Learning outcomes:

- **After completing this chapter You should be able to:**
  - Understand the generating of samples from a specified distribution as input to a simulation model.
  - Illustrate some widely-used techniques for generating random variates like, Inverse-transform technique, Acceptance-rejection technique, Special properties
  - Clarify some problems with empirical distributions
Time Driven vs. Event Driven Simulation Models

- **Time Driven Dynamics**
  
  \[ x(t^+) = \begin{cases} 
    x(t) + 1 & \text{if } u_1(t) = 1, u_2(t) = 0, \\
    x(t) - 1 & \text{if } u_1(t) = 0, u_2(t) = 1, \\
    x(t) & \text{otherwise} 
  \end{cases} \]

- **Event Driven Dynamics**
  
  \[ f(x, e') = \begin{cases} 
    x + 1 & \text{if } e' = a, \\
    x - 1 & \text{if } e' = d, 
  \end{cases} \]

- In this case, time is divided in intervals of length \( \Delta t \), and the state is updated at every step.

- State is updated only at the occurrence of a discrete event.

Simulation of Discrete Event Systems

```
INITIALIZE

STATE
Update State
\( x' = f(x, e_i) \)

EVENT CALENDAR
\[
\begin{array}{cc}
  e_1 & t_1 \\
  e_2 & t_2 \\
  \vdots & \\
\end{array}
\]

Delete Infeasible Events

Add New Feasible Events

TIME
Update Time
\( t = t_1 \)

CLOCK STRUCTURE RNG
```

[Diagram of simulation process]
Simulation of Discrete Event Systems

In the mathematical fields of probability and statistics, a random variate is a particular outcome of a random variable: the random variates which are other outcomes of the same random variable might have different values. A random deviate or simply deviate is the difference of random variate with respect to the distribution central location (e.g., mean), often divided by the standard deviation of the distribution.\[1\]

Random variates are used when simulating processes driven by random influences (stochastic processes). In modern applications, such simulations would derive random variates corresponding to any given probability distribution from computer procedures designed to create random variates corresponding to a uniform distribution, where these procedures would actually provide values chosen from a uniform distribution of pseudorandom numbers.

Preparation

- It is assumed that a source of uniform [0,1] random numbers exists.
- Linear Congruential Method (LCM)

- Random numbers $R, R_1, R_2, \ldots$ with
  - PDF
    \[
    f_R(x) = \begin{cases} 
    1 & 0 \leq x \leq 1 \\
    0 & \text{otherwise}
    \end{cases}
    \]
  - CDF
    \[
    F_R(x) = \begin{cases} 
    0 & x < 0 \\
    x & 0 \leq x \leq 1 \\
    1 & x > 1
    \end{cases}
    \]
Random variate generation

Random variate generation deals with the production of random values (e.g., 26, 54, 71, 10, …) for a given random variable in such a way that the values produced form the probability distribution of the random variable.

Four fundamental approaches for random variate generation:
1. Inverse Transformations
2. (Composition will not be discussed)
3. Acceptance/Rejection
4. Special Properties

Inverse-transform Technique

The concept:
- For cdf function: \( r = F(x) \)
- Generate \( r \) from uniform \((0,1)\)
- Find \( x \):
  \[
  x = F^{-1}(r)
  \]

NOTE: Numerical inversion routines for Normal, Gamma, and Beta probability distributions are available.
Inverse-transform Technique

The inverse-transform technique can be used in principle for any distribution.
- Most useful when the CDF $F(x)$ has an inverse $F^{-1}(x)$, which is easy to compute.
- **Required steps**
  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 uniform random numbers $R_1, R_2, R_3, ...$ and compute the desired random variate by

$$X_i = F^{-1}(R_i)$$
Inverse-transform Technique (3) Exponential Distribution (2) ($\lambda=1$)

Check: Does the random variable $X_1$ have the desired distribution?

\[
P(X_1 \leq x_0) = \int_{0}^{x_0} f(x) \, dx = 1 - e^{-\lambda x}
\]

Exponential Distribution:

- Exponential cdf:

\[
r = F(x) = 1 - e^{-\lambda x}
\]

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

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

**Figure:** Inverse-transform technique for $\text{exp}(\lambda = 1)$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$R_i$</th>
<th>$X_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1306</td>
<td>0.1400</td>
</tr>
<tr>
<td>2</td>
<td>0.0422</td>
<td>0.0431</td>
</tr>
<tr>
<td>3</td>
<td>0.6597</td>
<td>1.078</td>
</tr>
<tr>
<td>4</td>
<td>0.7965</td>
<td>1.592</td>
</tr>
<tr>
<td>5</td>
<td>0.7696</td>
<td>1.468</td>
</tr>
</tbody>
</table>
Exponential Distribution

Example: Generate 200 variates $X_i$ with distribution $\text{exp}(\lambda = 1)$

- Generate 200 $Rs$ with $U(0,1)$ and utilize the following equation:

$$X_i = -\frac{1}{\lambda} \ln R_i$$

- Where

$$F(x) = 1 - e^{-x} \sim \text{exp}(\lambda = 1)$$

- Check: Does the random variable $X_i$ have the desired distribution?

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

Inverse-transform Technique: Example

Example:
- Generate 200 or 500 variates $X_i$ with distribution $\text{exp}(\lambda = 1)$
- Generate 200 or 500 $R_i$ with $U(0,1)$, the histogram of $X_i$ becomes:
Exponential Distribution

- Histogram for 200 $R_i$:
  (empirical)

- Histogram for 200 $X_i$:
  (empirical)

Exponential Distribution

- [Inverse-transform]

- Histogram for 200 $R_i$:
  (theoretical)

- Histogram for 200 $X_i$:
  (theoretical)
**Exponential Distribution**

- Generate using the graphical view:

![Exponential Distribution Graph](image)

- Draw a horizontal line from $R_1$ (randomly generated), find its corresponding $x$ to obtain $X_1$.

- We can see that:
  - $X_1 \leq x_0$ when and only when $R_1 \leq F(x_0)$
    - so: $P(X_1 \leq x_0) = P(R_1 \leq F(x_0))$
  - $R_1 \sim U(0,1)$
    - so: $P(R_1 \leq F(x_0)) = F(x_0)$
Other Distributions

- Examples of other distributions for which inverse cdf works are:
  - Uniform distribution
    \[ X \sim U(a,b) \Rightarrow X = a + (b-a)R \]
  - Weibull distribution
    (\(v = 0\))
    \[
    f(x) = \begin{cases} \frac{\beta}{\alpha} x^{\beta-1} e^{-(x/a)^\beta}, & x \geq 0 \\ 0, & otherwise \end{cases}
    \]
    \[ \Rightarrow X = \alpha[-\ln(1-R)]^{1/\beta} \]

Other Distributions

- Triangular distribution
  \[
  f(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2-x, & 1 \leq x \leq 2 \\ 0, & otherwise \end{cases}
  \]
  \[
  F(x) = \begin{cases} \frac{x^2}{2}, & 0 \leq x \leq 1 \\ \frac{(2-x)^2}{2}, & 1 < x \leq 2 \\ 1, & x > 2 \end{cases}
  \]
  - \(X\) is generated by
    \[
    X = \begin{cases} \sqrt{2R}, & 0 \leq R \leq \frac{1}{2} \\ 2-\sqrt{2(1-R)}, & \frac{1}{2} < R \leq 1 \end{cases}
    \]
Empirical Continuous Dist’n

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

Given a small sample set (size \( n \)):

- Arrange the data from smallest to largest
- Define \( x_{(0)} = 0 \)
- Assign the probability \( \frac{1}{n} \) to each interval
- The resulting empirical cdf has a \( i^{th} \) line segment slope as \( a_i \)

\[
\begin{align*}
  a_i &= \frac{x_{(i)} - x_{(i-1)}}{i/n - (i-1)/n} = \frac{x_{(i)} - x_{(i-1)}}{1/n} \\
  \text{where } x_{(0)} \leq x_{(1)} \leq \ldots \leq x_{(n)}
\end{align*}
\]

The inverse cdf is calculated by

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

where \( (i-1)/n < R \leq i/n \); \( R \) is the random number generated
Empirical Continuous Dist’n

- Five observations of fire-crew response times (in mins.):
  - 2.76
  - 1.83
  - 0.80
  - 1.45
  - 1.24

<table>
<thead>
<tr>
<th>Interval (Hours)</th>
<th>Probability</th>
<th>Cumulative Probability, ( i/n )</th>
<th>Slope, ( a_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 ≤ x ≤ 0.80</td>
<td>0.2</td>
<td>0.20</td>
<td>4.00</td>
</tr>
<tr>
<td>0.8 ≤ x ≤ 1.24</td>
<td>0.2</td>
<td>0.40</td>
<td>2.20</td>
</tr>
<tr>
<td>1.24 ≤ x ≤ 1.45</td>
<td>0.2</td>
<td>0.60</td>
<td>1.05</td>
</tr>
<tr>
<td>1.45 ≤ x ≤ 1.83</td>
<td>0.2</td>
<td>0.80</td>
<td>1.90</td>
</tr>
<tr>
<td>1.83 ≤ x ≤ 2.76</td>
<td>0.2</td>
<td>1.00</td>
<td>4.65</td>
</tr>
</tbody>
</table>

Consider \( R_1 = 0.71 \):

\[
(i-1)/n = 0.6 < R_1 < i/n = 0.8
\]

\[
X_1 = x_{(4,1)} + a_4(R_1 - (4-1)/n)
\]

\[
= 1.45 + 1.90(0.71 - 0.60) = 1.66
\]
Empirical Continuous Dist’n

For a large sample set:

- What happens for large samples of data
  - Several hundreds or tens of thousand
  - First summarize the data into a frequency distribution with smaller number of intervals
  - Afterwards, fit continuous empirical CDF to the frequency distribution
  - Slight modifications
  - Slope

\[ X = F^{-1}(R) = x_{(i-1)} + a_i (R - c_{(i-1)}) \]

Where

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

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

<table>
<thead>
<tr>
<th>Interval (Hours)</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Cumulative Frequency, ( c_i )</th>
<th>Slope, ( a_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0.25 ≤ x ≤ 0.5</td>
<td>31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.81</td>
</tr>
<tr>
<td>2 0.5 ≤ x ≤ 1.0</td>
<td>10</td>
<td>0.10</td>
<td>0.41</td>
<td>5.0</td>
</tr>
<tr>
<td>3 1.0 ≤ x ≤ 1.5</td>
<td>25</td>
<td>0.25</td>
<td>0.66</td>
<td>2.0</td>
</tr>
<tr>
<td>4 1.5 ≤ x ≤ 2.0</td>
<td>34</td>
<td>0.34</td>
<td>1.00</td>
<td>1.47</td>
</tr>
</tbody>
</table>

Consider \( R_1 = 0.83 \):

\[ c_3 = 0.66 \leq R_1 < c_4 = 1.00 \]

\[ X_i = x_{(i-1)} + a_i (R_1 - c_{(i-1)}) \]

\[ = 1.5 + 1.47(0.83 - 0.66) \]

\[ = 1.75 \]
Inverse-transform Technique:
Empirical Continuous Distributions

- Problems with empirical distributions
  - The data in the previous example is restricted in the range $0.25 \leq X \leq 2.0$
  - The underlying distribution might have a wider range
  - Thus, try to find a theoretical distribution
  - Hints for building empirical distributions based on frequency tables
  - It is recommended to use relatively short intervals
    - Number of bins increase
  - This will result in a more accurate estimate

Inverse-transform Technique:
Discrete Distribution

- A number of continuous distributions do not have a closed form expression for their CDF, e.g.
  - Normal
  $F(x) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x - \mu}{\sigma \sqrt{2}} \right) \right]$
  - Gamma
  - Beta
- The presented method does not work for these distributions
- Solution
  - Approximate the CDF or numerically integrate the CDF
- Problem
  - Computationally slow
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
  - Gamma

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

- Data - Probability distribution:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p(x)$</th>
<th>$F(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>1.00</td>
</tr>
</tbody>
</table>

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

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

Consider $R_1 = 0.73$:

- $F(x_2) < R_1 \leq F(x_1)$
- $F(x_0) < 0.73 \leq F(x_1)$

Hence, $x_1 = 1$
Discrete Distribution

The inverse-transform technique as table-lookup procedure

\[ F^{-1}(X_{i-1}) = r_{i-1} < R \leq r_i = F^{-1}(X_i) \]

Set \( X = x_i \).

Method - Given \( X \), the generation scheme becomes:

\[
\begin{align*}
x & = \begin{cases} 
0, & R \leq 0.5 \\
1, & 0.5 < R \leq 0.8 \\
2, & 0.8 < R \leq 1.0
\end{cases}
\end{align*}
\]

Table for generating the discrete variate \( X \):

<table>
<thead>
<tr>
<th>( i )</th>
<th>( \text{Input} \ r_i )</th>
<th>( \text{Output} \ x_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>2</td>
</tr>
</tbody>
</table>

Consider \( R_i = 0.73 \):

\[ F(X_{0.9}) < R \leq F(X_{1.0}) \]

\[ F(X_{0.9}) < 0.73 \leq F(X_{1.0}) \]

Hence, \( X_i = 1 \)

Acceptance-Rejection technique

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

Procedure:

Step 1. Generate \( R \sim U(0,1) \)
Step 2. If \( R \geq \frac{1}{4} \), accept \( X = R \).
Step 3. If \( R < \frac{1}{4} \), reject \( R \), return to Step 1

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

Example: Generate three Poisson variates with mean $\alpha = 0.2$
- $\exp(-0.2) = 0.8187$
- Variate 1
  - Step 1: Set $n = 0$, $P = 1$
  - Step 2: $R_1 = 0.4357$, $P = 1 \times 0.4357$
  - Step 3: Since $P = 0.4357 < \exp(-0.2)$, accept $N = 0$
- Variate 2
  - Step 1: Set $n = 0$, $P = 1$
  - Step 2: $R_1 = 0.4146$, $P = 1 \times 0.4146$
  - Step 3: Since $P = 0.4146 < \exp(-0.2)$, accept $N = 0$
- Variate 3
  - Step 1: Set $n = 0$, $P = 1$
  - Step 2: $R_1 = 0.8353$, $P = 1 \times 0.8353$
  - Step 3: Since $P = 0.8353 > \exp(-0.2)$, reject $n = 0$ and return to Step 2 with $n = 1$
  - Step 2: $R_2 = 0.9952$, $P = 0.8353 \times 0.9952 = 0.8313$
  - Step 3: Since $P = 0.8313 > \exp(-0.2)$, reject $n = 1$ and return to Step 2 with $n = 2$
  - Step 2: $R_3 = 0.8004$, $P = 0.8313 \times 0.8004 = 0.6654$
  - Step 3: Since $P = 0.6654 < \exp(-0.2)$, accept $N = 2$
Acceptance-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!

<table>
<thead>
<tr>
<th>$N$</th>
<th>$R_n$</th>
<th>$P$</th>
<th>Accept/Reject</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4357</td>
<td>0.4357</td>
<td>$P &lt; \exp(-\alpha)$</td>
<td>Accept</td>
</tr>
<tr>
<td>0</td>
<td>0.4146</td>
<td>0.4146</td>
<td>$P &lt; \exp(-\alpha)$</td>
<td>Accept</td>
</tr>
<tr>
<td>0</td>
<td>0.8353</td>
<td>0.8353</td>
<td>$P \geq \exp(-\alpha)$</td>
<td>Reject</td>
</tr>
<tr>
<td>1</td>
<td>0.9952</td>
<td>0.8313</td>
<td>$P \geq \exp(-\alpha)$</td>
<td>Reject</td>
</tr>
<tr>
<td>2</td>
<td>0.8004</td>
<td>0.6654</td>
<td>$P &lt; \exp(-\alpha)$</td>
<td>Accept</td>
</tr>
</tbody>
</table>

NSPP

- 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

![Flowchart](https://via.placeholder.com/150)

Generate $E \sim \text{Exp}(\lambda^*)$

$t = t + E$

- **Condition**: $R \leq \lambda(t)$

- **Yes**
  - Output $E \sim t$

- **No**
Example: Generate a random variate for a NSPP

Data: Arrival Rates

<table>
<thead>
<tr>
<th>Mean Time Between Arrivals (min)</th>
<th>Arrival Rate ( \lambda(t) ) (#/min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t (min)</td>
<td>t (min)</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>60</td>
<td>12</td>
</tr>
<tr>
<td>120</td>
<td>7</td>
</tr>
<tr>
<td>180</td>
<td>5</td>
</tr>
<tr>
<td>240</td>
<td>8</td>
</tr>
<tr>
<td>300</td>
<td>10</td>
</tr>
<tr>
<td>360</td>
<td>15</td>
</tr>
<tr>
<td>420</td>
<td>20</td>
</tr>
<tr>
<td>480</td>
<td>20</td>
</tr>
</tbody>
</table>

Procedures:

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

Step 2. For random number \( R = 0.2130 \),
\[
E = -5 \ln(0.213) = 13.13
\]
\( t = 13.13 \)

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 \( R = 0.5530 \),
\[
E = -5 \ln(0.553) = 2.96
\]
\( t = 13.13 + 2.96 = 16.09 \)

Step 3. Generate \( R = 0.0240 \)
\[
\lambda(16.09)/\lambda^*=(1/15)/(1/5)=1/3
\]
Since \( R < 1/3 \), \( T_i = t = 16.09 \),
and \( i = i + 1 = 2 \)

Special Properties

- 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

- **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:
      \[
      Z_1 = B \cos \phi \\
      Z_2 = B \sin \phi
      \]
    - \(B^2 = Z_1^2 + Z_2^2 \sim \chi^2\) distribution with 2 degrees of freedom = \(Exp(\lambda = 2)\). Hence, \(B = (-2 \ln R)^{1/2}\)
    - The radius \(B\) and angle \(\phi\) are mutually independent.

\[
\begin{align*}
Z_1 &= (-2 \ln R)^{1/2} \cos(2\pi R_1) \\
Z_2 &= (-2 \ln R)^{1/2} \sin(2\pi R_2)
\end{align*}
\]

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

\[
X_i = \mu + \sigma Z_i
\]

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

\[
Y_i = e^{X_i}
\]
Summary

- Principles of random-variate generate via
  - Inverse-transform technique
  - Acceptance-rejection technique
  - Special properties

- Important for generating continuous and discrete distributions