Université Mohamed BOUDIAF M'Sila
Faculty of Mathematics and Computer Science

Department of Mathematics, 3rd Year Applied Mathematics, Semester 6

Module: Software Simulation and Practice

Module Coordinator:
Dr. BOUHADJAR El Mountasar Billah

Chapter 2: Sample Generation According to Different Probability Laws

Table of Contents

Chapter 1

Sample Generation According to Different Probability Laws

1.1 Introduction

In this chapter, we will explore methods for generating random samples from different probability distributions. These techniques are essential in simulation and modeling, as they allow us to reproduce random behavior in complex systems.

We will discuss three main methods:

- **The inverse transform method**: This method, applicable to both discrete and continuous cases, uses the cumulative distribution function to generate samples from a given probability distribution.
- **The rejection method**: This approach generates samples by "rejecting" values that do not satisfy certain conditions, while ensuring that the accepted samples follow the desired distribution.
- **The composition method**: This method combines several probability distributions to generate more complex samples.

By the end of this chapter, you will be able to choose and apply the most appropriate method for generating samples from a given probability distribution, a key skill in simulation and statistical analysis.

1.2 The Inverse Transform Method

Cumulative distribution function: Let X be a discrete or continuous random variable. The cumulative distribution function (CDF) of X is the function F such that, for all $n \in \mathbb{R}$:

$$\begin{aligned} F(n) &= P(X \leq n) \\ &= \int_{-\infty}^{n} f(t)\,dt. \end{aligned}$$

which satisfies:

- $F(n) \in [0,1]$, $\forall n \in \mathbb{R}$
- F is non-decreasing on $\mathbb{R}$
- $\lim_{n\to-\infty} F(n) = 0, \quad \lim_{n\to+\infty} F(n) = 1$
- F is right-continuous.

Exp1: Let $X \sim U[a,b]$, then

$$f(n) = \begin{cases} 0, & n \notin [a,b] \\ \frac{1}{b-a}, & n \in [a,b] \end{cases} = \frac{1}{b-a}\mathbf{1}_{[a,b]}$$

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

Example 1.2.1. *Let $T \sim U[0,1]$*

$$f(t) = \mathbf{1}_{[0,1]}(t) = \begin{cases} 0, & t \notin [0,1] \\ 1, & t \in [0,1] \end{cases}$$

$$F(t) = \begin{cases} 0, & t < 0 \\ t, & 0 \leq t \leq 1 \\ 1, & t \geq 1 \end{cases}$$

Remark 1. *If $X \sim U[a,b]$, then $T = \frac{X-a}{b-a}$ follows $U[0,1]$. Conversely, if $T \sim U[0,1]$, then $X = a + (b-a)T$ follows $U[a,b]$.*

Example 1.2.2. *If $X \sim \mathcal{N}(0,1)$, then the density of X is:*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R},$$

the cumulative distribution function F is:

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

F has no closed-form expression (i.e., it cannot be written in terms of elementary functions).

2.1 Principle of the Method

Let X be a real-valued random variable with cumulative distribution function F. Recall that F is non-decreasing (and therefore has at most a finite or countable number of discontinuity points). Moreover, it is right-continuous and has left limits at every point. In general, however, it is not bijective. We can nevertheless introduce the following notion.

Definition 1.2.3. *The generalized inverse function or quantile function F^- of F is defined by*

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

This function coincides with F^{-1} when F is bijective.

Lemma 1.2.4. *For all $x \in \mathbb{R}$ and $u \in [0,1]$, we have the equivalence*

$$u \leq F(x) \quad \Longleftrightarrow \quad F^-(u) \leq x.$$

In particular, $F \circ F^-(u) \geq u$, with equality if and only if $u \in F(\mathbb{R})$.

Example 1.2.5. *Let F be defined as follows:*

$$F(x) = \begin{cases} \frac{x}{3} & \textit{if } x \in [0,1[\\ \frac{x}{3} + \frac{1}{3} & \textit{if } x \in [1,2] \\ 1 & \textit{if } x \geq 2 \\ 0 & \textit{if } x \leq 0 \end{cases}$$

Then:

$$F^{-1}(t) = \begin{cases} 3t & \textit{if } t \in \left[0, \frac{1}{3}\right] \\ 1 & \textit{if } t \in \left[\frac{1}{3}, \frac{2}{3}\right] \\ 3t - 1 & \textit{if } t \in \left[\frac{2}{3}, 1\right] \end{cases}$$

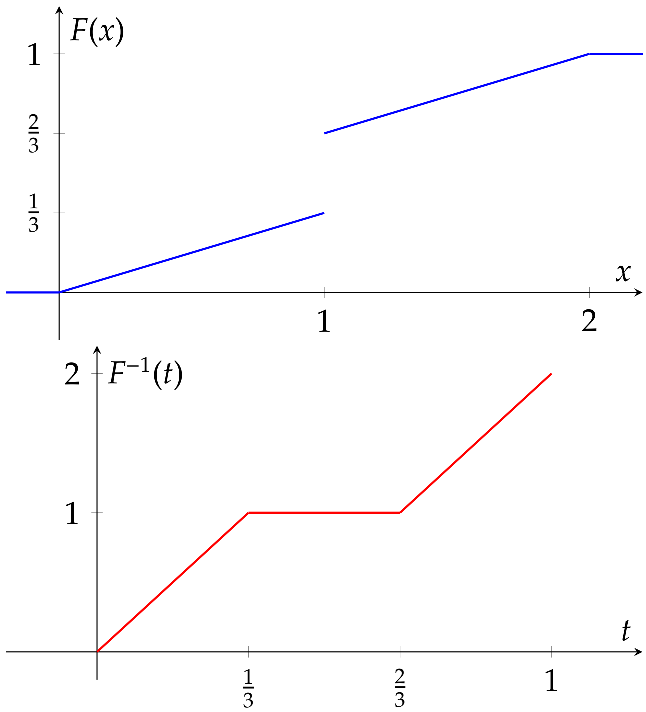

Figure 1.1: Graphs of $F(x)$ and $F^{-1}(t)$

Simulation by inversion of the cumulative distribution function

R program:

```
finv<-function(u)
{
if (u<1/3)
{z=3*u }
else if (u<2/3)
{z=1 }
else
{ z=3*u-1 }
return(z)
}
u=runif(1,0,1) # simulation of a uniform random variable (in [0,1])
# Print the result of applying finv to u
cat("Simulated value of u:", u, "\n" ) # The random uniform value
cat("Simulated value of F^(-1)(u):", finv(u), "\n") # The result
```

Matlab program:

```
u = rand(); % Generate a random uniform variable
fprintf('Simulated value of u: %f\n', u);
% Compute finv
if u < 1/3
z = 3 * u;
elseif u < 2/3
z = 1;
else
z = 3 * u - 1;
end

% Display the result
fprintf('Simulated value of F^(-1)(u): %f\n', z);
```

Proposition 1.2.6 (Inversion Method for Discrete Random Variables). *Let U be a random variable following the uniform distribution on $[0,1]$ and X be a discrete random variable with cumulative distribution function F_X. Let $X(\Omega) = \{x_1, x_2, x_3, \ldots\}$ with $x_1 < x_2 < x_3 < \ldots$. Then for all $k \geq 2$:*

- $p_1 := \mathbb{P}(X = x_1) = \mathbb{P}\left(U \leq F_X(x_1)\right)$
- $p_k := \mathbb{P}(X = x_k) = \mathbb{P}\left(F_X(x_{k-1}) < U \leq F_X(x_k)\right)$

Proof 1. *Since for every k, $F_X(x_k) \in [0,1]$, then from the cumulative distribution function of the uniform distribution we have*

$$\mathbb{P}\left(U \leq F_X(x_1)\right) = F_X(x_1).$$

Since $p_1 = F_X(x_1)$, it follows that

$$p_1 = \mathbb{P}\left(U \leq F_X(x_1)\right).$$

Moreover, for all $k \geq 2$, since

$$\mathbb{P}(X = x_k) = F_X(x_k) - F_X(x_{k-1}),$$

we also obtain

$$\mathbb{P}(X = x_k) = \mathbb{P}\left(U \leq F_X(x_k)\right) - \mathbb{P}\left(U \leq F_X(x_{k-1})\right) = \mathbb{P}\left(F_X(x_{k-1}) < U \leq F_X(x_k)\right).$$

Example 1.2.7. *We want to simulate the random variable X following a Bernoulli distribution with parameter $p \in (0,1)$.*

Solution. *According to the previous proposition, if U is a random variable following the uniform distribution on $[0,1]$, we have*

$$\mathbb{P}(X = 1) = \mathbb{P}\left(F_X(0) < U \leq F_X(1)\right) = \mathbb{P}(1 - p < U \leq 1) = \mathbb{P}(U > 1 - p).$$

The probability of the event $[X = 1]$ *is therefore the same as that of the event* $[U > 1 - p]$. *Thus, we can propose the following simulation:*

Matlab Program

```
% Ask the user to enter the value of p
p = input('Enter the value of p (0 < p < 1): ');

% Simulate the Bernoulli random variable X
if rand(1,1) > 1 - p
X = 1;
else
X = 0;
end

% Display the result
disp(['The value of X is: ', num2str(X)]);
```

R Program

```
p <- as.numeric(readline(prompt = "Enter the value of p (0 < p < 1): "))
if (runif(1,0,1) > 1 - p) {
X <- 1
} else {
X <- 0
}
cat("The value of X is:", X, "\n")
```

Exercise 1.2.8. *Use the inversion method to simulate a random variable X following a geometric distribution with parameter* $p \in (0, 1)$.

Solution

For a geometric distribution, the cumulative distribution function is given by

$$F_X(k) = \sum_{i=1}^{k} q^{i-1}p = p\sum_{i=0}^{k-1} q^i = 1 - (1-p)^k$$

According to the previous proposition, if U is a random variable following the uniform distribution on $[0, 1]$, we have for k:

$$p_k = \mathbb{P}\left(F_X(k-1) < U \leq F_X(k)\right) = \mathbb{P}\left(1 - q^{k-1} < U \leq 1 - q^k\right).$$

We obtain the same formula for $k = 1$,

$$p_1 = \mathbb{P}\left(0 < U \leq F_X(1)\right) = \mathbb{P}\left(1 - q^{1-1} < U \leq 1 - q^1\right).$$

Thus, the event $[X = k]$ occurs if and only if k is the smallest integer such that U is less than or equal to $1 - q^k$. As long as U exceeds $1 - q^k$, we increase k by 1.

We can therefore propose the following pseudo-code:

1. Ask the user to enter the value of p.
2. Initialize $k \leftarrow 1$ and generate $U \sim \mathcal{U}([0,1])$.
3. While $U > 1-(1-p)^k$, increment k.
4. Display the final value of k, which corresponds to X.

In MATLAB, this program can be written as follows:

```
p = input('Enter the value of p (0 < p < 1): ');
k = 1;
u = rand(1, 1);
while u > 1 - (1 - p)^k
    k = k + 1;
end
disp(['The value of X is: ', num2str(k)]);
```

Remark 2 (Another way to simulate the geometric distribution)**.** *We can continue the calculations, which gives for $k \in \mathbb{N}^*$*

$$\begin{aligned}\mathbb{P}(X=k) =& \mathbb{P}\left(1-q^{k-1} < U \leq 1-q^k\right)\\ =& \mathbb{P}\left(q^{k-1} > 1-U \geq q^k\right)\\ =& \mathbb{P}\left((k-1)\ln(q) > \ln(1-U) \geq k\ln(q)\right)\\ =& \mathbb{P}\left(k-1 < \frac{\ln(1-U)}{\ln(q)} \leq k\right) \quad \textit{since } \ln(q) < 0\\ =& \mathbb{P}\left(\left\lfloor \frac{\ln(1-U)}{\ln(q)} \right\rfloor = k-1\right)\\ =& \mathbb{P}\left(\left\lfloor \frac{\ln(1-U)}{\ln(q)} \right\rfloor + 1 = k\right)\end{aligned}$$

We can therefore propose the following simulation:

In Matlab:

```
p = input('Enter the value of p (0 < p < 1): ');
X = floor(log(1 - rand(1, 1)) / log(1 - p)) + 1;
disp(['The value of X is: ', num2str(X)]);
```

In R:

```
p <- as.numeric(readline(prompt="Enter the value of p (0 < p < 1):"))
X <- floor(log(1 - runif(1)) / log(1 - p)) + 1
cat("The value of X is :", X, "\n")
```

Proposition 1.2.9. *If F is continuous and strictly increasing on $\mathbb{R}$ and satisfies*

$$\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to+\infty} F(x) = 1,$$

and if $U \sim \mathcal{U}[0,1]$, then $X = F^{-1}(U)$ has density f.

Proof 2. *Let $a < b \in \mathbb{R}$. Since F is strictly increasing,*

$$\{a < X \leq b\} = \{a < F^{-1}(U) \leq b\} = \{F(a) < U \leq F(b)\}.$$

Therefore

$$\mathbb{P}(a < X \leq b) = \int_0^1 1_{\{F(a)<u\leq F(b)\}} du = F(b) - F(a) = \int_a^b f(y)dy.$$

This means that X has density f.

Example 1.2.10. *The Cauchy distribution with parameter $a > 0$ has density*

$$f(y) = \frac{a}{\pi(y^2+a^2)}, \quad y \in \mathbb{R}.$$

The associated cumulative distribution function is

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(y)dy = \frac{1}{\pi}\int_{-\infty}^x \frac{1}{(y/a)^2+1}\frac{dy}{a} \\ &= \frac{1}{\pi}\int_{-\infty}^{x/a} \frac{1}{1+z^2}dz = \frac{1}{\pi}[\arctan(z)]_{-\infty}^{x/a} = \frac{1}{\pi}\arctan(x/a) + \frac{1}{2}. \end{aligned}$$

To invert the cumulative distribution function, let $u \in (0,1)$ and find x such that $F(x) = u$:

$$F(x) = u \Longleftrightarrow x = a\tan(\pi(u - \tfrac{1}{2})).$$

Consequently, if $U \sim \mathcal{U}[0,1]$,

$$X = a\tan(\pi(U - \tfrac{1}{2})) \sim \mathcal{C}(a).$$

Since $\pi(U - \frac{1}{2}) \sim \mathcal{U}\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and the tangent function is periodic with period π, for any bounded function $f : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned} \mathbb{E}(f(\tan(\pi(U - \tfrac{1}{2})))) &= \frac{1}{\pi}\int_{-\pi/2}^0 f(\tan(\pi + x))dx + \frac{1}{\pi}\int_0^{\pi/2} f(\tan x)dx \\ &= \frac{1}{\pi}\int_0^{\pi} f(\tan x)dx = \mathbb{E}(f(\tan(\pi U))). \end{aligned}$$

Thus

$$\tan(\pi(U - \tfrac{1}{2})) \stackrel{\mathcal{L}}{=} \tan(\pi U)$$

and therefore

$$X = a\tan(\pi U) \sim \mathcal{C}(a).$$

Exercise 1.2.11. *Use the inversion method to simulate a random variable following an exponential distribution with parameter $\lambda > 0$, whose density is given by*

$$\lambda e^{-\lambda z}\mathbf{1}_{\{z>0\}}.$$

Solution

The associated cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} \lambda e^{-\lambda z}\mathbf{1}_{\{z>0\}}dz = 1_{\{x>0\}}\left(1 - e^{-\lambda x}\right).$$

For $u \in (0,1)$, we have

$$F(x) = u \Longleftrightarrow x = -\frac{1}{\lambda}\ln(1-u).$$

We deduce that if $U \sim \mathcal{U}[0,1]$, then

$$-\frac{1}{\lambda}\ln(1-U) \sim \mathcal{E}(\lambda).$$

We can simplify further by noticing that $1 - U \overset{\mathcal{L}}{\equiv} U$ (that is, $1-U$ has the same distribution as U). This implies that

$$-\frac{1}{\lambda}\ln(1-U) \overset{\mathcal{L}}{\equiv} -\frac{1}{\lambda}\ln(U),$$

which allows us to conclude that

$$S = -\frac{1}{\lambda}\ln(U) \sim \mathcal{E}(\lambda).$$

Matlab Program

```
function Y = rexp(m, n, lambda)
    % Generate an m*n matrix of exponential random variables
    Y = -log(rand(m, n)) / lambda;
end
```

R Program

```
rexp_matrix <- function(m, n, lambda) {
  # Generate an m x n matrix of exponential random variables
  Y <- -log(matrix(runif(m * n), nrow = m, ncol = n)) / lambda
  return(Y)
}

result <- rexp_matrix(3, 4, 2)
print(result)
```

1.3 Rejection Method

We want to simulate a random variable with density f (on $\mathbb{D}$), and we assume that there exists a distribution with density g, which is easy to simulate, such that

$$\forall x \in \mathbb{D}, \quad f(x) \leq k g(x)$$

where k is a constant. Let us define $\alpha(x) = \frac{f(x)}{kg(x)}$ on the set $\{x : g(x) > 0\}$.

Proposition 1.3.1. *Let $(X_n, U_n)_{n\geq 1}$ be a sequence of i.i.d. random variables such that for every n, X_n has density g and is independent of U_n, where U_n follows the uniform distribution on* $[0, 1]$.

Let

$$T = \inf\{n : U_n \leq \alpha(X_n)\}$$

(T is a random integer). Let $X = X_T$.

Then the random variable X simulated in this way has density f.

Proof 3. *Note that $\alpha(X_n)$ is defined for almost every ω. Indeed,*

$$\mathbb{P}(X_n \in \{x : g(x) = 0\}) = 0.$$

Let $\varphi \in C_b^+(\mathbb{D})$ (the set of bounded positive continuous functions). We have

$$\varphi(X_T) = \sum_{n=1}^{+\infty} \varphi(X_n) \mathbf{1}_{n=T}$$

(for every ω, this sum contains at most one non–zero term).

$$\begin{aligned}
\mathbb{E}(\varphi(X)) &= \sum_{n=1}^{+\infty} \mathbb{E}(\varphi(X_n) \mathbf{1}_{n=T}) \\
&= \sum_{n=1}^{+\infty} \mathbb{E}\left(\varphi(X_n) \prod_{i=1}^{n-1} \mathbf{1}_{U_i > \alpha(X_i)} \times \mathbf{1}_{U_n \leq \alpha(X_n)}\right) \\
&= \sum_{n=1}^{+\infty} \mathbb{E}\left(\varphi(X_n) \mathbf{1}_{U_n \leq \alpha(X_n)}\right) \prod_{i=1}^{n-1} \mathbb{E}\left(\mathbf{1}_{U_i > \alpha(X_i)}\right).
\end{aligned}$$

The last step follows from the independence of the pairs (X_i, U_i).

Let us compute, for every n,

$$\begin{aligned}
\mathbb{E}\left(\varphi(X_n) \mathbf{1}_{U_n \leq \alpha(X_n)}\right) &= \int_{x\in\mathbb{D}} \left(\int_{u\in[0,1]} \mathbf{1}_{u \leq \alpha(x)} du\right) \varphi(x) g(x) dx \\
&= \int_{x\in\mathbb{D}} \frac{f(x)}{kg(x)} \varphi(x) g(x) dx \\
&= \frac{1}{k} \int_{x\in\mathbb{D}} \varphi(x) f(x) dx.
\end{aligned}$$

We deduce that for every n,

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{U_n>\alpha(X_n)}\right) &= 1-\mathbb{E}\left(\mathbf{1}_{U_n\leq\alpha(X_n)}\right)\\ &= 1-\frac{1}{k}.\end{aligned}$$

Note that the inequality $f \leq kg$ implies

$$\int_{\mathbb{D}} f(x)dx \leq k\int_{\mathbb{D}} g(x)dx,$$

hence $k \geq 1$. Therefore

$$\begin{aligned}E(\varphi(X)) &= \sum_{n=1}^{+\infty}\int_{\mathbb{D}}\varphi(x)f(x)dx\times\frac{1}{k}\left(1-\frac{1}{k}\right)^{n-1}\\ &= \int_{\mathbb{D}}\varphi(x)f(x)dx\times\frac{1}{k}\times\frac{1}{1-\left(1-\frac{1}{k}\right)}\\ &= \int_{\mathbb{D}}\varphi(x)f(x)dx.\end{aligned}$$

Hence the result.

Example 1.3.2. *Let*

$$f(x)=\frac{e^{-x^3}}{Z}\mathbf{1}_{x\geq 1}$$

with

$$Z=\int_1^{\infty}e^{-x^3}dx$$

(the constant Z is unknown, but we will not need it).

Let

$$g(x)=\frac{1}{x^2}\mathbf{1}_{x\geq 1}.$$

The functions f and g are probability densities. It is easy to simulate from the density g using the inversion method for the cumulative distribution function.

We start by studying the function

$$h : x \geq 1 \mapsto x^2e^{-x^3}.$$

We have

$$h'(x) = (2x-3x^4)e^{-x^3} \leq 0$$

for every $x \geq 1$.

Therefore, for every $x \geq 1$,

$$h(x) < h(1) = e^{-1}.$$

We deduce (for $x \geq 1$)

$$f(x) \le \frac{1}{eZ} g(x).$$

Let us set

$$k = \frac{1}{eZ}.$$

Then

$$\alpha(x) = \frac{f(x)}{kg(x)} = x^2 e^{-x^3+1}.$$

Although we do not know Z, we can compute $\alpha(x)$ for every $x \ge 1$. Therefore, we can use the rejection sampling algorithm to simulate from the density f.

R Program

Method 1

```
b <- 0
while (b == 0) {
  u <- runif(1, 0, 1)
  v <- runif(1, 0, 1)
  y <- 1 / (1 - v)
  if (u <= y^2 * exp(-y^3) * exp(1)) {
    b <- 1
  }
}
print(y)
```

Using inversion of the cumulative distribution function:
for $x \ge 1$,

$$G(x) = \int_1^x g(u)du = \left[-\frac{1}{u}\right]_1^x = 1 - \frac{1}{x}$$

for $u \in [0, 1]$

$$G^{-1}(u) = \frac{1}{1-u}.$$

Method 2

```
while (TRUE) {
  u <- runif(1)
  v <- runif(1)
  y <- 1 / (1 - v)
  if (u <= y^2 * exp(-y^3) * exp(1)) {
     break
  }
```

```
}
print(y)
```