Random-Variate Generation

Overview

• Generation of samples from a specified distribution as input to a simulation model

• Widely-used techniques for generating random variates
  • Inverse-transform technique
  • Acceptance-rejection technique
  • Direct transformation
Inverse-transform Technique

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

Exponential Distribution

\[
F(x) = \int_{-\infty}^{x} f(t) \, dt = \begin{cases} 
1 - e^{-\lambda x}, & x \geq 0 \\
0, & x < 0 
\end{cases}
\]

Generate \( X_1, X_2, X_3, \ldots \) that follow the exponential distribution

\[ X = -\frac{1}{\lambda} \ln(1 - R) \]

This is called the random-variate generator for the exponential distribution

A common simplification

\( R \sim U[0, 1] \Rightarrow 1 - R \sim U[0, 1] \)

\[ X_i = -\frac{1}{\lambda} \ln R_i \]
Exponential Distribution

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

Generate $200 Rs \sim U(0,1)$

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

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

\[ F(x) = \begin{cases} 
0, & x < a \\
\frac{x-a}{b-a}, & a \leq x \leq b \\
1, & x > b 
\end{cases} \]

\[ F(X) = (X - a)/(b - a) = R \]

\[ X = a + (b - a)R \]

Empirical Continuous Distribution

When theoretical distribution is not applicable

Collect empirical data

Interpolate between observed data points to fill in the gaps

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

Arrange data from smallest to largest \( x(0) = 0 \leq x(1) \leq x(2) \leq \ldots \leq x(n) \)

Assign probability \( 1/n \) to each interval \( x(i-1) \leq x \leq x(i) \)

The slope of the \( i \)th segment

\[ a_i = \frac{x(i) - x(i-1)}{i/n - (i-1)/n} = \frac{x(i) - x(i-1)}{1/n} \]

The inverse cdf

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

When large amounts of data are available

Summarize data into a frequency distribution

Slope of the $i$th line segment $a_i = \frac{x_i - x_{i-1}}{c_i - c_{i-1}}$ (for inverse function)

$c_i$ = cumulative probability of first $i$ intervals, $x_{i-1} < x \leq x_i$

The inverse cdf $X = F^{-1}(R) = x_{i-1} + a_i (R - c_{i-1})$

Where $c_{i-1} < R \leq c_i$

---

Empirical Continuous Distribution

Example: data collected for 100 machine repair times

<table>
<thead>
<tr>
<th>$i$</th>
<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</td>
<td>$0.25 \leq x \leq 0.5$</td>
<td>31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>$0.5 &lt; x \leq 1.0$</td>
<td>10</td>
<td>0.10</td>
<td>0.41</td>
<td>5.0</td>
</tr>
<tr>
<td>3</td>
<td>$1.0 &lt; x \leq 1.5$</td>
<td>25</td>
<td>0.25</td>
<td>0.66</td>
<td>2.0</td>
</tr>
<tr>
<td>4</td>
<td>$1.5 &lt; x \leq 2.0$</td>
<td>34</td>
<td>0.34</td>
<td>1.00</td>
<td>1.47</td>
</tr>
</tbody>
</table>

Consider $R_1 = 0.83$

$c_3 = 0.66 < R_1 < c_4 = 1.00$

$x_1 = x_{a(1)} + a_1(R_1 - c_{a(1)})$

$= 1.5 + 1.47(0.83 - 0.66)$

$= 1.75$
Discrete Distribution

All discrete distributions can be generated via inverse-transform technique

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

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

<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:

Consider $R_i = 0.73$

$F(x_{i-1}) < R \leq F(x_i)$

$F(x_0) < 0.73 \leq F(x_1)$

Hence, $X_i = 1$
Acceptance-Rejection Technique

Useful particularly when inverse cdf does not exist in closed form, a.k.a. thinning

Illustration:
To generate random variates, $X \sim U(1/4, 1)$

Step 1. Generate $R \sim U[0,1]$
Step 2a. If $R \geq 1/4$, accept $X=R$
Step 2b. If $R < 1/4$, 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

Poisson has pmf $p(n) = P(N = n) = \frac{e^{-\alpha} \alpha^n}{n!}, \ n = 0, 1, 2, \ldots$

$N$=no. of arrivals in one time unit
Interarrival times $A_1, A_2, \ldots$ are exponentially distributed with rate $\alpha$
$\alpha$ = mean no. of arrivals per time unit

The relationship between the Poisson distribution and the exponential distribution is

$N=n$ if and only if $A_1 + A_2 + \ldots + A_n \leq 1 < A_1 + \ldots + A_n + A_{n+1}$

There were exactly $n$ arrivals before time 1
The $n$th arrival occurred before 1
The $n+1$st arrival occurred after 1
Generate exponential $A$'s until $A_{n+1}$ occurs after 1
Set $N=n$
Poisson Distribution

Simplify for efficiency

\[ A_i = (-\frac{1}{\alpha}) \ln R_i \]

\[ \sum_{i=1}^{n} -\frac{1}{\alpha} \ln R_i \leq 1 < \sum_{i=1}^{n+1} -\frac{1}{\alpha} \ln R_i \]

\[ \ln \prod_{i=1}^{n} R_i = \sum_{i=1}^{n} \ln R_i \geq -\alpha > \sum_{i=1}^{n+1} \ln R_i = \ln \prod_{i=1}^{n+1} R_i \]

\[ \prod_{i=1}^{n} R_i \geq e^{-\alpha} > \prod_{i=1}^{n+1} R_i \]

Final procedure

Step 1. \( n=0, P=1 \)

Step 2. Generate \( R_{n+1}, P=P(R_{n+1}) \)

Step 3. If \( P<e^{-\alpha} \), accept \( N=n \). Else, reject \( n, n=n+1 \), goto step 2.

To generate \( N=n \), \( n+1 \) random numbers are required, the average is

\[ E(n+1) = \alpha + 1 \]

For large \( \alpha \), this is inefficient
Example: Generate 3 Poisson variates with mean $\alpha = 0.2$

$e^{0.2} = 0.8187$

**Step 1.** Set $n = 0, P = 1$.

**Step 2.** $R_1 = 0.4357, P = 1 \cdot R_1 = 0.4357$.

**Step 3.** Since $P = 0.4357 < e^{-\alpha} = 0.8187$, accept $N = 0$.

**Step 1-3.** ($R_1 = 0.4357$ leads to $N = 0$.)

**Step 1.** Set $n = 0, P = 1$.

**Step 2.** $R_1 = 0.8353, P = 1 \cdot R_1 = 0.8353$.

**Step 3.** Since $P \geq e^{-\alpha}$, reject $n = 0$ and return to Step 2 with $n = 1$.

**Step 2.** $R_2 = 0.9952, P = R_1R_2 = 0.8313$.

**Step 3.** Since $P \geq e^{-\alpha}$, reject $n = 1$ and return to Step 2 with $n = 2$.

**Step 2.** $R_3 = 0.8004, P = R_1R_2R_3 = 0.6654$.

**Step 3.** Since $P < e^{-\alpha}$, accept $N = 2$.

---

**Poisson Distribution**

For large $\alpha$ ($\alpha \geq 15$),

$$Z = \frac{N - \alpha}{\sqrt{\alpha}}$$

is approximately normal standard

Generate a standard normal variate $Z$ (explained later)

Generate Poisson variate $N = \lfloor \alpha + \sqrt{\alpha}Z - 0.5 \rfloor$

Ceiling and 0.5 is to round-up to nearest integer
Non-Stationary Poisson Process (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, \( \lambda^* = \max \lambda(t) \)
"Accept" only a portion of arrivals, thinning out just enough to get the desired time-varying rate

Example: Generate a random variate for a NSPP

<table>
<thead>
<tr>
<th>Step</th>
<th>( \lambda^* = \max \lambda(t) = 1/5, t = 0 ) and ( i = 1 )</th>
<th>( \lambda(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>( T ) (min)</td>
<td>MIAT (min)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>For random number ( R = 0.2130 ), ( E = -5\ln(0.213) = 13.13 )</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( t = 0 + 13.13 = 13.13 )</td>
<td>60</td>
</tr>
<tr>
<td>Step 3</td>
<td>Generate ( R = 0.8830 )</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>( \lambda(13.13)/\lambda^* = (1/15)/(1/5) = 1/3 )</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>Since ( R &gt; 1/3 ), do not generate the arrival</td>
<td>( T_1 = t = 13.13 )</td>
</tr>
<tr>
<td>Step 2</td>
<td>For random number ( R = 0.5530 ), ( E = -5\ln(0.553) = 2.96 )</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>( t = 13.13 + 2.96 = 16.09 )</td>
<td>300</td>
</tr>
<tr>
<td>Step 3</td>
<td>Generate ( R = 0.0240 )</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>( \lambda(16.09)/\lambda^* = (1/15)/(1/5) = 1/3 )</td>
<td>420</td>
</tr>
<tr>
<td></td>
<td>Since ( R &lt; 1/3 ), ( T_1 = t = 16.09, i = i + 1 = 2 )</td>
<td>480</td>
</tr>
</tbody>
</table>
Direct Transformation for Normal Distribution

Inverse transform cannot be applied, since Normal does not have inverse

For normal(0, 1)
Consider two standard normal random variables, \( Z_1 \) and \( Z_2 \)

In polar coordinates:
\[
Z_1 = B \cos \theta \\
Z_2 = B \sin \theta
\]

\( B^2 = Z_1^2 + Z_2^2 \sim \text{chi-square with 2 DOF} = \text{Exp}(\lambda = 2) \)
\( B = (-2 \ln R)^{1/2} \)
The radius \( B \) and angle \( \theta \) are mutually independent.
\( Z_1 = (-2 \ln R_1)^{1/2} \cos(2\pi R_2) \)

\( R_1 \) – random to generate \( B \)
\( R_2 \) – random to generate \( \theta \)

Direct Transformation for Normal Distribution

Approach for \( N(\mu, \sigma^2) \)

Generate \( Z_i \sim N(0, 1) \)
\( X_i = \mu + \sigma Z_i \)

Example:
\( R_1 = 0.1758, R_2 = 0.1489 \)
\( Z_1 = (-2 \ln(0.1758))^{1/2} \cos(2\pi 0.1489) = 1.11 \)
\( Z_2 = (-2 \ln(0.1758))^{1/2} \sin(2\pi 0.1489) = 1.50 \)

\( \mu=10, \ \sigma=2 \)
\( X_1 = 10 + 2(1.11) = 12.22 \)
\( X_2 = 10 + 2(1.50) = 13.00 \)