"JUST THE MATHS"

SLIDES NUMBER

19.4

PROBABILITY 4
(Measures of location and dispersion)

by

A.J.Hobson

19.4.1 Common types of measure

UNIT 19.4 - PROBABILITY 4

MEASURES OF LOCATION AND DISPERSION

19.4.1 COMMON TYPES OF MEASURE

We include three common measures of location (or central tendency) used in the discussion of probability distributions and one common measure of dispersion (or scatter).

They are as follows:

(a) The Mean

(i) For Discrete Random Variables

If the values $x_1,\ x_2,\ x_3,\ \ldots.,\ x_n$ of a discrete random variable, x, have probabilities $P_1,\ P_2,\ P_3,\ \ldots.,P_n$ respectively, then P_i represents the expected frequency of x_i divided by the total number of possible outcomes.

For example, if the probability of a certain value of x is 0.25, then there is a one in four chance of its occurring.

The arithmetic mean, μ, of the distribution may therefore be given by the formula

$$\mu = \sum_{i=1}^{n} x_i P_i.$$

(ii) For Continuous Random Variables

In this case, we use the probability density function, $f(x)$, for the distribution, which is the rate of increase of the probability distribution function, $F(x)$.

For a small interval, δx of x-values, the probability that any of these values occurs is approximately $f(x)\delta x$, which leads to the formula

$$\mu = \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x.$$

(b) The Median

(i) For Discrete Random Variables

The median provides an estimate of the middle value of x, taking into account the frequency at which each value occurs.

More precisely, the median is a value, m, of the random variable, x, for which

$$P(x \leq m) \geq \frac{1}{2} \textbf{ and } P(x \geq m) \geq \frac{1}{2}.$$

The median for a discrete random variable may not be unique (see Example 1, following).

(ii) For Continuous Random Variables

The median for a continuous random variable is a value of the random variable, x, for which there are equal chances of x being greater than or less than the median itself.

More precisely, it may be defined as the value, m, for which $P(x \leq m) = F(m) = \frac{1}{2}$.

Note:
Other measures of location are sometimes used, such as **"quartiles"**, **"deciles"** and **"percentiles"**, which divide the range of x values into four, ten and one hundred equal parts respectively.

For example, the third quartile of a distribution function, $F(x)$, may be defined as a value, q_3, of the random variable, x, such that

$$F(q_3) = \frac{3}{4}.$$

(c) The Mode

The mode is a measure of the most likely value occurring of the random variable, x.

(i) For Discrete Random Variables

In this case, the mode is any value of x with the highest probability, and, again, it may not be unique (see Example 1, following).

(ii) For Continuous Random Variables

In this case, we require a value of x for which the probability density function (measuring the concentration of x values) has a maximum.

(d) The Standard Deviation

The most common meaure of dispersion (or scatter) for a probability distribution is the **"standard deviation"**, σ.

(i) For Discrete Random Variables

In this case, the standard deviation is defined by the formula

$$\sigma = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2 P(x)}.$$

(ii) For Continuous Random Variables

In this case, the standard deviation is defined by the formula

$$\sigma = \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, \mathrm{d}x},$$

where $f(x)$ denotes the probability density function.

Each measures the dispersion of the x values around the mean, μ.

Note:
σ^2 is known as the **"variance"** of the probability distribution.

EXAMPLES

1. Determine (a) the mean, (b) the median, (c) the mode and (d) the standard deviation for a simple toss of an unbiased die.

   Solution

   (a) The mean is given by

$$\mu = \sum_{i=1}^{6} i \times \frac{1}{6} = \frac{22}{6} = 3.5$$

   (b) Both 3 and 4 on the die fit the definition of a median since

$$P(x \leq 3) = \frac{1}{2}, \quad P(x \geq 3) = \frac{2}{3}$$

   and

$$P(x \leq 4) = \frac{2}{3}, \quad P(x \geq 4) = \frac{1}{2}.$$

   (c) All six outcomes count as a mode since they all have a probability of $\frac{1}{6}$.

(d) The standard deviation is given by

$$\sigma = \sqrt{\sum_{i=1}^{6} \frac{1}{6}(i - 3.5)^2} \simeq 2.917$$

2. Determine (a) the mean, (b) median and (c) the mode and (d) the standard deviation for the distribution function

$$F(x) \equiv \begin{cases} 1 - e^{-\frac{x}{2}} & \text{when } x \geq 0; \\ 0 & \text{when } x < 0. \end{cases}$$

Solution

First, we need the probability density function, $f(x)$, which is given by

$$f(x) \equiv \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} & \text{when } x \geq 0; \\ \\ 0 & \text{when } x < 0 \end{cases}$$

Hence,

(a)

$$\mu = \int_0^\infty \frac{1}{2} x e^{-\frac{x}{2}} \,\mathrm{d}x.$$

On integration by parts, this gives

$$\mu = \left[-xe^{-\frac{x}{2}}\right]_0^\infty + \int_0^\infty e^{-\frac{x}{2}} \,\mathrm{d}x = \left[-2e^{-\frac{x}{2}}\right]_0^\infty = 2.$$

(b) The median is the value, m, for which

$$F(m) = \frac{1}{2}.$$

That is,

$$1 - e^{-\frac{m}{2}} = \frac{1}{2},$$

giving

$$-\frac{m}{2} = \ln\left[\frac{1}{2}\right].$$

Hence, $m \simeq 1.386$.

(c) The mode is zero since the maximum value of the probability density function occurs when $x = 0$.

(d) The standard deviation is given by

$$\sigma^2 = \int_0^\infty \frac{1}{2}(x-2)^2 e^{-\frac{x}{2}}\ \mathrm{d}x.$$

On integration by parts, this gives

$$\sigma^2 = -\left[(x-2)^2 e^{-\frac{x}{2}}\right]_0^\infty + \int_0^\infty 2(x-2)e^{-\frac{x}{2}}\ \mathrm{d}x$$

$$= 4 - \left[4(x-2)e^{-\frac{x}{2}}\right]_0^\infty\ 4e^{-\frac{x}{2}}\ \mathrm{d}x = 4$$

Thus $\sigma = 2$.