Data: Arrival Rates

***t* (min)**	***Mean Time Between Arrivals* (min)**	***Arrival Rate* $\lambda(t)$ (#/min)**
0	15	1/15
60	12	1/12
120	7	1/7
180	5	1/5
240	8	1/8
300	10	1/10
360	15	1/15
420	20	1/20
480	20	1/20

Procedures:

Step 1. $\lambda^* = max\ \lambda(t) = 1/5$, $t = 0$ and $i = 1$.

Step 2. For random number $R = 0.2130$,

$E = -5ln(0.213) = 13.13$

$t = 13.13$

Step 3. Generate $R = 0.8830$

$\lambda(13.13)/\lambda^*=(1/15)/(1/5)=1/3$

Since $R>1/3$, do not generate the arrival

Step 2. For random number $R = 0.5530$,

$E = -5ln(0.553) = 2.96$

$t = 13.13 + 2.96 = 16.09$

Step 3. Generate $R = 0.0240$

$\lambda(16.09)/\lambda^*=(1/15)/(1/5)=1/3$

Since $R<1/3$, $T_1 = t = 16.09$,

and $i = i + 1 = 2$

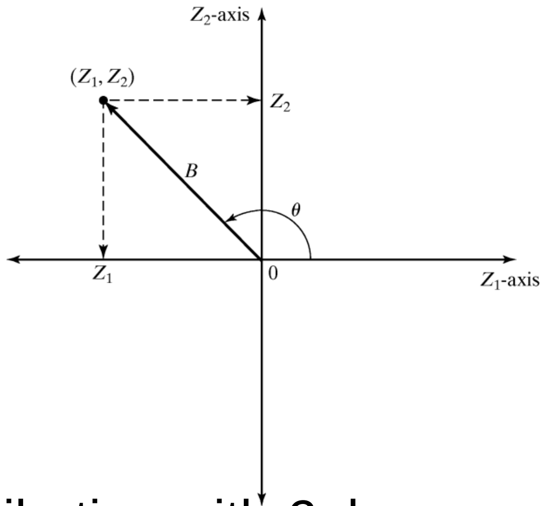
Special Properties

- Variate generate techniques that are based on features of particular family of probability distributions
- For example:
  - Direct Transformation for *normal* and *lognormal* distributions
  - Convolution
  - Beta distribution (from gamma distribution)

Direct Transformation

[Special Properties]

- Approach for *Normal(0,1)*:
  - Consider two standard normal random variables, Z_1 and Z_2, plotted as a point in the plane:



In polar coordinates:

$$Z_1 = B \cos \theta$$
$$Z_2 = B \sin \theta$$

  - $B^2 = Z^2_1 + Z^2_2$ ~ chi-square distribution with *2* degrees of freedom that is equivalent to *Exp*($\lambda = 2$). Hence, $B = (-2 \ln R)^{1/2}$
  - The radius B and angle θ are mutually independent.

$$\boxed{\begin{aligned} Z_1 &= (-2 \ln R)^{1/2} \cos(2\pi R_2) \\ Z_2 &= (-2 \ln R)^{1/2} \sin(2\pi R_2) \end{aligned}}$$

Direct Transformation [Special Properties]

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

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

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

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

Summary

- Principles of random-variate generate via
  - Inverse-transform technique
  - Acceptance-rejection technique
  - Special properties
- Important for generating continuous and discrete distributions