Essentials of Data Science With R Software - 1

Probability Theory and Statistical Inference

Univariate Random Variables

:::

Lecture 24

Discrete Random Variables, Probability Mass Function and Cumulative Distribution Function

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Discrete Random Variables:

A random variable *X* is defined to be discrete if its probability space is either finite or countable, i.e. if it takes only a finite or countable number of values.

Note that a set *V* is said to be countable, if its elements can be listed, i.e. there is a one-to-one correspondence between *V* and the positive integers.

Discrete Random Variables:

Example:

- **Let *X* denote the random variable to be the number of defective components in an electronic device.**
- **The device possess a number of characteristics, the random variable *X* summarizes the device only in terms of the number of defects.**
- **The possible values of *X* are integers from zero up to some large value that represents the maximum number of defects that can be found on one of the device.**
- **If this maximum number is very large, we might simply assume that the range of *X* is the set of integers from zero to infinity.**

Probability Distribution:

The probability distribution of a random variable *X* is a description of the probabilities associated with the possible values of *X*.

For a discrete random variable, the distribution is specified by a list of the possible values along with the probability of each.

In some cases, it is convenient to express the probability in terms of a formula.

In practice, a random experiment can often be summarized with a random variable and its distribution. The details of the sample space can often be omitted.

Discrete Random Variables : Example 1

Consider tossing a coin where each trial results in either a head (*H*) or a tail (*T*), each occurring with the same probability 0.5.

The sample space is $\Omega = \{H, T\}$.

Let *X* be a function such that

$$f(x) = \begin{cases} 1 & \text{if outcome is } H \\ 0 & \text{if outcome is } T. \end{cases}$$

Discrete Random Variables : Example 2

When two coins are tossed, observe the outcome

Ω = {ω : ω is *HH*, *HT*, *TH* or *TT*}

Let *X* : number of heads.

$$X(\omega) = \begin{cases} 0 & \text{if } \omega \text{ is } TT \\ 1 & \text{if } \omega \text{ is } TH \text{ or } HT \\ 2 & \text{if } \omega \text{ is } HH. \end{cases}$$

Clearly, the space of *X* is the set (0, 1 , 2).

We can see that *X* is a discrete random variable because its space is finite and can be counted.

Discrete Random Variables : Example 2

Let *X* be a function such that

$$X(\omega) = \begin{cases} 0 & \text{if } \omega \text{ is } TT \\ 1 & \text{if } \omega \text{ is } TH \text{ or } HT \\ 2 & \text{if } \omega \text{ is } HH \end{cases}$$

Let *X* : number of heads.

Clearly, the space of *X* is the set (0, 1 , 2).

We can also assign certain probabilities to each of these values, e.g.

$P(X = 1) = \frac{2}{4} = \frac{1}{2}$

Discrete Random Variables : Example 2

When two coins are tossed, observe the outcome

Ω = {ω : ω is *TT*, *HT*, *TH* or *TT*}

Let C_1 = {ω : ω is *TT*} **C_2 = {ω : ω is *TH*}**

C_3 = {ω : ω is *HT*} **C_4 = {ω : ω is *HH*}**

$C_1,\ C_2,\ C_3,\ C_4 \subset \Omega$

Using independence and equally likely assumptions for events,

$P(C_i) = \frac{1}{4}$ **for each set C_i = 1, 2, 3, 4.**

Then

$P(C_1) = \frac{1}{4},$

$P(C_2 \cup C_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$

$P(C_4) = \frac{1}{4}.$

Discrete Random Variables : Example 2

X : number of heads.

$P(X = 0) = \frac{1}{4}$ because $P(C_1) = \frac{1}{4}$,

$P(X = 1) = \frac{1}{2}$ because $P(C_2 \cup C_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$,

$P(X = 2) = \frac{1}{4}$ because $P(C_4) = \frac{1}{4}$,

The following depicts the distribution of probability over the elements of range of X

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Discrete Random Variables : Example 2

Observe that for each *x*, *P*(*X* = *x*) > 0

$$\sum_{x=0}^{2} P(X = x) = 1$$

$$\sum_{x \le 1} P(X = x) = \frac{1}{4} + \frac{1}{2} \equiv F(1)$$

$$\sum_{x \le 2} P(X = x) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1 \equiv F(2)$$

Discrete Random Variables : Example 3

When the coin is tossed multiple times, observe sequences such as

H, T, H, H, T, H, H, T, T, ...

Let X : number of trials required to get the third head, then $X = 4$ for the given sequence.

Clearly, the space of X is the set (3, 4, 5, . . .).

We can see that X is a discrete random variable because its space is finite and can be counted.

We can also assign certain probabilities to each of these values, e.g. $P(X = 3) = p_1$ and $P(X = 4) = p_2$.

Probability Mass Function (PMF) :

The probability distribution of a random variable *X* is a description of the probabilities associated with the possible values of *X*.

For a discrete random variable, the distribution is specified by a list of the possible values along with the probability of each.

In some cases, it is convenient to express the probability in terms of a formula.

Probability Mass Function (PMF) :

Let X be a discrete random variable which takes k different values.

The probability mass function (PMF) of X is given by

$$p(X) = P(X = x_i) = p_i \text{ for each } i = 1, 2, \ldots, k.$$

It is required that the probabilities p_i satisfy the following conditions:

(1) $0 \leq p_i \leq 1$,

(2) $\sum_{i=1}^{k} p_i = 1$

Cumulative Distribution Function (CDF) of Discrete Variable:

The cumulative distribution function CDF of a discrete random variable as

$$F(X) = \sum_{i=1}^{k} I_{\{xi \leq x\}}\, p_i$$

where I is an indicator function defined as

$$I_{\{xi \leq x\}} = \begin{cases} 1 & \text{if } x_i \leq x \\ 0 & \text{otherwise.} \end{cases}$$

The CDF of a discrete variable is always a step function.

Working with the CDF for Discrete Random :

We can easily calculate various types of probabilities for discrete random variables using the CDF.

Let *a* and *b* be some known constants, then

- $P(X \leq a) = F(a)$
- $P(X < a) = P(X \leq a) - P(X = a) = F(a) - P(X = a)$
- $P(X > a) = 1 - P(X \leq a) = 1 - F(a)$

Working with the CDF for Discrete Random :

- $P(X \geq \mathrm{a}) = 1 - P(X < a) = 1 - F(a) + P(X = a)$

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

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

- $P(a < X < b) = F(b) - F(a) - P(X = b)$

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

Working with the CDF for Discrete Random : Example 1

There are six possible outcomes of rolling a die.

Define X : Number of dots observed on the upper surface of the die,

then the six possible outcomes can be described as

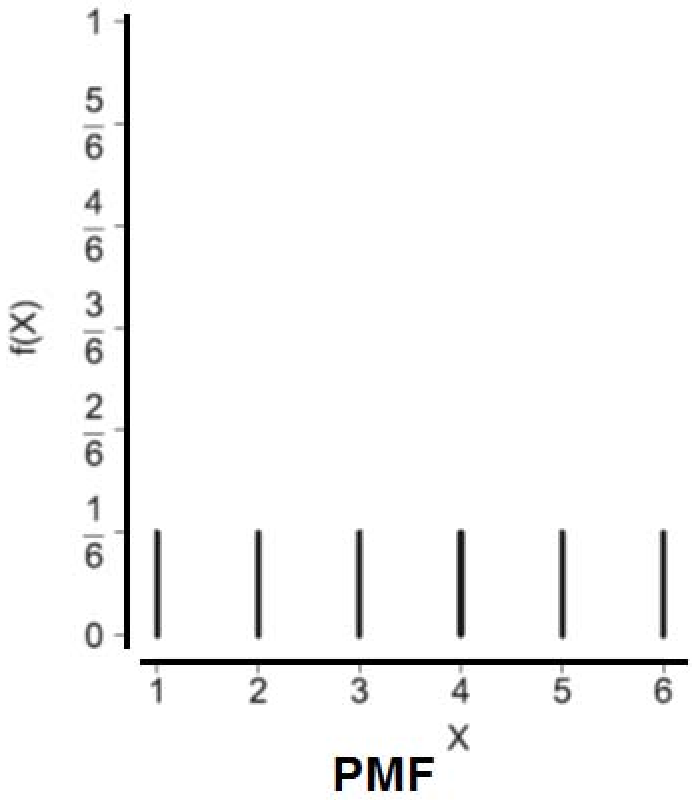
$x_1 = 1, x_2 = 2, . . . , x_6 = 6$ with $P(X = x_i) = 1/6$; $i = 1, 2, . . . , 6$.

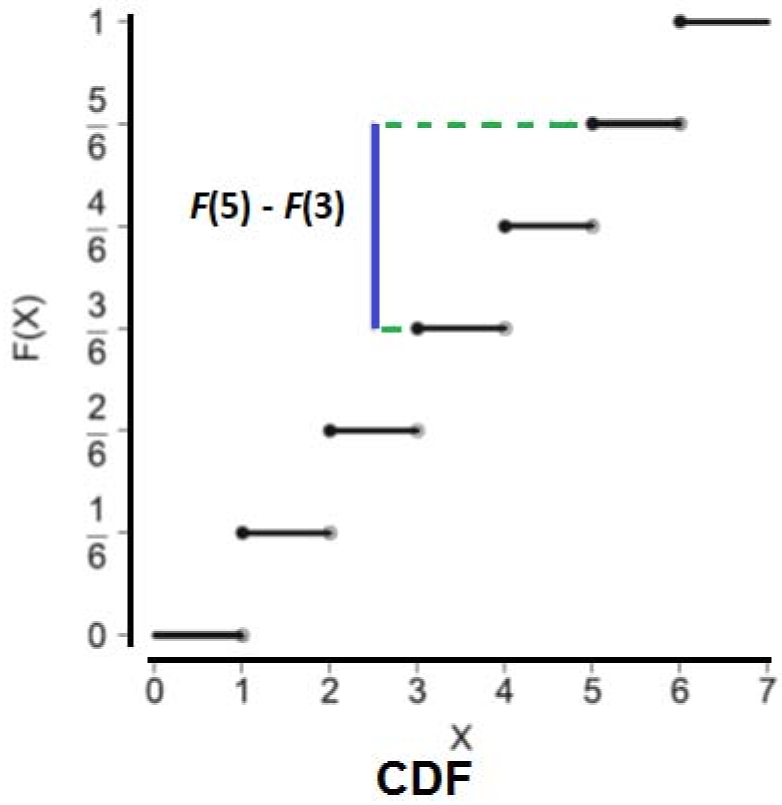
The PMF and CDF are therefore defined as follows:

$$p(x) = \begin{cases} 1/6 & \text{if } x = 1 \\ 1/6 & \text{if } x = 2 \\ 1/6 & \text{if } x = 3 \\ 1/6 & \text{if } x = 4 \\ 1/6 & \text{if } x = 5 \\ 1/6 & \text{if } x = 6 \\ 0 & \text{elsewhere} \end{cases}$$

$$F(x) = \begin{cases} 1/6 & \text{if } -\infty < x < 1 \\ 1/6 & \text{if } 1 \le x < 2 \\ 1/6 & \text{if } 2 \le x < 3 \\ 1/6 & \text{if } 3 \le x < 4 \\ 1/6 & \text{if } 4 \le x < 5 \\ 1/6 & \text{if } 5 \le x < 6 \\ 1/6 & \text{if } 6 \le x < \infty \end{cases}$$

Working with the CDF for Discrete Random : Example 1





$$P(3 < X \leq 5) = F(5) - F(3) = (5/6) - (3/6) = 2/6$$

Working with the CDF for Discrete Random : Example 2

Suppose m and n are the two numbers such that $m = 1, 2, 3$ and

$n = 1, 2$.

$\Omega = \{(m, n) \ : m = 1, 2, 3 \ ; n = 1, 2\}$

Define X : Pair of numbers (m, n) and random variable X as

$X(m, n) = m + n$

$P(\{m, n\}) = \frac{1}{6}$

Clearly, the space of X is the set (2, 3 , 4, 5).

Working with the CDF for Discrete Random : Example 2

The distribution function is

$$F(x) = P(\omega : X(\omega) \leq x)$$

$$= P(\omega : m + n \leq x)$$

$$= \begin{cases} 0 & \textbf{if } x < 2 \quad \textbf{No points in } \Omega \\ 1/6 & \textbf{if } x < 3 \quad \rightarrow \textbf{ One point in } \Omega, \textbf{ i.e., } (1,1) \\ 3/6 & \textbf{if } x < 4 \quad \rightarrow \textbf{ Three points in } \Omega, \textbf{ i.e., } (1,1),(1,2),(2,1) \\ 5/6 & \textbf{if } x < 5 \quad \rightarrow \textbf{ Five points in } \Omega, \textbf{ i.e., } (1,1),(1,2),(2,1),(3,1),(2,2) \\ 1 & \textbf{if } x \geq 5 \quad \rightarrow \textbf{ All points in } \Omega \end{cases}$$

Observe that *F* is a step function increasing only by jumps.

If *F* is a step function, then the corresponding random variable is discrete.

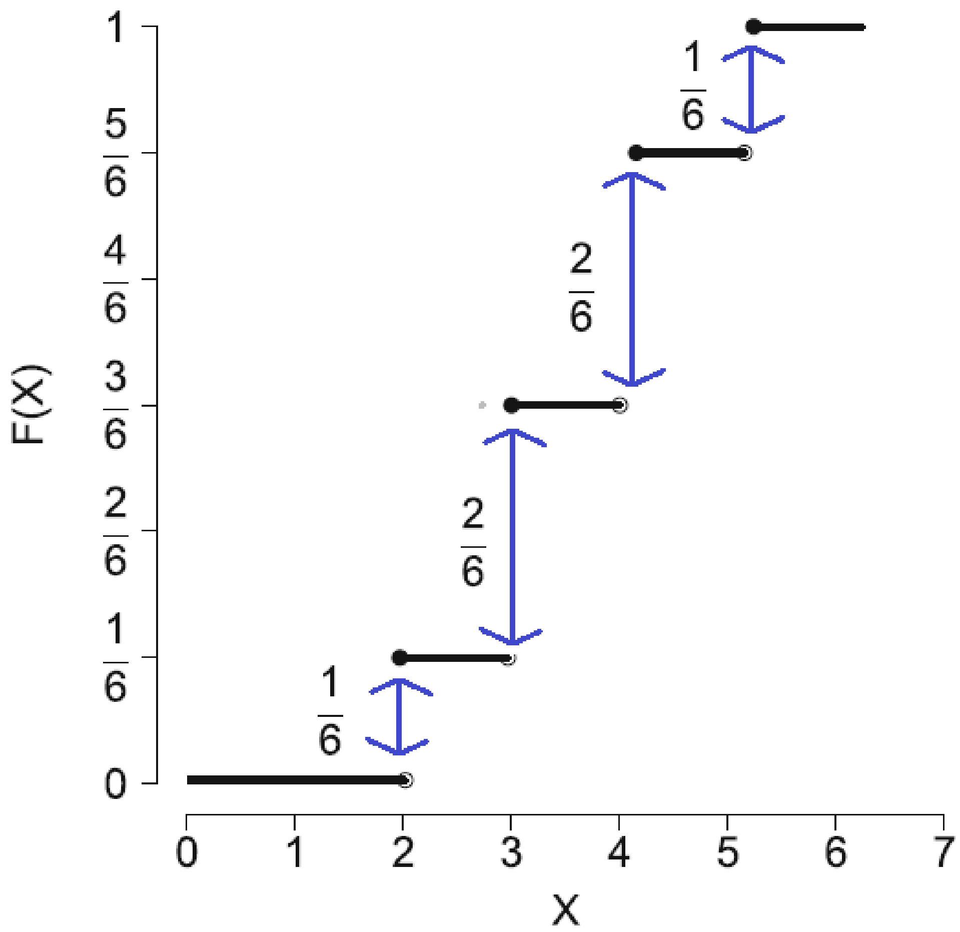
Working with the CDF for Discrete Random : Example 2

The distribution function is

Jumps are at 2, 3, 4, and 5

of sizes 1/6, 2/6, 2/6 and

1/6 respectively.



CDF