Inversion Samplers

STAT221

Probability Distributions

If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the probability distribution of X.

A function can serve as a probability distribution of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions

- $f(x) \geq 0$ for each value within its domain;
- $\sum_x f(x) = 1$, where the summation extends over all the values within its domain.

Cumulative Distribution

If X is a discrete random variable, the function given by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \qquad \text{for} - \infty < x < \infty$$

where $f(t)$ is the value of the probability distribution of X at t, is called the distribution function, or the cumulative distribution of X.

The values $F(x)$ of the distribution function of a discrete random variable X satisfy the conditions

- $F(-\infty) = 0$ and $F(\infty) = 1$;
- if $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b.

Continuous Random Variables

A function with values $f(x)$, defined over the set of all real numbers, is called a probability density function of the continuous random variable X if and only if

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

for any real constants a and b with $a \leq b$.

Analogously to discrete random variables, we can state the following properties:

A function can serve as a probability density of a continuous random variable X if its values, $f(x)$, satisfy the conditions

- $f(x) \geq 0$ for $-\infty < x < \infty$;
- $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

And we can define the distribution function as:
If X is a continuous random variable and the value of its probability density at t is $f(t)$, then the function given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt \qquad \text{for } -\infty < x < \infty$$

is called the cumulative distribution function of X.

If $f(x)$ and $F(x)$ are the values of the probability density and the distribution function of X at x, then

$$P(a \leq X \leq b) = F(b) - F(a)$$

for any real constants a and b with $a \leq b$, and

$$f(x) = \frac{dF(x)}{dx}$$

where the derivative exists.

The cdf is Uniformly Distributed

If X is a continuous random variable with cdf $F_X(x)$, then $U = F_X(X) \sim Uniform(0, 1)$.

Let us first define the inverse transformation

$$F_X^{-1}(u) = \inf\{x : F_X(x) = u\}, \qquad 0 < u < 1.$$

Then we can show that if $U \sim Uniform(0, 1)$, then for all $x \in \mathbb{R}$

$$\begin{aligned} P\left(F_X^{-1}(U) \leq x\right) &= P\left(\inf\{t : F_X(t) = U\} \leq x\right) \\ &= P\left(U \leq F_X(x)\right) \\ &= F_U(F_X(x)) \\ &= F_X(x) \end{aligned}$$

It follows that $F_X^{-1}(U)$ has the same distribution as X.

Therefore, to generate a random observation X:

- Derive the inverse function $F_X^{-1}(u)$
- Generate a random u from $Uniform(0, 1)$
- Obtain $x = F_X^{-1}(u)$

Example: Exponential Distribution

The cdf of the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x} \qquad \text{for } x > 0$$

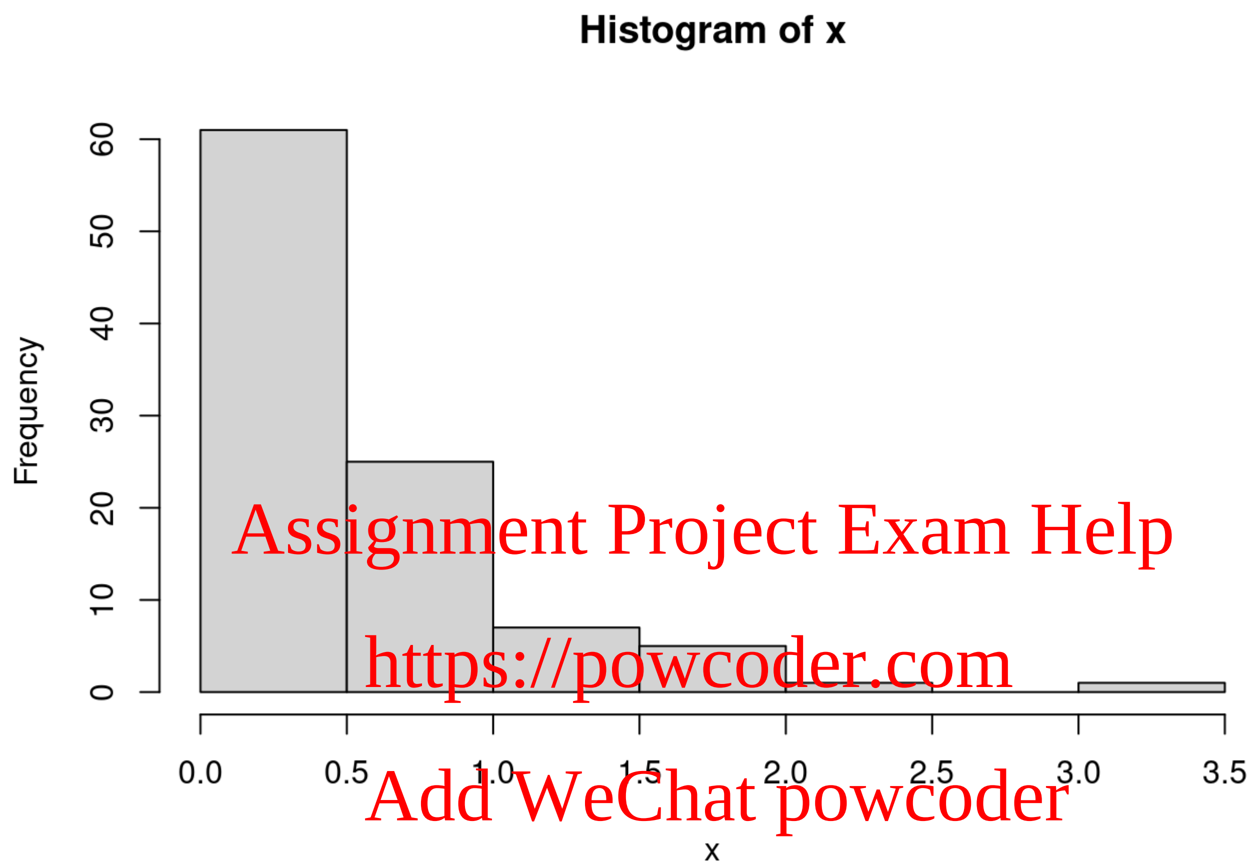
Then the inverse transformation is

$$F_X^{-1}(u) = -\frac{1}{\lambda}\log(1 - u)$$

Because U and $1 - U$ have the same distribution, we can use

$$x = -\frac{1}{\lambda}\log(u)$$

```
# generate a random sample of size 100
# from an exponential distribution with lambda = 2
u <- runif(100, 0, 1)
x <- -log(u)/2
hist(x)
```



Discrete Random Variables

If X is a discrete random variable and

$$\cdots < x_{i-1} < x_i < x_{i+1} < \cdots$$

are the points of discontinuity of $F_X(x)$ the the inverse transformation is $F_X^{-1}(u)$, where

$$F_X(x_{i-1}) < u \leq F_X^{-1}(x_i)$$

Example: Bernoulli Distribution

The probability mass function of the Bernoulli distribution can be written as

$$f_X(x; p) = p^x (1-p)^{1-x} \qquad \text{for } x \in \{0, 1\}$$

Then, $F_X(0) = f_X(x) = 1 - p$ and $F_X(1) = f_X(0) + f_X(1) = 1$.

It follows, that

$$
\begin{aligned}
F_X^{-1}(u) &= 1 \qquad \text{if } u > 1-p \\
F_X^{-1}(u) &= 0 \qquad \text{if } u \le 1-p
\end{aligned}
$$

```
# generate a random sample of size 100
# from a Bernoulli distribution with p = 0.25
u <- runif(100, 0, 1)
x <- u > 0.75
table(x)
```

```
## x
## FALSE  TRUE
##    69    31
```

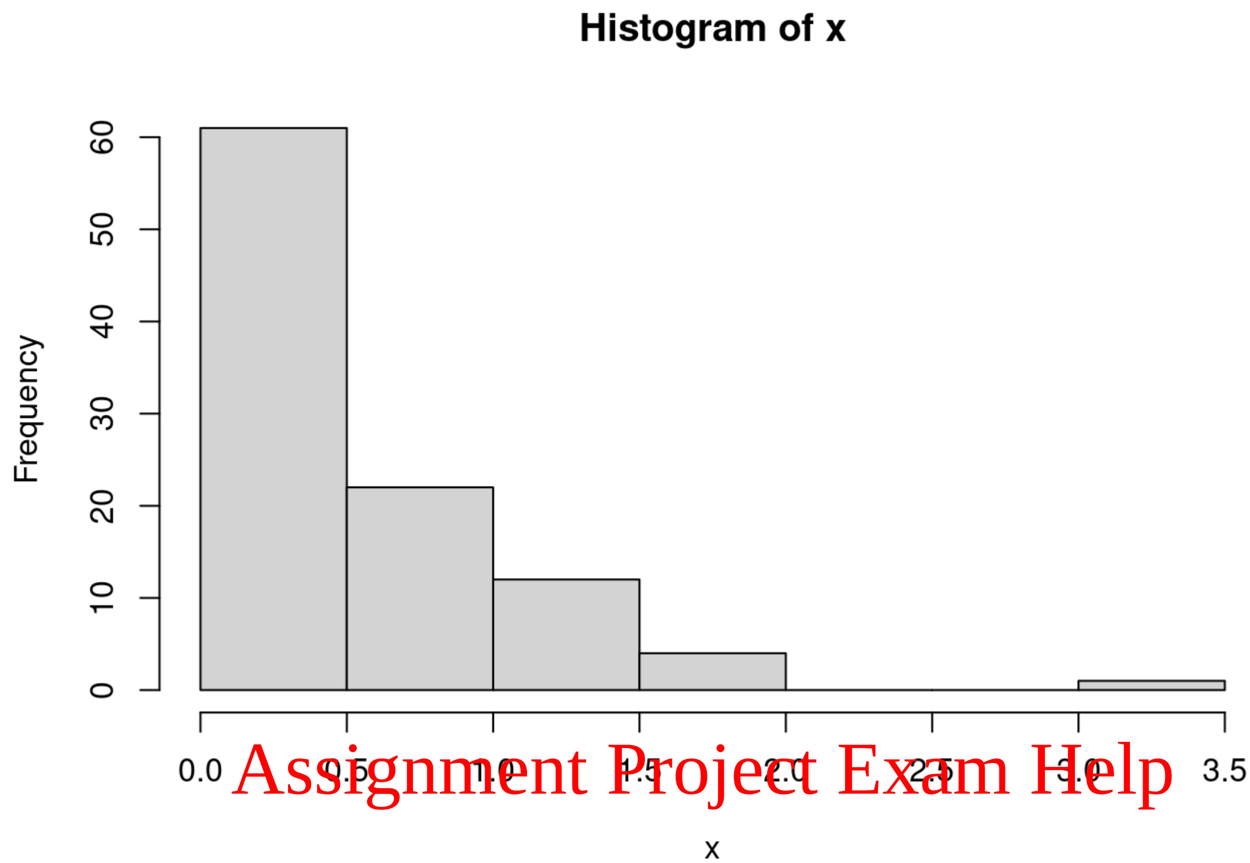
```
mean(x)
```

```
## [1] 0.31
```

Distributions in R

Several function are already available in R to obtain randum values, densities, probabilities, and quantiles from certain distributions.


```
# exponetial distribution
x <- rexp(100, rate=2)
hist(x)
```




```
# Bernoulli distribution
x <- rbinom(100, size=1, prob=0.25)
table(x)
```

```
## x
##  0  1
## 83 17
```

```
mean(x)
```

```
## [1] 0.17
```