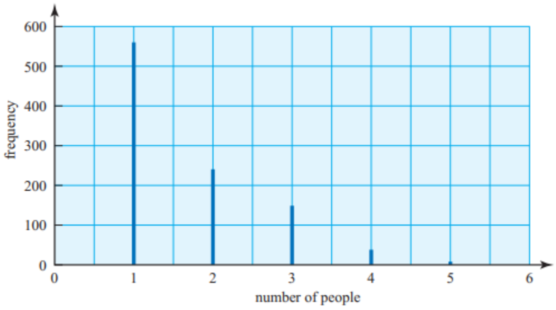
Discrete Random Variables

A. DISCRETE RANDOM VARIABLES

A traffic survey, at critical points around the town. The survey involved 1000 cars. The number of people in each car was noted, with the following results.

Number of people per car	1	2	3	4	5	> 5
Frequency	560	240	150	40	10	0

The numbers of people per car are necessarily discrete. A discrete frequency distribution is best illustrated by a vertical line chart.



The survey involved 1000 cars. This is a large sample and reasonable to use the result to estimate the *probabilities* of the various possible outcomes. Divide the *frequency* by 1000 to obtain the *relative frequency*, or *probability* of each outcome (number of people).

Outcome (Number of people)	1	2	3	4	5	> 5
Probability (Relative frequency)	0.56	0.24	0.15	0.04	0.01	0

- Notation and condition for a discrete random variable

A discrete random variable is usually denoted by an upper case letter such as *X, Y, Z*. The particular values that variable takes are denoted by lower case letters, such as *r*. Thus $P(X = r_1)$ means the probability that the discrete random variable X takes a particular value r_1. The expression $P(X = r)$ is used to express a more general idea.

Another, shorter way of writing probabilities is P_1, P_2, P_3, … . If a finite discrete random variable has *n* distinct outcomes $r_1, r_2, \ldots, r_n$, with associated probabilities $p_1, p_2, \ldots, p_n$, then the sum of the probabilities must equal 1. Since the various outcomes cover all possibilities, they are *exhaustive*.

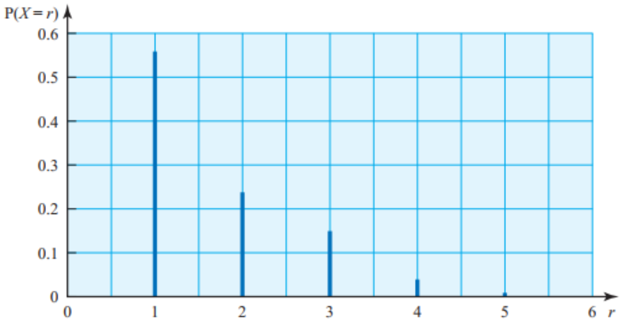
Formally, we have

$$P_1 + P_2 + \ldots + P_n = 1$$
$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} P(X = r_i) = 1$$

If there is no ambiguity, then $\sum_{i=1}^{n} P(X = r_i)$ is often aberted to $\sum P(X = r)$ *or* P_r

- Diagrams of discrete random variable

Just as with frequency distribution for discrete data, the most appropriate diagram to illustrate a discrete random variable is a vertical line chart.



The diagram shows probability distribution of X, the number of people per car.

B. EXPECTATION AND VARIANCE

Next survey, involved 800 cars. The number of people is shown in the table.

Number of people per car	1	2	3	4	5	> 5
Frequency	280	300	164	52	4	0

Use the results to estimate the probabilities of the various possible outcome as before.

The most useful measure of central tnedency is the *mean* or *expectation* of the random variable and the most useful measure of spread is the variance.

Using relative frequencies generates an alternative approach which gives the *expectation* $E(X) = \mu$ and *variance* $Var(X) = \sigma^2$ for a discrete random variable.

We define the expectation, E(X) as

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

and Variance, $Var(X)$ as

$$\sigma^2 = E([X - \mu]^2) = \sum(r - \mu^2)P_r$$

Or $$\sigma^2 = E(X^2) - \mu^2 = \sum r^2 P_r - [\sum rP_r]^2$$

These formulae can also be written as:

$$E(X) = \sum xp$$
$$Var(X) = \sum x^2 p - [E(X)]^2$$

When calculating the expectation and variance of a discrete probability distribution, you will find helpful to set your work out systematically in a table.

			(a)	(b)
r	p_r	rp_r	r^2p_r	$(r-\mu)^2p_r$
1	0.35	0.35	0.35	0.35
2	0.375	0.75	1.5	0
3	0.205	0.615	1.845	0.205
4	0.065	0.26	1.04	0.26
5	0.005	0.025	0.125	0.045
Totals	$\Sigma p_r = 1$	$\mu = \mathrm{E}(X) = 2$	4.86	$\mathrm{Var}(X) = 0.86$

In this case:

$$\mathrm{E}(X) = \mu = \sum r\,p_r$$
$$= 1 \times 0.35 + 2 \times 0.375 + 3 \times 0.205 + 4 \times 0.065 + 5 \times 0.005$$
$$= 2$$

And *either* from **(a)**

$$\mathrm{Var}(X) = \sigma^2 = \sum r^2 p_r - \left[\sum r p_r\right]^2$$

This is μ.

$$= 1^2 \times 0.35 + 2^2 \times 0.375 + 3^2 \times 0.205 + 4^2 \times 0.065 + 5^2 \times 0.005 - 2^2$$
$$= 4.86 - 4$$
$$= 0.86$$

or from **(b)**

$$\mathrm{Var}(X) = \sigma^2 = \sum (r-\mu)^2 p_r$$
$$= (1-2)^2 \times 0.35 + (2-2)^2 \times 0.375 + (3-2)^2 \times 0.205 + (4-2)^2 \times 0.065 + (5-2)^2 \times 0.005$$
$$= 0.86$$

In practice, method **(a)** is to be **preferred** since the computation is usually easier, especially when the expectation is other than a whole number.

Sample question:

[9709_s14_qp_63_003]

A pet shop has 6 rabbits and 3 hamsters. 5 of these pets are chosen at random. The random variable X represents the number of hamsters chosen.

(i) Show that the probability that exactly 2 hamsters are chosen is $\frac{10}{21}$.

(ii) Draw up the probability distribution table for X.

Solution:

(i) $P(2) = \frac{C_3^6 \times C_2^3}{C_5^9}$

(ii)

x	0	1	2	3
Prob	2/42	15/42	20/42	5/42

$$P(0) = \frac{C_5^6}{C_5^9} = \frac{6}{126}$$
$$P(1) = \frac{C_4^6 \times C_1^3}{C_5^9} = \frac{45}{126}$$
$$P(3) = \frac{C_2^6 \times C_1^3}{126}$$

[9709_w14_qp_61_002]

The number of phone calls, *X*, received per day by Sarah has the following probability distribution.

x	0	1	2	3	4	$\geqslant 5$
$\mathrm{P}(X = x)$	0.24	0.35	$2k$	k	0.05	0

(i) Find the value of *k*.
(ii) Find the mode of *X*.
(iii) Find the probability that the number of phone calls received by Sarah on any particular day is more than the mean number of phone calls received per day.

Solution:

(i) $0.24 + 0.35 + 2k + k + 0.05 = 1$

$k = 0.12$

(ii) Model number is 1

(iii) mean $= 0 \times 0.24 + 1 \times 0.35 + 2 \times 0.24 + 3 \times 0.12 + 4 \times 0.05 = 1.39$

$P(> 1.39) = P(2,3,4) = 0.41$

[9709_s13_qp_63_002]

The 12 houses on one side of a street are numbered with even numbers starting at 2 and going up to 24. A free newspaper is delivered on Monday to 3 different houses chosen at random from these 12. Find the probability that at least 2 of these newspapers are delivered to houseswith number greater

Solution:

$P(at\ least\ 2) = P(2,3)\ or\ 1 - P(0,1)$

$= \frac{5}{12} \times \frac{4}{11} \times \frac{7}{10} \times C_2^3 + \frac{5}{12} \times \frac{4}{11} \times \frac{3}{10}$

$= \frac{4}{11}\ (0.364)$

OR $\frac{(C_3^5)+(C_2^5 C_1^7)}{C_3^{12}}$