Chapter 7

Random variables and discrete probability distributions

Random variables and probability distributions

A **random variable** is a function or a rule that assigns a numerical value to each outcome of an experiment.

Alternatively, the ***value*** of a random variable is a numerical event.

There are two types of random variables:

- **discrete random variables**
- **continuous random variables**

Two Types of Random Variables…

Discrete random variable

One that takes on a ***countable*** number of values

E.g. sum of values on the roll of two dice: 2, 3,…, 12.

Continuous random variable

One whose values are ***not discrete***, not countable

E.g. time (30 mins, …, 30.01 mins, …, 30.02 mins, …)

Analogy:

Integers are discrete, while real numbers are continuous.

Discrete and Continuous Random Variables

A random variable is ***discrete*** if it can assume only a countable number of values.

A random variable is ***continuous*** if it can assume an uncountable number of values.

Discrete random variable

After the first value is defined the second value, and any value thereafter, are known.

Therefore, the number of values is countable.

Continuous random variable

After the first value is defined, any number can be the next one.

Therefore, the number of values is uncountable.

Probability Distributions

A ***probability distribution*** is a table, formula, or graph that describes the values of a random variable and the probability associated with these values.

Since we're describing a **random variable** (which can be discrete or continuous) we have two types of probability distributions:

- discrete probability distributions (this chapter) and
- continuous probability distributions (Chapter 8).

Probability Notation

An upper-case letter will represent the ***name*** of the random variable, usually **X**.

Its lower-case counterpart will represent the ***value*** of the random variable.

The probability that the random variable **X** will equal x is:

$P(\mathbf{X} = x)$, or more simply $p(x)$

Discrete probability distributions

A table, formula or graph that lists all possible values a discrete random variable can assume, together with their associated probabilities, is called *a discrete probability distribution*.

To calculate P(**X** = x), the probability that the random variable **X** assumes the value x, add the probabilities of all the simple events for which **X** is equal to x.

Example 1

Find the probability distribution of the random variable describing the number of heads that turn up when a coin is flipped twice.

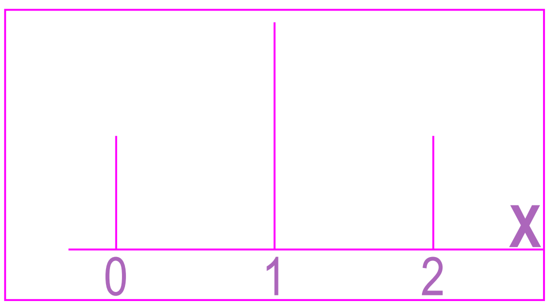
Example 1: Solution

Possible outcomes: {HH, HT, TH, TT}

X = Number of heads = {0, 1, 2}

Simple event	x	Probability
HH	2	1/4
HT	1	1/4
TH	1	1/4
TT	0	1/4

x	p(x)
0	1/4
1	1/2
2	1/4



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

The probability distribution can be used to calculate probabilities of different events.

Example 1 continued

$$P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

Requirements of a Discrete Probability Distribution

If a discrete random variable can take values $X = x_i$, then the following must be true:

$$1.\ 0 \le p(x_i) \le 1 \text{ for all } x_i$$

$$2.\ \sum_{\text{all } x_i} p(x_i) = 1$$

Probabilities as relative frequencies

In practice, probabilities are often estimated from relative frequencies.

Example 2

The number of cars a dealer is selling daily was recorded over the last 200 days. The data are summarised as follows:

Daily sales	Frequency
0	10
1	30
2	70
3	50
4	40
	200

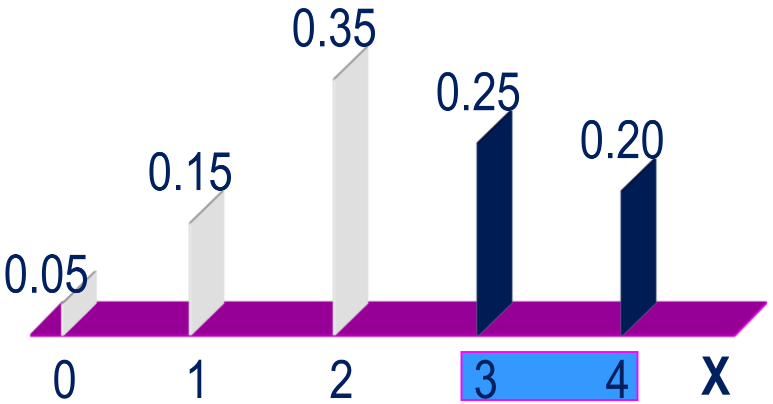
a. Estimate the probability distribution.

b. State the probability of selling more than 2 cars a day.

Solution

a. From the table of frequencies we can calculate the relative frequency, which becomes our estimated probability distribution.

Daily sales	Relative frequency
0	10/200 = 0.05
1	30/200 = 0.15
2	70/200 = 0.35
3	50/200 = 0.25
4	40/200 = 0.20
	1.00



b. The probability of selling more than 2 cars a day is

$$P(X>2) = P(X=3) + P(X=4) = 0.25 + 0.20 = 0.45.$$

Expected value and variance

The discrete probability distribution represents a ***population***.

In Example 2, the population of number of cars sold per day, and in Example 3, the population of sales call outcomes.

Since we have ***populations***, we can describe them by computing various ***parameters***.

E.g. the population mean and population variance.

The Population Mean (or Expected Value)

Population mean μ is also known as the ***expected value*** of X, denoted by E(X).

The expected value of a random variable **X** is the weighted average of the possible values it can assume, where the weights are the corresponding probabilities of each x_i.

Given a discrete random variable **X** with values x_i, that occur with probabilities $p(x_i)$, the expected value of **X** is

$$\boxed{\mu = E(X) = \sum_{all\ x_i} x_i \cdot p(x_i)}$$

Variance

The population variance is calculated in a similar manner.

The variance is the weighted average of the squared deviations of the values of **X** from their mean μ, where the weights are the corresponding probabilities of each x_i.

Let **X** be a discrete random variable with possible values x_i that occur with probabilities $p(x_i)$, and let the mean $E(X) = \mu$. The variance of **X** is defined to be

$$\sigma^2 = V(X) = E\left[(\mathbf{X}-\mu)^2\right] = \sum_{all\ x_i} (x_i - \mu)^2 p(x_i)$$

Short cut formula for σ^2

$$\sigma^2 = E\left[(X-\mu)^2\right] = \sum (x_i - \mu)^2 p(x_i) = \sum_{all\ x_i} x_i^2 p(x_i) - \mu^2$$

or

$$\sigma^2 = E(X^2) - \mu^2$$

where $E(X^2) = \sum x_i^2 p(x_i)$

Standard Deviation

The *standard deviation* of a random variable **X**, denoted σ, is the positive square root of the variance of **X**.

$$\sigma = SD(X) = \sqrt{VAR(X)}$$

Example 3

The total number of cars to be sold next week is described by the following probability distribution.

Determine the expected value and standard deviation of **X**, the number of cars sold.

x	0	1	2	3	4
p(x)	0.05	0.15	0.35	0.25	0.20

Example 3: Solution

x	0	1	2	3	4
p(x)	0.05	0.15	0.35	0.25	0.20

$$E(X) = \mu = \sum_{i=1}^{5} x_i p(x_i)$$

$$= 0(0.05) + 1(0.15) + 2(0.35) + 3(0.25) + 4(0.20)$$

$$= 2.40 \text{ cars}$$

$$V(X) = \sigma^2 = \sum_{i=1}^{5} (x_i - 2.4)^2 p(x_i)$$

$$= (0-2.4)^2(.05) + (1-2.4)^2(.15) + (2-2.4)^2(.35)$$

$$+ (3-2.4)^2(.25) + (4-2.4)^2(.20) = 1.24\,(\text{cars})^2$$

$$\sigma = \sqrt{1.24} = 1.11 \text{ cars}$$

Laws of Expected Value and Variance

Let X be a random variable and c and b are constants.

Laws of expected value

- $E(c) = c$
- $E(\mathbf{X} + c) = E(\mathbf{X}) + c$
- $E(c\mathbf{X}) = cE(\mathbf{X})$
- $E(c\mathbf{X} + b) = cE(\mathbf{X}) + b$

Laws of variance

- $V(c) = 0$
- $V(\mathbf{X} + c) = V(\mathbf{X})$
- $V(c\mathbf{X}) = c^2V(\mathbf{X})$
- $V(c\mathbf{X} + b) = c^2V(\mathbf{X})$

Example 4…

With the probability distribution of cars sold per week, assume a salesman earns a fixed weekly wage of \$150 plus \$200 commission for each car sold. What is his expected wage, and the variance of the wage, for the week?

Solution

- The weekly wage $Y = 200X + 150$
- $E(Y) = E(200X + 150) = 200E(X) + 150$ $E(cX + b) = cE(X) + b$

  $= 200(2.4) + 150 = \$630$
- $V(Y) = V(200X + 150) = 200^2V(X)$ $V(cX+b) = c^2V(X)$

  $= 200^2(1.24) = 49\,600\ (\$)^2$
- $SD(Y) = \sqrt{49600} = \$222.71$

Example 4

The monthly sales at a computer store have a mean of \$25,000 and a standard deviation of \$4,000. Profits are 30% of the sales less fixed costs of \$6,000.

Find the mean and standard deviation of the monthly profit.

Solution

- E(Sales) = \$25000

  SD(Sales) = \$4000

  V(Sales) = $(4000)^2$ $(\$)^2$
- Profit = 0.30(Sales) – 6000