IST1990 Probability Theory And Statistics

- **Lecture 4**

- Random Variables
- Discrete Random Variables
- Probability Mass Function for Discrete Random Variables
- Cumulative Distribution Functions of Discrete R.V’s.

Random Variables

A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

Remark on notation: We will use capital letters (e.g. X) to denote random variables. After the random experiment is performed the observed value of the random variable is denoted by a lower case (e.g. x).

Example:

- X is the random variable corresponding to the temperature of the room at time t.
- x is the measured temperature of the room at time t.

Discrete Random Variables

A **discrete random variable** is a random variable with a finite or countable range. For example: number of scratches in the surface of a disk; number of TCP connections to a server; number or errors in a transmission.

Many situations can be adequately modeled using **discrete random variables**:

- Number of failed hard-drives in a RAID system
- Number of failing levees (dijken) in a redundant water protection system
- Number of customers attempting to access a webserver
- Number of times you need to call an airline company until you get someone competent on the other side…

Probability Distributions

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

For a discrete random variable X this is actually quite simple, as X can only take a finite/countable number of possible values.

Definition: Probability *Mass* Function (p.m.f.)

For a discrete random variable X with possible values $x_1, x_2, \ldots x_n$ the **probability mass function** the function with domain $\{x_1, \ldots, x_n\}$ such that

(i) $f(x_i) \geq 0$

(ii) $\sum_{i=1}^{n} f(x_i) = 1$

(iii) $P(X = x_i) = f(x_i)$

Probability Distributions

If the space of possible outcomes is countably infinite, the previous definition can be easily generalized:

Definition: Probability Mass Function (p.m.f.)

For a discrete random variable X with possible values $x_1, x_2, \ldots$ the **probability mass function** is the function $f : \{x_1, x_2, \ldots\} \to [0, 1]$ satisfying

(i) $f(x_i) \geq 0$

(ii) $\sum_{i=1}^{\infty} f(x_i) = 1$

(iii) $P(X = x_i) = f(x_i)$

Example

Consider a situation in which a user can make one of three possible requests to a GUI. We can view this as a random experiment with the following sample space:

$$\text{Sample Space}=\{\text{Print}, \text{Save}, \text{Cancel}\}$$

We can identify each request with a number. For instance identify Print, Save and Cancel with 0,1, and 2, respectively. Let X be the random variable corresponding to the random experiment above.

Let's construct a p.m.f. associated with variable X

$$P(X=0)=0.2, \quad P(X=1)=0.5, \quad P(X=2)=0.3$$

In this case $\{x_1, x_2, x_3\} = \{0, 1, 2\}$ and

$$f(x) = P(X=x) = \begin{cases} 0.2 & \text{if } x=0 \\ 0.5 & \text{if } x=1 \\ 0.3 & \text{if } x=2 \end{cases}$$

It's very easy to check this is a valid probability mass function.

Example

Determine the value of c such that the following functions can serve as a probability distributions of the discrete random variable *X*:

a)

X	-2	-1	2	3	5
$f(x)$	0.1	0.15	c	0.30	0.05

$$0.1 + 0.15 + c + 0.30 + 0.05 = 1 \longrightarrow c = 0.40$$

b) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2, 3$

$$f(0) + f(1) + f(2) + f(3) = 1$$

$$c(0^2 + 4) + c(1^2 + 4) + c(2^2 + 4) + c(3^2 + 4) = 1$$

$$4c + 5c + 8c + 13c = 1 \longrightarrow c = \frac{1}{30}$$

Example

a) Determine the value of k such that the following function can serve as a p.m.f. of discrete random variable X:

$$f(x) = \begin{cases} \dfrac{k}{x^2} & x = 1, 2, 3, 4 \\ 0 & otherwise \end{cases}$$

b) Find the pmf of X.

Solution:

i) $f(xi) \geq 0$ *if* k *is* positive, that is, $k>0$.

ii) The sum of pmf's should be 1: $\Sigma\, f(xi) = 1$

$$\sum_{i=1}^{4} f(x_i) = 1 \Longrightarrow \frac{k}{1^2} + \frac{k}{2^2} + \frac{k}{3^2} + \frac{k}{4^2} = 1$$

$$\frac{k}{1} + \frac{k}{4} + \frac{k}{9} + \frac{k}{16} = 1$$

$$\frac{144k + 36k + 16k + 9k}{144} = 1$$

$$\frac{205k}{144} = 1 \Longrightarrow k = \frac{144}{205} \ .$$

b) The pmf of X is

$$f(x) = \begin{cases} \dfrac{144}{205x^2} & x = 1, 2, 3, 4 \\ 0 & otherwise\ (o.w) \end{cases}$$

$$f(x) = \begin{cases} \dfrac{144}{205} & x=1 \\ \dfrac{36}{205} & x=2 \\ \dfrac{16}{205} & x=3 \\ \dfrac{9}{205} & x=4 \\ 0 & o.w. \end{cases}$$

Example

In a semiconductor manufacturing process, two wafers from a lot are tested. Each wafer is classified as *pass* or *fail*. Assume that the probability that a wafer passes the test is 0.8 and that wafers are *independent*. The sample space for the experiment and associated probabilities are shown in following Table. For example, because of the independence, the probability of the outcome that the first wafer tested passes and the second wafer tested fails, denoted as *pf*, is

$$P(pf) = 0.8(0.2) = 0.16$$

The random variable *X* is defined to be equal to the number of wafers that pass. The
last column of the table shows the values of *X* that are assigned to each outcome in the experiment.

Table Wafer Tests

Outcome			
Wafer 1	Wafer 2	Probability	x
Pass	Pass	0.64	2
Fail	Pass	0.16	1
Pass	Fail	0.16	1
Fail	Fail	0.04	0

Example

Independent trials consisting of the flipping of a coin having probability p of coming up heads are continually performed until a head occurs. If we let X denote the number of times the coin is flipped until a head occurs, then find the probability mass function of X.

Solution:

Let X be the random variable that denotes the number of times the coin is flipped until a head occurs. Then X = 1,2,3,... Then, the probability mass function for X is:

Sample Space	# of trials (X)	Probability
H	1	p
TH	2	(1-p)p
TTH	3	$(1-p)^2 p$
TTTH	4	$(1-p)^3 p$
TTTTH	5	$(1-p)^4 p$
...	...	...
$\underbrace{T\,T\,...\,T}_{n-1}\,H$	n	$(1-p)^{n-1} p$
...	...	...

$$P(X=1)=P(\text{H})=\text{p}$$

$$P(X=2)=P((\text{T,H}))=(1\text{-p})\text{p}$$

$$P(X=3)=P((\text{T,T,H}))=(1\text{-p})^2\text{p}$$

$$\vdots$$

$$P(X=n-1)=P((\underbrace{\text{T,T,...,T}}_{\text{n-2}}\text{,H}))=(1\text{-p})^{\text{n-2}}\text{p}$$

$$P(X=n)=P((\underbrace{\text{T,T,...,T}}_{\text{n-1}}\text{,H}))=(1\text{-p})^{\text{n-1}}\text{p}$$

$$\vdots$$

The probability distribution of the random variable X is given as:

X=x	1	2	3	…	x	…
f(x)=P(X=x)	p	(1-p)p	$(1-p)^2p$		$(1-p)^{x-1}p$	

Then, the pmf of X is:

$f(x)= (1-p)^{x-1}p, \quad x=1,2,\ldots$

Let's show that the sum of this p.m.f is equal to 1:

$$\sum_{x=1}^{\infty} P\{X = x\} = \sum_{x=1}^{\infty} (1-p)^{x-1} p \stackrel{?}{=} 1$$

$$= \sum_{x=1}^{\infty} (1-p)^{x} \frac{p}{(1-p)}$$

$$= \frac{p}{(1-p)} [1-p+(1-p)^2+\ldots]$$

$$= \frac{p}{(1-p)} [(1-p)\{1+(1-p)+(1-p)^2+\ldots\}]$$

$$= \frac{p}{(1-p)} \frac{(1-p)}{1-(1-p)} = 1$$

Example

Let X be the number of heads when a fair coin is tossed five times. Then, find the probability mass function (pmf) of X.

Solution:

S={HHHHH, HHHHT,...,TTTTT} the total number of possible outcomes is $2^5=32$.

X is a discrete random variable taking 5 possible values: 0,1,2,3,4,5

When a fair coin is tossed five times, we can get «the probability of number of x heads is obtained» just by counting the number of outcomes above which have the desired number of heads, and dividing by the total number of possible outcomes, 32. You can easily count the desired number of outcomes by using combinations →→

Number of Head	Total number of Possible Outcomes	Desired number of Outcomes	Probability
0	32	1	1/32
1		5	5/32
2		10	10/32
3		10	10/32
4		5	5/32
5		1	1/32
Total		**32**	**1**

Desired number of outcomes

0 H , 5 T $\quad {}_5C_0 = 1$

1 H , 4 T $\quad {}_5C_1 = 5$

2 H , 3 T $\quad {}_5C_2 = 10$

3 H , 2 T $\quad {}_5C_3 = 10$

4 H , 1 T $\quad {}_5C_4 = 5$

5 H , 0 T $\quad {}_5C_5 = 1$

Then, the pmf of X is:

$$f(x) = \begin{cases} \dfrac{\binom{5}{x}}{2^5} & x = 0,1,2,3,4,5 \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Function

For any random variable the possible outcomes of a random variable are real numbers. It is then often useful to describe the probability distribution using a different function...

Definition: Cumulative Distribution Function (c.d.f.)

The **cumulative distribution function** of a random variable X is denoted by $F(x) : \mathbb{R} \rightarrow [0,1]$, and is given by

$$F(x) = P(X \leq x), \qquad \text{where } x \in \mathbb{R}$$

Cumulative Distribution Function

Properties:

Let X be a discrete random variable X with p.m.f. given by $f(\cdot)$. For any $x \in \mathbb{R}$ we have

$$F(x) = P(X \leq x) = \sum_{i: x_i \leq x} f(x_i)$$

Furthermore $F(x)$ satisfies the following properties

a. F is increasing: if $x \leq y$ then $F(x) \leq F(y)$.
b. $F(x^+) = F(x)$ for $x \in \mathbb{R}$. Thus, F is *continuous from the right.*
c. $F(x^-) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$. Thus, F has *limits from the left.*
d. $F(-\infty) = 0$.
e. $F(\infty) = 1$.

$$F(x^+) = \lim_{t \downarrow x} F(t), \; F(x^-) = \lim_{t \uparrow x} F(t), \; F(\infty) = \lim_{t \to \infty} F(t), \; F(-\infty) = \lim_{t \to -\infty} F(t)$$

The following result shows how the distribution function can be used to compute the probability that X is in an interval →

Suppose that F(x) is the c.d.f. of a real-valued discrete random variable X. If a,b are Real numbers such that a<b, then

a. $\mathbb{P}(X = a) = F(a) - F(a^-)$

b. $\mathbb{P}(a < X \leq b) = F(b) - F(a)$

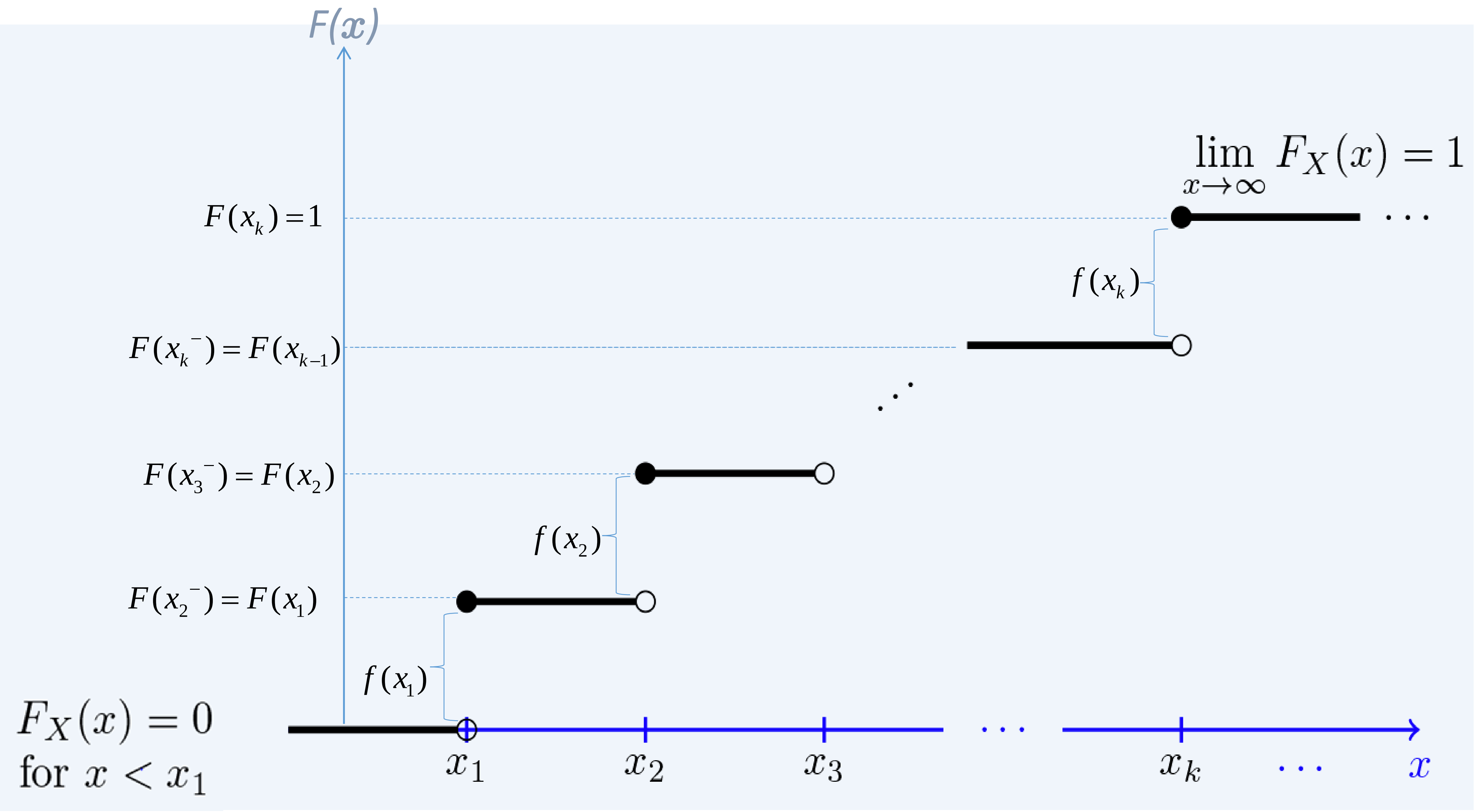
c. $\mathbb{P}(a < X < b) = F(b^-) - F(a)$

d. $\mathbb{P}(a \leq X \leq b) = F(b) - F(a^-)$

e. $\mathbb{P}(a \leq X < b) = F(b^-) - F(a^-)$

Remember that: $F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$ and $F(x^-) = \mathbb{P}(X < x)$ for $x \in \mathbb{R}$.

Note that: c.d.f is defined for any real number, and not only for the values X can take!.



$F(x)$
$\lim_{x\to\infty} F_X(x) = 1$
$F(x_k) = 1$
$f(x_k)$
$F(x_k^-) = F(x_{k-1})$
$F(x_3^-) = F(x_2)$
$f(x_2)$
$F(x_2^-) = F(x_1)$
$f(x_1)$
$F_X(x) = 0$
for $x < x_1$
x_1
x_2
x_3
x_k
x

C.D.F. Example

The probability distribution of a random variable X, is given by

X	1	2	3	4
f(x)	0.11	0.23	0.37	0.29

Construct the cumulative distribution of X; F(X).

$$F(x) = \begin{cases} 0 & for \quad x < 1 \\ f(1) & for \quad 1 \le x < 2 \\ f(1) + f(2) & for \quad 2 \le x < 3 \\ f(1) + f(2) + f(3) & for \quad 3 \le x < 4 \\ f(1) + f(2) + f(3) + f(4) & for \quad 4 \le x \end{cases}$$

$$F(x) = \begin{cases} 0 & for \quad x < 1 \\ 0.11 & for \quad 1 \le x < 2 \\ 0.34 & for \quad 2 \le x < 3 \\ 0.71 & for \quad 3 \le x < 4 \\ 1 & for \quad 4 \le x \end{cases}$$

C.D.F. Example

For the experiment of flipping three coins with the random variable number of heads, find the probability distribution and cumulative distribution function.

Sample space	r.v. (X)
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

Thus, the probability distribution of X *is*

X	0	1	2	3	
f(x)	1/8	3/8	3/8	1/8	= 1

the cumulative distribution function of X *is*

$$F(x) = \begin{cases} 0 & for \quad x < 0 \\ f(0) & for \quad 0 \le x < 1 \\ f(0) + f(1) & for \quad 1 \le x < 2 \\ f(0) + f(1) + f(2) & for \quad 2 \le x < 3 \\ f(0) + f(1) + f(2) + f(3) & for \quad 3 \le x \end{cases}$$

$$F(x) = \begin{cases} 0 & for \quad x < 0 \\ 1/8 & for \quad 0 \le x < 1 \\ 4/8 & for \quad 1 \le x < 2 \\ 7/8 & for \quad 2 \le x < 3 \\ 1 & for \quad 3 \le x \end{cases}$$