Math 416/516: Stochastic Simulation

Haijun Li

lih@math.wsu.edu
Department of Mathematics
Washington State University

Week 5
Outline

1. Inverse Transform Method
2. Acceptance-Rejection Method
Generalized Inverse Function

Recall: Any cumulative distribution function (cdf) $F(x)$ is a non-decreasing, right-continuous function from zero at $-\infty$ to 1 at $+\infty$.

**Definition**

Let $F(x)$ be a cdf. Then the generalized inverse of $F(x)$ is defined by

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad \forall \ u \in [0, 1].$$

- If $F(x)$ is continuous and strictly increasing, then $F^{-1}(u)$ is the usual inverse function of $F(x)$. That is,

$$F(F^{-1}(u)) = u, \quad \text{for } 0 \leq u \leq 1; \quad F^{-1}(F(x)) = x, \quad \text{for } x \in \mathbb{R}.$$

- $F(x) \geq u$ if and only if $x \geq F^{-1}(u)$.

- In general,

$$F(F^{-1}(u)) \geq u, \quad \text{for } 0 \leq u \leq 1; \quad F^{-1}(F(x)) \leq x, \quad \text{for } x \in \mathbb{R}.$$
Inverse Transform Method

**Theorem**

Let $F(x)$ be a cdf and $U$ be uniformly distributed over $[0, 1]$. Then

$$X = F^{-1}(U)$$

has the cdf $F(x)$.

**Algorithm (Sampling from $F$):**

1. Draw a number $u$ from $[0, 1]$ at random.
2. Calculate $x = F^{-1}(u)$. 
Example: Sampling from exponential distributions

- Let $X \sim \exp(\lambda)$; that is, $F(x) = 1 - e^{-\lambda x}$, $x \geq 0$.
- $E(X) = \lambda^{-1}$, $\text{var}(X) = \lambda^{-2}$. $\lambda^{-1}$ = scale parameter.
- Simulate a set of samples $x_1, \ldots, x_n$ from this distribution.

Algorithm:
1. Draw $n$ random numbers $u_1, \ldots, u_n$ from $[0, 1]$ independently.
2. Calculate

$$x_i = F^{-1}(u_i) = -\frac{1}{\lambda} \log(1 - u_i), \quad i = 1, 2, \ldots, n.$$
Example: Sampling from Fréchet distributions

Let $X \sim \text{Frechet}(\alpha)$; that is, $F(x) = e^{-x^{-\alpha}}$, $x \geq 0$. The shape parameter $\alpha > 0$ is known as the heavy tail index.

If $0 < \alpha \leq 1$, $E(X) = \infty$. If $0 < \alpha \leq 2$, $\text{var}(X) = \infty$.

For financial applications, $\alpha > 2$.

In Internet data analysis, $\alpha$ can be smaller than 1.


Algorithm:

1. Draw $n$ random numbers $u_1, \ldots, u_n$ from $[0, 1]$ independently.
2. Calculate

$$x_i = F^{-1}(u_i) = (- \log u_i)^{-1/\alpha}, \quad i = 1, 2, \ldots, n.$$

3. A set of samples $\{x_1, \ldots, x_n\}$ is generated from Frechet($\alpha$).
Sampling from Bernoulli($p$) via Inverse Transform

Let $X$ be discrete with a Bernoulli distribution; that is,

$P(X = 1) = p, \quad P(X = 0) = 1 - p$, where $0 \leq p \leq 1$.

$E(X) = p$, and $\text{var}(X) = p(1 - p)$.

Let $F(x)$ denote the cdf of $X$, then

$$F^{-1}(u) = \begin{cases} 
0 & \text{if } 0 \leq u \leq 1 - p \\
1 & \text{if } 1 - p < u \leq 1.
\end{cases}$$

Algorithm:

1. Draw $n$ random numbers $u_1, \ldots, u_n$ from $[0, 1]$ independently.
2. Calculate

$$y_i = F^{-1}(u_i) = \begin{cases} 
0 & \text{if } 0 \leq u_i \leq 1 - p \\
1 & \text{if } 1 - p < u_i \leq 1.
\end{cases}, \quad i = 1, 2, \ldots, n.$$

3. A set of samples $\{y_1, \ldots, y_n\}$ is generated from Bernoulli($p$).
Discrete Distributions

Let $X$ be discrete with probability mass function:

$$p_i = P(X = v_i), \ i = 1, \ldots, m,$$

where $v_1 \leq v_2 \leq \cdots \leq v_m$ are all the possible values of $X$.

Let $F(x)$ denote its cdf, then

$$F(x) = \begin{cases} 
0 & \text{if } x < v_1 \\
\sum_{i=1}^{k} p_i & \text{if } v_k \leq x < v_{k+1} \\
1 & \text{if } v_m \leq x.
\end{cases}$$

$F^{-1}(u) = v_k$ if $\sum_{i=1}^{k-1} p_i < u \leq \sum_{i=1}^{k} p_i$. 
Sampling from Discrete Distributions

Note that $\sum_{i=1}^{m} p_i = 1$, and $p_1, \ldots, p_m$ are all non-negative.

Algorithm:

1. Draw a random numbers $u$ from $[0, 1]$.
2. Calculate

$$x = F^{-1}(u) = \begin{cases} v_1 & \text{if } 0 \leq u \leq p_1 \\ v_k & \text{if } \sum_{i=1}^{k-1} p_i < u \leq \sum_{i=1}^{k} p_i, \ 1 < k < m \\ v_m & \text{if } \sum_{i=1}^{m-1} p_i < u \leq 1 \end{cases}$$

3. Repeat (1) and (2) $n$ times to generate a set of samples \{x_1, \ldots, x_n\} from the discrete cdf $F(x)$. 
Remarks

- The inverse transform method is useful and effective especially for sampling from univariate distributions.
- The inverse transform method can be applied to generating samples from multivariate distributions. For example, to generate a sample from $(X, Y)$ with joint cdf $F(x, y)$, using the following two-step procedure:

  1. Generating a value $x$ from the marginal distribution $F_1(x)$ of $X$.
  2. Generating a value $y$ from the conditional distribution $F_{2|1}(y|x)$ of $Y$ conditioning on $X = x$.

Both steps can be implemented by using the inverse transform method, but this method becomes intractable for high-dimensional distributions.
Generate more “easy samples”, then reject some.

- Suppose that a (target) probability density function $f(x)$ is difficult to simulate directly.
- But we know how to simulate samples from an alternative probability density $g(x)$, where $g(x) \geq f(x)$.
- We simulate more samples from $g(x)$, and then accept a portion of them based on some rules.
Acceptance-Rejection Method

Again, let \( f(x) \) and \( g(x) \) denote two probability density functions such that \( cg(x) \geq f(x) \) for some constant \( c \geq 1 \).

**Theorem**

Let \( X \sim f(x) \), and \( Y \sim g(y) \). If \( U \) is uniformly distributed over \([0, 1]\) and independent of \( Y \), then for any subset \( A \),

\[
P(X \in A) = P\left( Y \in A \left| U \leq \frac{f(Y)}{cg(Y)} \right. \right).
\]

That is,

\[
X \overset{d}{=} \left[ Y \left| U \leq \frac{f(Y)}{cg(Y)} \right. \right]
\]
Proof. Observe that

\[ P\left( U \leq \frac{f(Y)}{cg(Y)} \right) = \int P\left( U \leq \frac{f(y)}{cg(y)} \mid Y = y \right) g(y)dy \]

\[ = \int P\left( U \leq \frac{f(y)}{cg(y)} \right) g(y)dy = \int \frac{f(y)}{cg(y)} g(y)dy = \frac{1}{c}. \]

Similarly, for any set \( A \),

\[ P\left( Y \in A, U \leq \frac{f(Y)}{cg(Y)} \right) = \int_A P\left( U \leq \frac{f(y)}{cg(y)} \right) g(y)dy \]

\[ = \int_A \frac{f(y)}{cg(y)} g(y)dy = \frac{1}{c} \int_A f(y)dy. \]

Therefore,

\[ P\left( Y \in A \middle| U \leq \frac{f(Y)}{cg(Y)} \right) = \frac{P\left( Y \in A, U \leq \frac{f(Y)}{cg(Y)} \right)}{P\left( U \leq \frac{f(Y)}{cg(Y)} \right)} \]

\[ = \int_A f(y)dy = P(X \in A). \]

Q.E.D.
Acceptance-Rejection Algorithm

Again, let \( f(x) \) and \( g(x) \) denote two probability density functions such that \( cg(x) \geq f(x) \) for some constant \( c \geq 1 \).

Algorithm:

1. Generate a trial sample value \( y \) from the density \( g(y) \).
2. Generate a random number \( u \) from the uniform distribution over \([0, 1] \).
3. Accept \( y \) and set \( x := y \) if

\[
u \leq \frac{f(y)}{cg(y)};
\]

otherwise, discard \( y \) and go to (1).

Haijun Li
Math 416/516: Stochastic Simulation
Week 5 14 / 23
Remark

- The overall probability of accepting a trial sample value is $1/c$, and so on average, about $c$ sample values from $g(y)$ are needed to generate one acceptable sample value from $f(x)$.
- $c$ can be selected as

$$c^* = \max_{\text{all } x\text{'s}} \frac{f(x)}{g(x)}.$$

A smaller $c$ that is close to 1 is more efficient.
- The uniform and exponential distributions are often chosen as the alternative distribution $g(x)$.
- The Acceptance-Rejection method can be applied to Monte Carlo simulations from multi-dimensional distributions.
Gamma Distribution

- A random variable $X$ is said to have a **gamma distribution** if it has the pdf
  
  $$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}, \ x > 0,$$
  
  where $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} \, dx$.

- If $k$ is a positive integer, then the gamma function $\Gamma(k) = (k - 1)!$.

- Widely used in reliability engineering and telecommunication.

- $E(X) = \frac{k}{\lambda}$, $V(X) = \frac{k}{\lambda^2}$.

- $k = \text{shape parameter}$, $\lambda = \text{rate parameter}$, $\theta = \lambda^{-1} = \text{scale parameter}$. 
A special case: When $k = n$ (integer), the distribution is called an **Erlang distribution** (widely used in telecommunication).

$k = 1$: Exponential distribution. The pdf is given by $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$. 

![Graph of Erlang distributions for different values of $k$](image)
Example: Sampling from Gamma Dist

Consider the case where $k = 3/2$, $\lambda = 1$ (note: $\Gamma(3/2) = \sqrt{\pi}/2$):

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

To simulate samples from this distribution, we use the exponential distribution as the alternative distribution $g(x) = \lambda e^{-\lambda x}, \ x \geq 0$.

Find the constant $c$ such that

$$c = \max_{x \geq 0} \frac{f(x)}{g(x)} = \max_{x \geq 0} \frac{2\sqrt{x}e^{-x}}{\sqrt{\pi} \lambda e^{-\lambda x}}.$$  

The maximum is attained at $x^* = 0.5/(1 - \lambda)$ and

$$c = \frac{1}{\sqrt{2e\pi\lambda^2(1 - \lambda)}}.$$  

The smallest $c$ is most efficient. $c$ is minimized at $\lambda^* = 2/3$ with minimal value $c^* = 1.257$.

That is, $g(x) = \frac{2}{3} e^{-2x/3}$ and $c^* = 1.257.$
Example: Sampling from Gamma Dist (cont’d)

Algorithm:

1. Generate a trial sample $y$ from the exponential distribution with rate $2/3$.
2. Generate a random number $u$ from the uniform distribution over $[0, 1]$.
3. Accept $y$ and set $x := y$ if

$$u \leq \frac{f(y)}{cg(y)} = 2\sqrt{\frac{2ey}{3}}e^{-y/3};$$

otherwise, discard $y$ and go to (1).

Remark: In order generate 100 samples from $f(x)$, we need to generate about 125.7 trial samples from $g(y)$. 
Dealing With Unknown Normalizing Constant

Suppose that a probability density function

\[ f(x) = k \, h(x), \quad \text{denoted by } f(x) \propto h(x) \]

where \( h(x) \) is a known function, but the normalizing constant \( k \) is unknown.

Let \( g(y) \) be a density function from which we know how to draw samples. If

\[ \frac{h(y)}{g(y)} \leq c, \quad \forall \ y, \]

then we can use the acceptance rule

\[ u \leq \frac{h(y)}{cg(y)} \]

where \( y \) is drawn from \( g(y) \) and \( u \) is drawn from the uniform distribution over \([0, 1]\) independently.
Example: Unknown Normalizing Constant

Simulate samples from the following density function:

\[ f(x) \propto e^{-x^2} \left(1 - e^{-\sqrt{1+x^2}}\right), \quad x \in \mathbb{R}. \]

Observe that

\[ e^{-x^2} \left(1 - e^{-\sqrt{1+x^2}}\right) \leq \frac{1}{\sqrt{2\pi}} \phi(x) \]

where \( \phi(x) \) is the standard normal density function.

Algorithm:

1. Generate a trial sample \( y \) from the normal distribution \( N(0, 1) \).
2. Generate a random number \( u \) from the uniform distribution over \([0, 1]\).
3. Accept \( y \) and set \( x := y \) if

\[ u \leq \frac{h(y)}{\sqrt{2\pi} \phi(y)} = \frac{e^{-y^2} \left(1 - e^{-\sqrt{1+y^2}}\right)}{e^{-y^2/2}}; \]

otherwise, discard \( y \) and go to (1).
Example: 2-Dimensional Simulation

Simulate samples from the following density function:

\[ f(x, y) = \begin{cases} 1/\pi & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Let

\[ g(x, y) = \begin{cases} 1/4 & \text{if } -1 \leq x, y \leq 1 \\ 0 & \text{otherwise} \end{cases} \]

Note that \( g(x, y) \) is the joint pdf of two uniformly distributed random variables over \([-1, 1]\).

Observe that

\[ c = \max_{-1 \leq x, y \leq 1} \frac{f(x, y)}{g(x, y)} = \frac{4}{\pi} \approx 1.273. \]

It is easy to see that

\[ \frac{f(x, y)}{c g(x, y)} = \begin{cases} 1 & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \]
Algorithm:

1. Generate a pair of independent trial samples \((x, y)\) from the uniform distribution over \([-1, 1]\).
2. Generate a random number \(u\) from the uniform distribution over \([0, 1]\).
3. Accept the pair \((x, y)\) if
   \[
   u \leq \frac{f(x, y)}{cg(x, y)} = 1, \text{ equivalently } x^2 + y^2 \leq 1;
   \]
   otherwise, discard \((x, y)\) and go to (1).

Remark: In order to generate 100 samples, we need to generate about 127.3 trial pairs from \(g(x, y)\).