**WK 13:** probability

**11.3: RANDOM variables & probability functions**

Random variables

A random variable (r.v.) is a numeric quantity that changes from trial to trial in a random process.

- We often use a capital letter, e.g. X or Y, to denote a random variable, e.g.:
  - X = number of home team wins in World Series
  - Y = sum of two dice rolls
  - G= grade on next statistics exam
  - T = time to run 1500m
  - W = weight of a rat
- We want to answer probability questions about events determined by a random variable
  - E.g. the probability that the sum of 2 dice rolls is 8 (Y = 8)

DISCRETE v CONTINUOUS RANDOM VARIABLES

- A random variable is discrete if it has a finite set of possible values
  - E.g. result of a die roll [1, 2, 3, 4, 5, or 6]
  - X = home wins in World Series = {0, 1, 2,…}
  - Y = sum of two dice rolls = {2, 3,…, 12}
- A random variable is continuous if it can take any value within some interval
  - E.g. the amount of weight change and time needed to read the text
  - T = time to run 1500m
  - W = weight of a rat
- The distinction between discrete and continuous random variables if important for determining how we express their probabilities
  - Continuous random variables: we use a density curve, where the probability of being in some region is found as the area under the density curve
  - Discrete: we determine probabilities by specifying a probability for each possible value of the random variable

    → We will assume that a random variable is discrete with a relatively small set of possible values

Probability functions

For a discrete random variable, a probability function gives the probability for each of its possible values. We often use notation such as *p(2)* as shorthand to denote the probability of the event, *P(X = 2)*

- In some cases, the probability function may be given as a mathematical expression, but often we simply give a table of the probabilities for each possible value

PROBABILITY FUNCTION FOR A DISCRETE RANDOM VARIABLE:

A probability function assigns a probability, between 0 and 1, to every value of a discrete random variable. The sum of all of these probabilities must be one, i.e. $\Sigma p(x) = 1$

TEXTBOOK EXAMPLE: suppose that we roll a fair six-sided die and let the random variable X be the value showing at the top of the die. Find the probability function for X

There are six possible values, {1, 2, 3, 4, 5, 6}, that are all equally likely, so the probability of each is 1/6. We can express this as a table:

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $p(x)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

Or simply write $p(x) = \frac{1}{6}$ for $x = 1, 2, \dots, 6$

LECTURE EXAMPLE: X = # heads in two coin flips

- The 4 (equally likely) outcomes are {HH, TH, HT, TT}

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

What is the probability that there is exactly one H from two flips?

$$P(X = 1) = p(1) = \frac{1}{2}$$

LECTURE XAMPLE: sum of two die

| $x$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(x)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

Find the probability of rolling less than 5

$$P(X < 5) = p(2) + p(3) + p(4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = 1/6$$

Find the probability of not rolling a 7

$$P(X \neq 7) = 1 - p(7) = 1 - \frac{6}{36} = \frac{5}{6}$$

<u>Mean of a random variable</u>

TEXTBOOK EXAMPLE: RAFFLE TICKETS. The grand prize is \$500, two second prizes of \$100 and ten third prizes of \$20 each. They plan to sell 1000 tickets at \$2/ticket. What is the average amount of money won with each ticket in the lottery?

The total amount of prize money is \$500 . 1 + \$100 . 2 + \$20 . 10= \$900. Since there are 1000 tickets sold, the average amount won per ticket is \$900/1000 = \$0.90.

We can formalise this process to find the mean of any random variable if we know its probability function. Letting X represent the amount won with a raffle ticket, the probability function is shown below.

**Table 11.9** ***Probability function for raffle winnings***

| $x$ | 500 | 100 | 20 | 0 |
|---|---|---|---|---|
| $p(x)$ | $\frac{1}{1000}$ | $\frac{2}{1000}$ | $\frac{10}{1000}$ | $\frac{987}{1000}$ |

The process of calculating the total winnings and dividing by the number of tickets sold is equivalent to multiplying each value of the random variable by its corresponding probability and adding the results:

$$500 \cdot \frac{1}{1000} + 100 \cdot \frac{2}{1000} + 20 \cdot \frac{10}{100} + 0 \cdot \frac{987}{1000} = 0.90$$

- We call this the mean or expected value of the random variable, X
  - Since this represents the average value over the "population" of all tickets, we use the notation $\mu = 0.90$ to represent this mean

In general, we find the mean for a random variable from its probability function by multiplying each of the possible values by the probability of getting that value and summing the results

MEAN OF A RANDOM VARIABLE

For a random variable $X$ with probability function $p(x)$, the mean, $\mu$ is:

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

TEXTBOOK EXAMPLE: find the mean of the sum of two dice rolls using the probability function given in the below table:

| $x$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(x)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

We multiply each of the possible values from the sum of two dice rolls by its corresponding probability given in the Table, and add up the results:

$$\mu = 2 \,.\frac{1}{36} + 3.\frac{2}{36} + 4.\frac{3}{36} + 5.\frac{4}{36} + 6.\frac{5}{36} + 7.\frac{6}{36} + 8.\frac{5}{36} + 9.\frac{4}{36} + 10.\frac{3}{36} + 11.\frac{2}{36} + 12.\frac{1}{36} = 7$$

LECTURE EXAMPLE: marks in a quiz

| x | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| p(x) | 0.23 | 0.18 | 0.24 | 0.22 | 0.13 |

$$\mu = 1(0.23) + 2(0.18) + 3(0.24) + 4(0.22) + 5(0.13) = 2.84$$

LECTURE EXAMPLE: lottery

A $1 lottery game has prizes given by

| x | $55 | $5 | $1 | $0 |
|---|---|---|---|---|
| p(x) | 0.003 | 0.043 | 0.213 | 0.741 |

Find the mean prize for a single ticket

$$\mu = 55(0.003) + 5(0.043) + 1(0.213) + 0(0.741)$$

Is it worth the $1 cost to play?

On average, you would loose about 41c for every $1 ticket

Standard deviation of a random variable

Standard deviation is a way to measure the variability in a sample:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

- We apply similar reasoning to measure the standard deviation in a population that is defined by a random variable with probability function *p(x)*
  - To do this, we find avg. squared deviation from mean, $\mu$, and then take a square root

STANDARD DEVIATION OF A RANDOM VARIABLE

The variance, $\sigma^2$, for a discrete r.v. $X$ with probability function $p(x)$and mean $\mu$ is given by:

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

The standard deviation is given by:

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

TEXTBOOK EXAMPLE: find the standard deviation of the random variable $X = the\ sum\ of\ two\ dice\ rolls$

We found above that the mean of $X$ is $\mu = 7$. To find the variance, $\sigma^2$, we compute the mean of the squared deviations from $u = 7$:

$$\sigma^2 = (2-7)^2\frac{1}{36} + (3-7)^2\frac{2}{36} + (4-7)^2\frac{3}{36} + \cdots + (11-7)^2\frac{2}{36} + (12-7)^2\frac{1}{36}$$
$$= 25\,.\frac{1}{36} + 16.\frac{2}{36} + 9.\frac{3}{36} + \cdots + 16\,.\frac{2}{36} + 25.\frac{1}{36}$$
$$= 5.8333$$

The standard deviation of X is

$$\sigma = \sqrt{5.833} = 2.42$$