Chapter 3. Discrete Random Variables and Probability Distributions

Chapter 3:
Discrete Random Variables and Probability Distributions

- 3.1 Random Variables
- 3.2 Probability Distributions for Discrete Random Variables
- 3. 3 Expected Values of Discrete Random Variables
- 3.4 The Binomial Probability Distribution
- 3.5 Hypergeometric and Negative Binomial Distributions
- 3.6 The Poisson Probability Distribution

Introduction

Why we introduce random variables in this chapter?

As before, we use the letter denote the event,

e.g. When we flip a coin twice, we use the letter "H" and "T" denote the coin's head side and tail sides occurs respectively

So the simple event denote as:

E={HH, HT,TH, TT }

Sometimes ,we may not interest in each simple event, but in some numerical value associate with the event. e.g, we may be interest in the number of heads turn up

Introduction

E.g. in selecting a random sample of students ,we may be interest in **the proportion that are women** rather than which particular students are women.

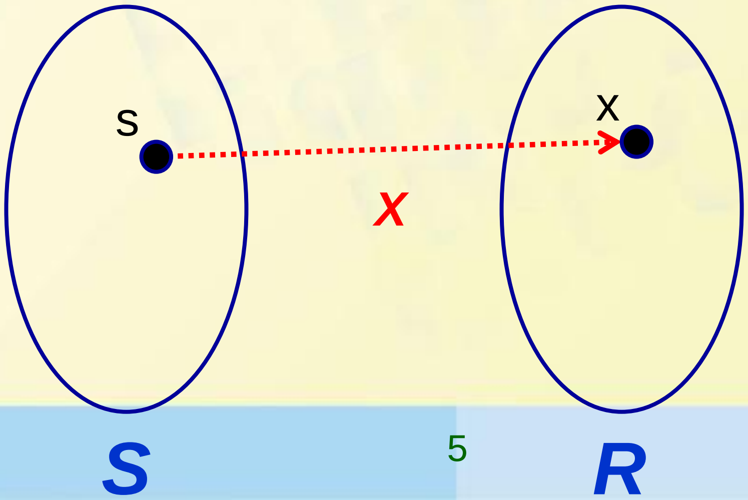
In these example, we have a rule that assigns s to each simple event in S a single real number. Mathematically speaking, we are dealing with a function. Historically, this particular type of function has been called a "random variable"

3.1 Random Variables

- Random Variable (rv)

For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S.

In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real number.



3.1 Random Variables

Random variables are denoted by **uppercase letters**, such as **X** and **Y**.

We use lowercase letters ,such as x, to denote a variable. So here we use lowercase letters to represent some **particular value of the corresponding random variable**

The notation $X(s)=x$ meas that x is the value associated with the outcome s by the rv **X**.

3.1 Random Variables

- **Two Types of Random Variables**

➢ **Discrete Random Variable (Chap. 3)**

A discrete random variable is an rv whose possible values either **constitute a finite set** or else can be **listed in an infinite sequence** in which there is a first element, a second element, and so on.

3.1 Random Variables

➢Continuous Random Variable (Chap. 4)

A random variable is continuous if its set of possible values consists of an entire interval on the number line.

Note: there is no way to create an infinite listing them!

2.1 Sample Spaces and Events

- Example 3.3 (Ex. 2.3 Cont')
- Example 2.3

Two gas stations are located at a certain intersection. **Each one has six gas pumps**. Consider the experiment in which **the number of pumps in use** at a particular time of day is determined for each of the stations. The possible outcomes:

	0	1	2	3	4	5	6
0	(0 0)	(0 1)	(0 2)	(0 3)	(0 4)	(0 5)	(0 6)
1	(1 0)	(1 1)	(1 2)	(1 3)	(1 4)	(1 5)	(1 6)
2	(2 0)	(2 1)	(2 2)	(2 3)	(2 4)	(2 5)	(2 6)
3	(3 0)	(3 1)	(3 2)	(3 3)	(3 4)	(3 5)	(3 6)
4	(4 0)	(4 1)	(4 2)	(4 3)	(4 4)	(4 5)	(4 6)
5	(5 0)	(5 1)	(5 2)	(5 3)	(5 4)	(5 5)	(5 6)
6	(6 0)	(6 1)	(6 2)	(6 3)	(6 4)	(6 5)	(6 6)

Example 3.3 (Ex. 2.3 Cont')

Now we define rv's X,Y, and U by

X = the total number of pumps in use at the two stations

Y = the difference between the number of pumps in use at station 1 and the number in use at station 2

U = the maximum of the numbers of pumps in use at the two stations.

When s=(2,3) then

$$X(2,3)=2+3=5$$

$$Y(2,3)=2-3=-1$$

$$U(2,3)=\max(2,3)=3$$

Note: X, Y, and U have finite values.

Example 3.4 (Ex. 2.4 Cont')

Example 2.4

If a new type-D flashlight battery has a voltage that is outside certain limits, that battery is characterized as a failure(F); otherwise, it is a success(S).

Suppose an experiment consists of testing each battery as it comes off an assembly line **until we first observe a success**. The sample space is

$$S = \{S, FS, FFS, FFFS, \ldots\}$$

which contains an **infinite number** of possible outcomes.

Example 3.4 (Ex. 2.4 Cont')

Now we define an rv X by

X = the number of batteries examined before the experiment terminates

Then X(S)=1, X(FFS)=2, X(FFFS)=3,….,X(FFFFFFS)=7, and so on.

Note: X can be listed in an infinite sequence

3.1 Random Variables

- **Example 3.5**

  suppose that in some random fashion, a location (latitude and longitude) in the continental United States is selected. Define an rv Y by

  Y= the height above sea level at the selected location

Note: Y contains all values in the range [A, B]

A: the smallest possible value

B: the largest possible value

- **Bernoulli Random Variable**

Any random variable whose only possible are 0 and 1 is called Bernoulli random variable.

Bernoulli Random Variable

- **Example 3.1**

When a student attempts to log on to a computer time-sharing system, either all ports are busy (F), in which case the student will fail to obtain access, or else there is at least one port free (S), in which case the student will be successful in accessing the system.

With $\boldsymbol{S}$={S,F}, define an rv X by

$$X(S) = 1, X(F) = 0$$

The rv X indicates whether (1) or not (0) the student can log on.

Bernoulli Random Variable

- **Example 3.2**

Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer, and define an rv Y by

Y(n)=1 if the selected number *n* is unlisted

Y(n)=0 otherwise

e.g.

if 5282966 **appears** in the telephone directory, then

Y(5282966) = 0, whereas Y(7727350)=1 tells us that the number 7727350 is unlisted.

3.2 Probability Distributions for Discrete Random Variables

- **Probability Distribution**

The probability distribution or probability mass function (pmf) of a discrete rv is defined for every number x by

$$p(x) = P(X=x) = P(\text{all } s \text{ in } S: X(s)=x)$$

Note:

$$p_i = p(x_i) \geq 0, \sum_{i=1}^{n} p_i = 1$$

x	x_1	x_2	…	X_n
$p(x)$	p_1	p_2	…	p_n

Probability Distribution

- **Example 3.8**

Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

One of these lots is to be randomly selected for shipment to a particular customer.

Let X=the number of defectives in the selected lot.

X	0 (lot 1,3 or 6)	1 (lot 4)	2 (lot 2 or 5)
Probability	0.5	0.167	0.333

Probability Distribution

- **Example 3.10**

Consider a group of five potential blood donors—A,B,C,D, and E—of whom only A and B have type O+ blood. Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified.

Let the rv Y= the number of typings necessary to identify an O+ individual.

Then the pmf of Y is

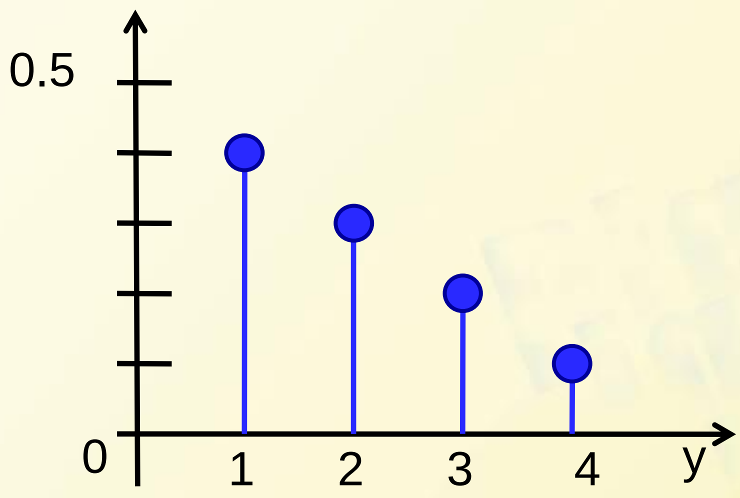
y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1

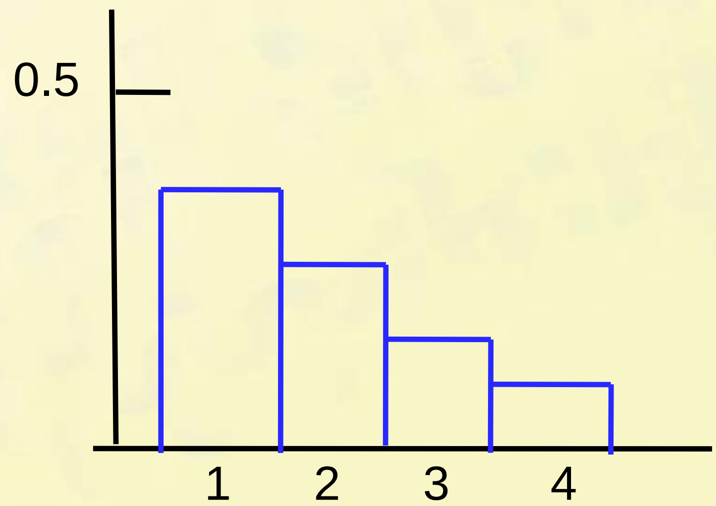
Where any y value not listed relieves zero probability

Probability Distribution

- Line Graph and Probability Histogram

y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1





Example 3.9

Suppose we go to a university bookstore during the first week of classes and observe whether the next person buying a computer buys a laptop or a desktop model.

$$X = \begin{cases} 1 & \text{If the customer purchase a laptop computer} \\ 0 & \text{If the customer purchase a desktop computer} \end{cases}$$

If 20% of all purchasers during that week select a laptop, the pmf for X is

$$p(x) = \begin{cases} 0.8, & if \quad x = 0 \\ 0.2, & if \quad x = 1 \\ 0, & if \quad x \neq 0 \; or \; 1 \end{cases}$$

x	0	1
$p(x)$	0.8	0.2

A **Parameter** of a Probability Distribution

Suppose $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution.

The collection of all probability distributions for different values of the parameter is called **a family of probability distributions.**

***e.g.* Ex. 3.9**

x	0	1
$p(x)$	0.8	0.2

⇨

x	0	1
$p(x)$	$1-\alpha$	α

A Parameter of a Probability Distribution

- **Example 3.12**

Starting at a fixed time, we observe the gender of each newborn child at a certain hospital until a boy (B) is born. Let p=P(B), assume that successive births are independent, and define the rv X by **X=number of births observed**. Then

$p(1) = P(X=1) = P(B) = p$

$p(2) = P(X=2) = P(GB) = P(G)\,P(B) = (1-p)p$

$p(3) = P(X=3) = P(GCB) = P(G)P(G)\,P(B) = (1-p)^2p$

…

$\mathbf{p(k) = P(X=k) = P(G\ldots GB) = (1-p)^{k-1}p}$

Cumulative Distribution Function

The cumulative distribution function (cdf) F(x) of a discrete rv variable X with pmf p(x) is defined for every number x by

$$F(x) = P(X \le x) = \sum_{y:y \le x} p(y)$$

For any number x, F(x) is the probability that the observed value of X will be at most x.

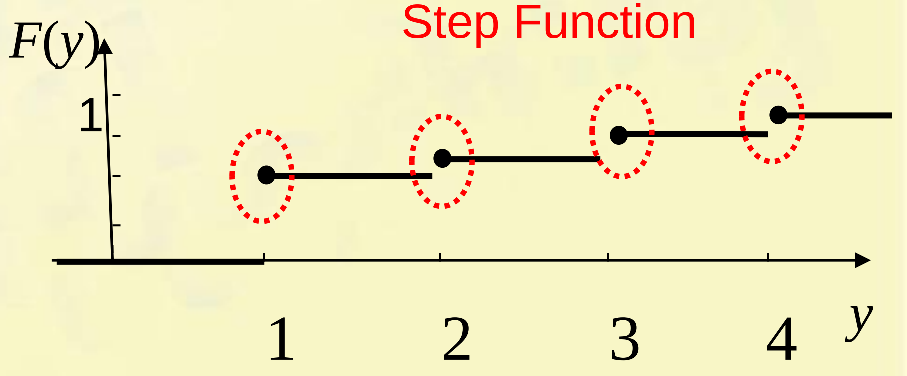
- **Example (Ex. 3.10 continued)**

The pmf of Y (the number of blood typings) in Example 3.10 was

y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1

Then the corresponding cdf is

$$F(y) = \begin{cases} 0, & if \ \ y < 1 \\ 0.4 & if \ 1 \le y < 2 \\ 0.7 & if \ 2 \le y < 3 \\ 0.9 & if \ 3 \le y < 4 \\ 1, & if \ 4 \le y \end{cases}$$



- **Example 3.14 (Ex. 3.12 Cont')**

$$p(x) = \begin{cases} (1-p)^{x-1} p, & x = 1,2,3,\dots \\ 0 & otherwise \end{cases}$$

For a positive integer x,

$$F(x) = \sum_{y \le x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1} p = p \sum_{y=1}^{x} (1-p)^{y-1}$$

$$F(x) = p \cdot \frac{1-(1-p)^x}{1-(1-p)} = 1-(1-p)^x$$

For any real value x,

$$F(x) = 1-(1-p)^{\lfloor x \rfloor}$$ $\lfloor x \rfloor$ is the largest integer ≤ x

- **Proposition**

For any two numbers a and b with $a \le b$,

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

where "a-" represents the largest possible X value that is strictly less than a.

In particular, if the only possible values are integers and if a and b are integers, then

$$\begin{aligned} P(a \le X \le b) &= P(X = a \text{ or } a+1 \text{ or } \dots \text{ or } b) \\ &= F(b) - F(a-1) \end{aligned}$$

Taking a=b yields $P(X = a) = F(a) - F(a-1)$ in this case.

- **Example 3.15**

Let X= the number of days of sick leave taken by a randomly selected employee of a large company during a particular year.

If the maximum number of allowable sick days per year is 14, possible values of X are 0, 1, …, 14. With $F(0)=0.58$, $F(1)=0.72$, $F(2)=0.76$, $F(3)=0.81$, $F(4)=0.88$, and $F(5)=0.94$. Then :

$P(2 \le X \le 5) = ?$

$P(X=3)=?$

Solution:

$P(2 \leq X \leq 5) = P(X=2,3,4 \text{ or } 5) = F(5) - F(1) = 0.22$

$P(X=3) = F(3) - F(2) = 0.05$

- **Three Properties of cdf (discrete/continuous cases)**

1. Non-decreasing, ***i.e.*** if $x_1 < x_2$ then $F(x_1) \le F(x_2)$

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

3. $F(x+0) = F(x)$

Note: Any function that satisfies the above properties would be a cdf.