Continuous Random Variables

- use relative frequencies and histograms obtained from data to estimate probabilities associated with a continuous random variable

Discrete Distribution

- A probability distribution is a formula, or a table used to assign probabilities to each possible value of a random variable, X.

- A probability distribution may be either discrete or continuous.

    - A discrete distribution means that X can assume one of a countable (usually finite) number of values.

    - A continuous distribution means that X can assume one of an infinite (uncountable) number of different values.

- With discrete probability distribution, each possible value of the discrete random variable, X can be associated with a non-zero probability.

- An example if the event of tossing a coin three times and counting the number of heads that land face-up. This is described in the table below:

Number of Heads, x	0	1	2	3
Probability, $p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

- $p(x) = P(X = x)$; in other words, the probability that the random discrete variable X is one of the possible values (outcomes) of the event. For example, based on the table above, the probability of 2 heads face-up is $\frac{3}{8}$.

- The expected value, $E(x)$ or μ, of a discrete distribution is the weighted mean of all the possible outcomes of an event. They are weighted according to their probabilities. Mathematically, this is expressed as:

$$\mu = E(X) = \sum x \times p(x)$$

- Variance, $Var(X) = \sigma^2$, measures the spread of the outcomes. It is the weighted mean of the difference between the value and the expected outcome. Again, they are weighted according to their probabilities. Mathematically this is expressed as:

$$\begin{aligned} Var(X) &= \sum (x-\mu)^2 \times p(x) \\ &= E[(X-\mu)^2] \end{aligned}$$

Variance, $Var(X)$, also has an alternate form that is preferable (not compulsory) to use when the mean is not an integer:

$$\begin{aligned} Var(X) &= \left[\sum x^2 \times p(x)\right] - \mu^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

WORKED EXAMPLE

A grocery store manager decided to complete an audit on all the tomato crates in the store's inventory. He found that 95% carry no rotten tomatoes, 2% carry one rotten tomato, 2% carry two rotten tomatoes, and 1% carry three rotten tomatoes.

a) Draw up an appropriate table to show the probabilities described above.

b) Calculate the mean, $E(X)$, variance $Var(X)$, and standard deviation, σ.

SOLUTIONS

a)

Rotten tomatoes, x	0	1	2	3
Probability, $p(x)$	0.95	0.02	0.02	0.01

b)

Rotten tomatoes, x	0	1	2	3	Sum
Probability, $p(x)$	0.95	0.02	0.02	0.01	1
$x \times p(x)$	0	0.02	0.02	0.03	0.07
$x^2 \times p(x)$	0	0.02	0.04	0.09	0.15

$E(X^2)$

$E(X)$

To find the mean:

$$E(X) = \sum x \times p(x)$$

$$= 0.07$$

To find the variance:

$$Var(X) = E(X^2) - E(X)^2$$

$$= 0.15 - 0.07^2$$

$$= 0.1451$$

To find the standard deviation:

$$\sigma = \sqrt{Var(X)}$$

$$= \sqrt{0.1451}$$

$$= 0.38092$$

CUMULATIVE DISTRIBUTION FUNCTION

- The cumulative distribution function, $F(x)$, is obtained by adding all the probabilities up to a certain point.

- It is useful to answer questions such as: "What is the probability that the first two outcomes occur?"

- Let's add another row to a previous table that describes the event of tossing a coin three times and adding the number of heads that land face-up.

Number of Heads, x	0	1	2	3
Probability, $p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$F(x)$	$\frac{1}{8}$	$\frac{4}{8}=\frac{1}{2}$	$\frac{7}{8}$	$\frac{8}{8}=1$

  - Hence, if the question asked what is the probability that there are $0, 1$ or 2 heads landing face up in this event, the answer would be:

$$F(2) = p(0) + p(1) + p(2)$$

$$= \frac{4}{8}$$

$$= \frac{1}{2}$$

- You may also see $F(X)$ expressed as $P(X \leq x)$. The latter provides a more intuitive understanding of its function.

Relative Frequency

- Relative frequencies, f_r, are estimates of the probabilities of the outcomes of an experiment – often referred to as experimental probabilities.
- Unless the experiment is biased, the greater the number of repetition (trials) of an experiment, the closer and closer the relative frequency will be to the theoretical probabilities; $f_r = \frac{f}{n}$ where n is the total number of 'trials'.
- For example, consider a survey that was conducted with a sample size of 50 employed people in a small town. They were asked about their annual incomes.

Annual Income	\$20 000	\$30 000	\$40 000	Sum
Frequency	16	11	23	50
Relative Frequency	$\frac{16}{50}$	$\frac{11}{50}$	$\frac{23}{50}$	1

We can then find the sample mean, $\bar{x}$, variance s^2, and standard deviation s. The calculations are as follows by adding two extra rows to the table:

x	\$20 000	\$30 000	\$40 000	Sum
f	16	11	23	50
f_r	$\frac{16}{50}$	$\frac{11}{50}$	$\frac{23}{50}$	1
xf_r	6 400	6 600	18 400	31 400
x^2f_r	128 000 000	198 000 000	736 000 000	1 062 000 000

To find sample mean, $\bar{x}$:

$$\bar{x} = \sum xf_r$$

$$= 31\,400$$

To find variance, s^2:

$$s^2 = \sum x^2 f_r - \bar{x}^2$$

$$= 1\,062\,000\,000 - (31\,400)^2$$

$$= 76\,040\,000$$

To find standard deviation, s:

$$s = \sqrt{s^2}$$

$$= \sqrt{76\,040\,000}$$

$$= 8\,720.09$$

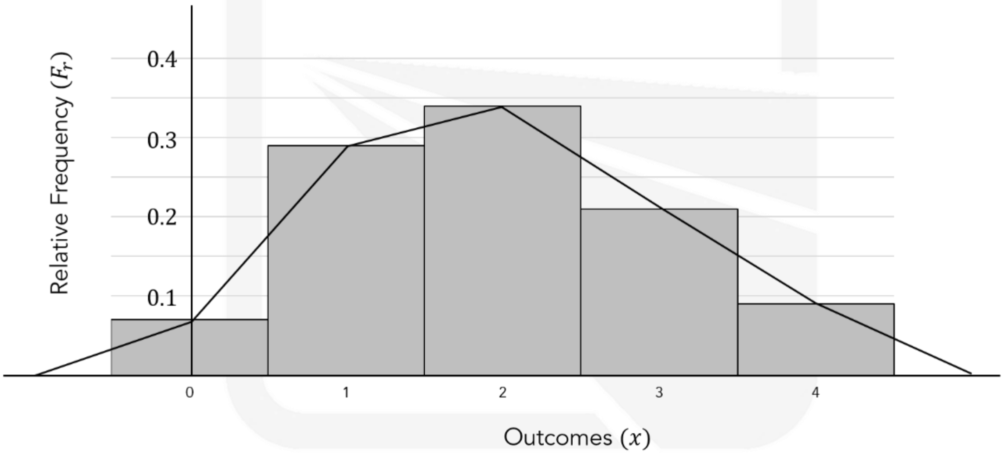
- Note that the above values are only estimates of the provability distribution:
    - $\bar{x}$ is an estimate of the expected value, $\mu = E(X)$
    - s^2 is an estimate of variance, $\sigma^2 = Var(X)$
    - s is an estimate of standard deviation, σ
- They are estimates because the calculations used the relative frequency, f_r, of each value which is an approximation of the theoretical probabilities.
- Rarely will f_r be exactly the same as the theoretical probabilities, although it does get closer and closer as the number of trials increase (as mentioned earlier).

HISTOGRAMS AND POLYGONS USING RELATIVE FREQUENCIES

- The relative frequencies of the outcomes of an event can be graphed using a relative frequency histogram and polygon.

- Consider the event of tossing a coin 100 times and counting the number of heads landing face-up. The results are tabulated below, followed by the corresponding relative frequency histogram and polygon.

x	0	1	2	3	4	Sum
f	7	29	34	21	9	100
f_r	0.07	0.29	0.34	0.21	0.09	1



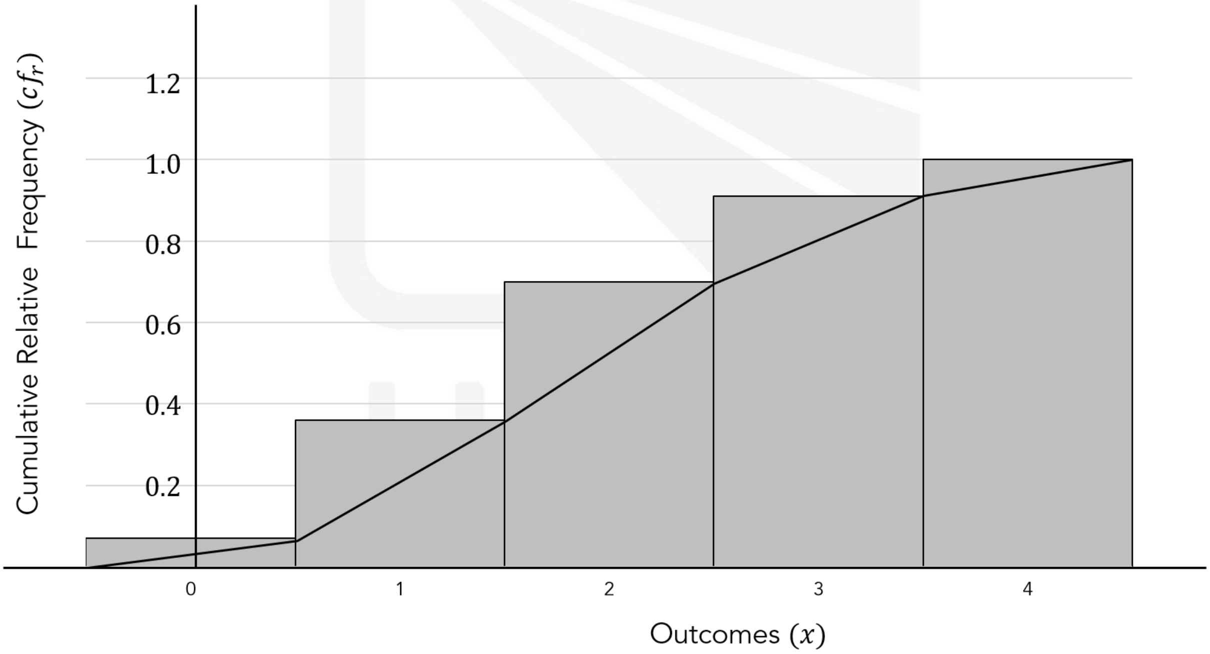
- There are some important properties to note about this graph:
  - The area underneath the polygon and histogram rectangles are equal
  - When the rectangles have a width of 1, then the sum of the areas of the rectangle is 1 because the sum of the probabilities (y −axis) also equal 1. This is shown mathematically:

$$\begin{aligned} Area &= (1)(0.07) + (1)(0.29) + (1)(0.34) + (1)(0.21) + (1)(0.09) \\ &= (1)(0.07 + 0.29 + 0.34 + 0.21 + 0.09) \\ &= (1)\left(\sum f_r\right) \\ &= (1)(1) \\ &= 1 \end{aligned}$$

CUMULATIVE RELATIVE FREQUENCIES

- As we have learnt in the first theory booklet for 'Statistical Analysis', a cumulative frequency polygon is called an ogive.
- Each cumulative frequency provides an estimate for the probability of obtaining all the outcomes up to a certain point. It is a proxy for the cumulative probability, $F(X)$.
- Continuing from the example of the tossing of coins and observing the number of heads that land face-up, we can add an additional row at the bottom of the table to include the cumulative frequency.

x	0	1	2	3	4	Sum
f	7	29	34	21	9	100
f_r	0.07	0.29	0.34	0.21	0.09	1
cf_r	0.07	0.36	0.70	0.91	1	$-$



- For example, the estimated probability of tossing 2 or fewer heads will be 0.70. In other words, there is approximately a 70% chance.

QUANTILES

- In statistics and probability, quantiles are cut points that divide the range (y −axis) of a probability distribution into continuous intervals with equal probabilities.

- The following are a list of useful quantiles that you are likely to come across:

  - Quartiles: divides the number of data points into four equal groups

  - Decile: divides the number of data points into ten equal groups

  - Percentile: divides the number of data points into one hundred equal groups

- When a cumulative frequency graph is drawn, finding a quantile is as easy as drawing a horizontal line that intercepts the cumulative frequency polygon.

- As an arbitrary example, the 4^{th} decile and the 90^{th} percentile are indicated in the graph below: