

4 Discrete Probability Distributions (P.41)

Last week we have learned that probability (or chance) plays an important role in many real world problems. For example, a health department wishes to know the probability of contacting a particular disease. It is therefore important to study how the probabilities are associated with experimental data. Now we devote our attention to study probability distributions. There are two main types of *random variables* and *probability distributions*. They are

- discrete random variables/discrete probability distributions
- continuous random variables/continuous probability distributions

4.1 Random variables

Consider the experiment of tossing a fair coin three times. Let X be the number of times the heads come up. Clearly, X can take one of the values 0, 1, 2, or 3, and no one can tell the value of X before they see the outcome. A variable of this type is called a *random variable*, since the values of X are *uncertain*.

Note: A random variable is a variable that assumes *numerical values* associated with random outcomes of an experiment.

4.2 Discrete random variables

Definition: A random variable (rv), X , which may take on only a *countable number of distinct values*, such as $0, 1, 2, 3, 4, \dots$, is said to be a *discrete random variable*. In general, the set of possible values of a discrete random variable X is denoted by S and can be written as $S = \{x_1, \dots, x_n, \dots\}$.

Examples: The following are discrete random variables:

- number of children per family;
- attendance of MATH1015 lectures;
- number of patients admitted to an ER each day, etc..

Note: Random variables which consist of measurements are usually not discrete. For example, height of students in a tutorial class. They are called *continuous* random variables and can take on any real value in a given interval, rather than being restricted to integers. Further examples include weight of babies, the amount of sugar in a blood sample, etc.

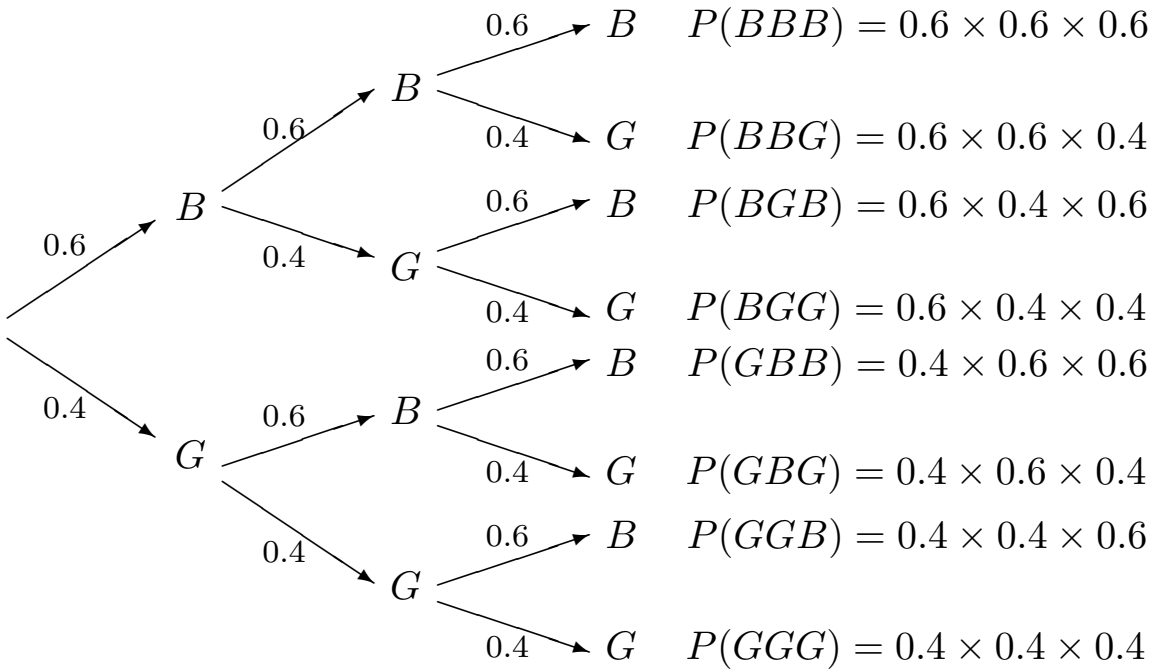
This type continuous random variables will be studied in the next chapter of this course.

4.3 Discrete Probability Distributions

Suppose that the random variable X denotes the number of boys in all families with three children. Assume that the probability of a boy is 0.6 and of a girl is 0.4. Look at the probability distribution of such children:



Tree diagram for the gender of three children



Define $X = \{\text{number of boys}\}$.

Since X can take only whole numbers as values, it is a discrete random variable. Tabulate the probabilities associated with each value of X .

Solution: From the above probability tree we have:

- $P(X = 0) =$ _____
- $P(X = 1) =$ _____
- $P(X = 2) =$ _____
- $P(X = 3) =$ _____

Now summarise the above probability distribution in the table below:



Value of X	0	1	2	3
Probability				

NOTE: The above table represents the values of X (outcomes) and its associated probabilities. This is called the *discrete probability distribution of X* , or the *probability mass function of X* .

Example: Find $\sum_{r=0}^3 P(X = r)$, the sum of all probabilities.

Solution: $\sum_{r=0}^3 P(X = r) =$ _____

Remark: It is easy to see that the sum of all probabilities is 1. This is true for any discrete probability distribution.

Notation: Let $p(r) = P(X = r)$ denotes the probability that X is taking the value r .

Example: In the above example, for $r = 3$, we have $p(3) = P(X = 3) = 0.216$.

Example: Find $P(X < 1)$, $P(X \leq 2)$, and $P(X \geq 2)$ for the above example.

Solution:

$$P(X < 1) = P(X = 0) = \underline{\hspace{2cm}}$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) \stackrel{\text{or}}{=} 1 - P(X = 3)$$

$$= \underline{\hspace{2cm}} \stackrel{\text{or}}{=} \underline{\hspace{2cm}}$$

$$P(X \geq 2) = P(X = 2) + P(X = 3) = \underline{\hspace{2cm}}$$



Exercise 1: A general practitioner doctor is interested in knowing how many years her patients stay with her. Let $X =$ the number of years a patient will stay with the doctor. Over the years, she has established the following probability distribution:

r	1	2	3	4	5	6	7
$P(X = r)$	0.1	0.05	0.1	?	0.3	0.2	0.1

1. Find $P(X = 4)$.

Ans: $P(X = 4) = 1 - 0.1 - 0.05 - 0.1 - 0.3 - 0.2 - 0.1 = 0.15$.

2. Find $P(X < 4)$.

Ans: $P(X < 4) = 0.1 + 0.05 + 0.1 = 0.25$.

Exercise 2: Read example P.42 of the textbook.

Example: Suppose that Peter plans to survey a sample of 100 families with 3 children. How many families do you expect in this sample having (i) two boys (ii) no boys?

Solution:

(i) The expected number of families with two children is _____ or about ____.

(ii) The expected number of families with no children is _____ or about ____.

4.4 Expected value of a discrete random variable

The long run average of a random variable is called the expected value. This represents the population (or true) mean, denoted μ . This population mean value, or the expected value of a random variable X is also denoted by $E(X)$.

Definition: When the random variable is discrete, its *expected value* or *mean* is given by the following sum over the all possible values of r :

$$\mu = E(X) = \sum_r rP(X = r).$$

Note: The sample mean, $\bar{X} = \frac{1}{n} \sum_i x_i$ is an estimate of the population mean, μ using the sample information $\{x_1, \dots, x_n\}$.

Example: Find the expected value of X , $\mu = E(X)$ for the random variable defined in the example with the family with 3 children. Recall that r takes values 0, 1, 2, and 3 with respective probabilities.

Solution:

$$E(X) = \sum rP(X = r)$$

= _____

= _____

= _____

Interpretation: Assume the probabilities for boys and girls are as described for all families. Then, there are, on average, 1.8 boys per family with three children.

Exercises: (1) Read Book P.42.
(2) Find $E(X)$ for Exercise 1 above.

Note: In general, for any number k , we can define $E(X^k)$ as follows:

$$E(X^k) = \sum_r r^k P(X = r).$$

Example: Find the expected value of X^2 , $E(X^2)$ for the random variable defined in the example with the family with 3 children.

Solution:

$$\begin{aligned} E(X^2) &= \sum_r r^2 P(X = r) \\ &= \underline{\hspace{10cm}} \\ &= \underline{\hspace{10cm}} \\ &= \underline{\hspace{10cm}} \end{aligned}$$

Example: Verify that $E(X^2) \neq [E(X)]^2$.

Solution: $[E(X)]^2 = \underline{\hspace{2cm}} \neq \underline{\hspace{2cm}} = E(X^2)$



4.5 Variance of a discrete random variable (p. 33)

In this section, we look at the concept of the population variance. This is denoted by $\text{Var}(X)$ or σ^2 :

Definition: The population *variance*, σ^2 , of X is given by

$$\text{Var}(X) = \sigma^2 = \text{E} \left[\left(X - E(X) \right)^2 \right] = \text{E} (X^2) - [E(X)]^2.$$

Therefore, the population standard deviation, σ , is just the square root of $\text{Var}(X)$, or $\sigma = \sqrt{\text{Var}(X)}$.

Note: Recall that s^2 is the *sample* variance, calculated from n observations by the formula

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right]$$

We say that s^2 is an estimator of σ^2 .

Example: Find the variance of number of boys in the example with the family with 3 children.

Solution: Recall that from

Value of X	0	1	2	3
Probability	0.064	0.288	0.432	0.216

we have $E(X) = 1.8$, and $E(X^2) = 3.96$. Therefore,

$$\begin{aligned} \sigma^2 &= \text{Var}(X) = \text{E} (X^2) - [E(X)]^2 \\ &= \underline{\hspace{10em}} \end{aligned}$$

Example: The standard deviation in the above example is

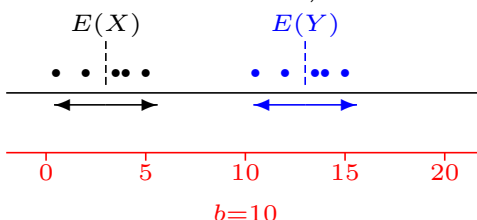
$$\sigma = \underline{\hspace{2cm}}$$

Exercise: Let $Y = 2X$. Using the above probability distribution, verify that $E(Y) = 2 \times 1.8 = 3.6$ and $Var(Y) = 4 \times 0.72 = 2.88$.

Remember:

- If $Y = X + b$, a shift of X , where b is a fixed constant, then

$$E(Y) = E(X + b) = E(X) + b$$

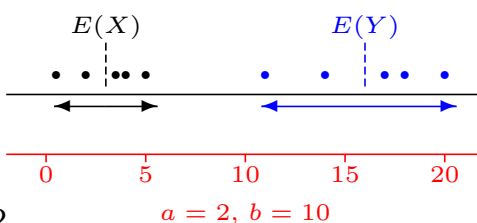
$$Var(Y) = Var(X + b) = Var(X)$$


Note that the variance is unchanged by a shift of location by b since

$$Var(Y) = E[(Y - E(Y))^2] = E[(X + b - E(X) - b)^2] = Var(X).$$

- If $Y = aX + b$, a linear function of X , where a and b are fixed constants, then

$$E(Y) = E(aX + b) = aE(X) + b$$

$$Var(Y) = Var(aX + b) = a^2Var(X)$$


Note that the variance is scaled by a^2 .

Example: The probability distribution of a random variable X is given below. Find

1. $E(X)$; 2. $Var(X)$; 3. $E(2X - 3)$; 4. $Var(2X - 3)$.

r	-2	-1	0	1	2
$p(r)$	0.2	0.1	0.3	0.3	0.1



Solution:

1. $E(X) = \sum_r r p(r)$

= _____

2. $E(X^2) = \sum_r r^2 p(r)$

= _____

$\text{Var}(X) = E(X^2) - [E(X)]^2 =$ _____

3. $E(2X - 3) = 2E(X) - 3 =$ _____

4. $\text{Var}(2X - 3) = 2^2 \text{Var}(X) =$ _____

Exercise: Let X be the number of adults having hypertension in a family. Suppose that the probability distribution of X is given by

Value of X, r	0	1	2
Probability, $P(X = r)$	0.72	0.26	0.02

Find $E(X)$, $E(X^2)$, and $\text{Var}(X)$. What is the standard deviation of X ?

Solution:

$E(X) = \sum_r r p(r) =$ _____

$E(X^2) = \sum_r r^2 p(r) =$ _____

$\text{Var}(X) = E(X^2) - [E(X)]^2 =$ _____

$\sigma =$ _____

Exercise: If two carriers of the gene for albinism marry, each of their children has probability of being albino is 0.25. It is known that the probabilities of any two children of such a couple of being

albino is independent. Suppose that a couple carrying the gene for albinism has five children and let X be the number of children being albino in this family. Find the probability distribution of X .

Solution:

Clearly X is a discrete valued RV as it can take one of the values from 0, 1, 2, 3, 4 or 5.

- $P(X = 0) = P(\text{no albino children among 5})$
= _____
- $P(X = 5) = P(\text{all 5 are albino children}) =$ _____
- $P(X = 1) = P(\text{only 1 is an albino child}) = ?$

Now we need to find the number of ways that one child can be albino. It could be either the first, second, third, fourth or the fifth child. That is there 5 possibilities and thus

$$P(X = 1) = P(\text{only 1 is an albino child})$$
$$= \underline{\hspace{2cm}}$$

- $P(X = 2) = P(\text{two of them are albino among 5}) = ?$

Now we need to find the number of ways that two of them can be albino. It could be any two of the five. That is: 1 and 2 or 1 and 3 or 1 and 4 or 1 and 5 or 2 and 3 or 2 and 4 or 2 and 5 or 3 and 4 or 3 and 5 or 4 and 5. There 10 possibilities and thus



$$P(X = 2) = P(\text{two of them are albino among 5})$$

$$= \underline{\hspace{10em}}$$

Similarly,

- $P(X = 3) = P(\text{three of them are albino among 5})$
 $= \underline{\hspace{10em}}$ and

- $P(X = 4) = P(\text{four of them are albino among 5})$
 $= \underline{\hspace{10em}}$

We summarize the above information in a tabular form:

r	0	1	2	3	4	5
$P(X = r)$	0.237	0.396	0.264	0.088	0.015	0.001

Exercise:

Find $E(X)$ and $\text{Var}(X)$ from the above distribution table.

Solution:

- $E(X) = \underline{\hspace{10em}}$

Ans 1.253

- $E(X^2) = \underline{\hspace{10em}}$

$\text{Var}(X) = \underline{\hspace{10em}}$

Ans 0.939

The above arguments will be considered to develop a very popular discrete distribution called **The Binomial Distribution**. Read your notes before the next lecture.