Binomial Distribution

A binomial distribution is one that consists of n repeated independent Bernoulli trials, where the probability of success is constant from trial to trial.

X represents a random variable denoting the number of successes in n Bernoulli trials

$$\boxed{P(X = x) = \binom{n}{x}(1 - p)^{n - x}p^{x}}$$

where: p is the probability of success in an individual trial

x is the number of successes

n is the number of trials

x	0	1	2	…	n	Σ
$p(x)$	$\binom{n}{0}p^0q^n$	$\binom{n}{1}p^1q^{n-1}$	$\binom{n}{2}p^2q^{n-2}$	…	$\binom{n}{n}p^nq^0$	1
$xp(x)$	0	$\binom{n}{1}p^1q^{n-1}$	$2\binom{n}{2}p^2q^{n-2}$	…	$n\binom{n}{n}p^nq^0$	$\sum_{k=0}^{n} k\binom{n}{k}p^kq^{n-k}$
$x^2p(x)$	0	$\binom{n}{1}p^1q^{n-1}$	$4\binom{n}{2}p^2q^{n-2}$	…	$n^2\binom{n}{n}p^nq^0$	$\sum_{k=0}^{n} k^2\binom{n}{k}p^kq^{n-k}$

$$E(X) = \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}$$

$$= 0\binom{n}{0}q^n + \sum_{k=1}^{n} k\binom{n}{k}p^k q^{n-k}$$

$$= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k q^{n-k}$$

$$= np\sum_{k=1}^{n} \binom{n-1}{k-1}p^{k-1} q^{n-k}$$

$$= np\sum_{k=0}^{n-1} \binom{n-1}{k}p^{k} q^{n-k-1}$$

$$= np(q+p)^{n-1}$$

$$= np$$

$$
\begin{aligned}
Var(X) &= E(X^2) - \mu^2 \\
&= E(X(X-1) + X) - n^2p^2 \\
&= E(X(X-1)) + np - n^2p^2
\end{aligned}
$$

$$
\begin{aligned}
E(X(X-1)) &= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} \\
&= 0\binom{n}{0}q^n + 0\binom{n}{1}pq^{n-1} + \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k q^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2}p^k q^{n-k} \\
&= n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}q^{n-k}
\end{aligned}
$$

$$E(X(X-1)) = n(n-1)p^2\sum_{k=0}^{n-2}\binom{n-2}{k}p^k q^{n-k-2}$$
$$= n(n-1)p^2(p+q)^{n-2}$$
$$= n(n-1)p^2$$

$$Var(X) = E(X(X-1)) + np - n^2p^2$$
$$= n(n-1)p^2 + np - n^2p^2$$
$$= n^2p^2 - np^2 + np - n^2p^2$$
$$= np - np^2$$
$$= np(1-p)$$
$$\underline{= npq}$$

If $X \sim \text{Bin}(n,p)$

$$E(X) = np$$

$$Var(X) = np(1-p)$$

$p < \frac{1}{2}$: distribution is positive skewed

$p = \frac{1}{2}$: distribution is symmetric about $x = \frac{1}{2}$

$p > \frac{1}{2}$: distribution is negative skewed e.g. (i) State whether a binomial distribution could be used in each of the following problems. If the binomial distribution is an acceptable model, define the random variable clearly and state its parameters.

a) A fair cubical dice is rolled 10 times. Find the probability of getting three fours, four fives and three sixes.

Here we are interested in three different outcomes, not binomial

Note: if we wanted just the probability of three fours it would be binomial (or four fives, or three sixes)

b) A fair coin is tossed until a head occurs. Find the probability that eight tosses are necessary.

Here the number of tosses is unknown, binomial distribution requires a fixed number of events, not binomial

Note: the random variable here is the number of tosses as that is what is of interest.

c) A jar contains 49 balls numbered 1 to 49. Six of the balls are selected at random. Find the probability that four of the six have an even score.

This depends upon whether the selections are made with, or without replacement.

Without replacement, then the probability is not the same for each selection so it is not binomial

With replacement, then it would be a binomial distribution.

X= number of balls with an even score

$$X \sim \mathrm{Bin}\left(6, \frac{24}{49}\right)$$

$$P(X=4) = \binom{6}{4}\left(\frac{25}{49}\right)^2\left(\frac{24}{49}\right)^4$$

$$= \frac{\binom{6}{4} 25^2\, 24^4}{49^6} \quad = 0.2247 \text{ (to 4 dp)}$$

(ii) In three trials of a binomial experiment, the probability of x successes is as follows;

x	0	1	2	3
$p(x)$	0.343	0.441	0.189	0.027

a) Find the probability of success at one trial

$X \sim \mathrm{Bin}(3, p)$

$$P(X = 3) = 0.027$$

$$\binom{3}{3}(1-p)^0 p^3 = 0.027$$

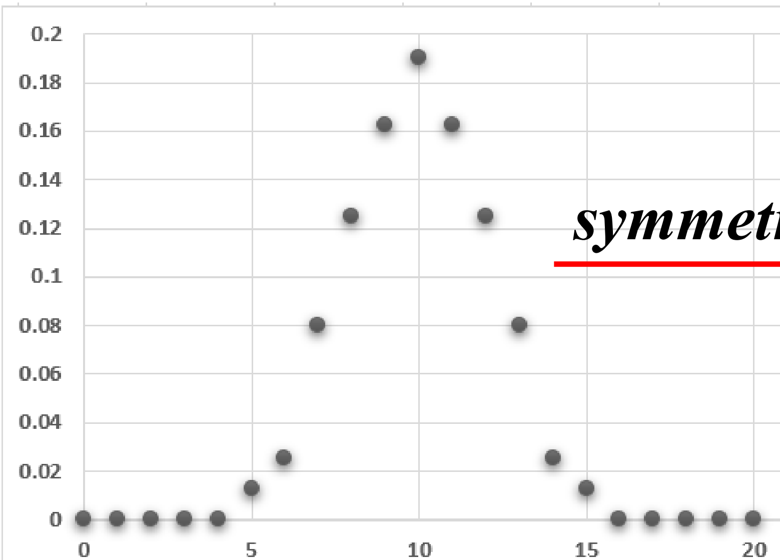
$$p^3 = 0.027$$

$$\underline{p = 0.3}$$

b) What is the variance of this distribution?

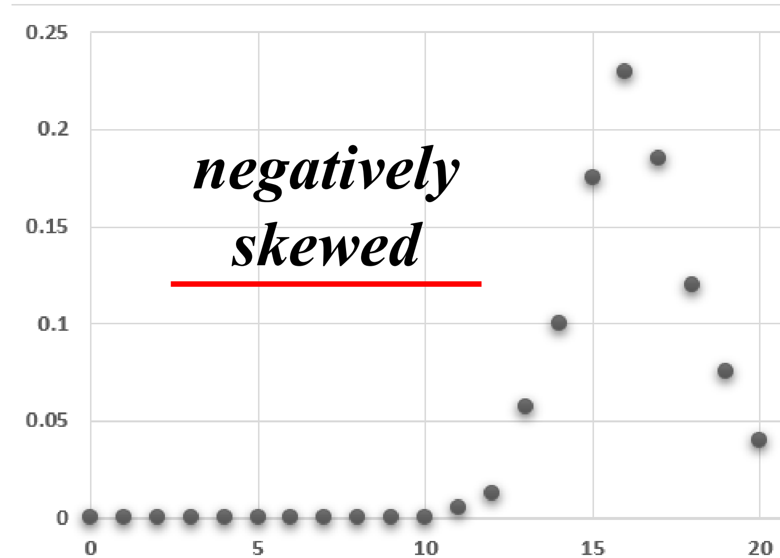
$$\begin{aligned} Var(X) &= np(1-p) \\ &= (3)(0.3)(0.7) \\ &= \underline{0.63} \end{aligned}$$

(iii) Classify each of the following distributions as positively skewed, negatively skewed or symmetric



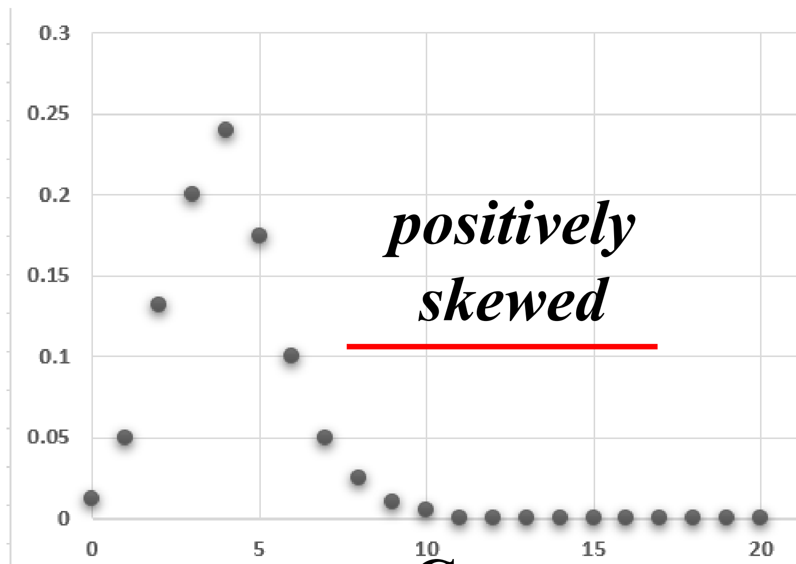
symmetric

A



negatively skewed

B



positively skewed

C

Which distribution is most likely to have an individual trial with a probability $> \frac{\pi}{5}$?

B

(iv) X is random variable whose distribution is defined as

$$X \sim \text{Bin}\left(150, \frac{2}{5}\right)$$

a) Find the mean and standard deviation of the distribution

$$\mu = np$$
$$= 150 \times \frac{2}{5}$$
$$= 60$$

$$\sigma^2 = np(1-p)$$
$$= 150 \times \frac{2}{5} \times \frac{3}{5}$$
$$= 36$$
$$\sigma = 6$$

b) How many more trials would be needed to increase the mean to 100?

$$100 = n \times \frac{2}{5}$$
$$n = 250$$

Another 100 trials would be required

c) What effect does this have on the standard deviation and distribution?

$$\sigma^2 = 250 \times \frac{2}{5} \times \frac{3}{5}$$
$$= 60$$

$\sigma = 7.75$ (to 2 dp)

The data would spread out more i.e. move away from the mean

(v) An experiment is devised where two dice are rolled ten times and the number of times the sum equals seven is noted.

a) Find the mean and standard deviation of the distribution

Let X = the number of times the sum is seven

$$X \sim \text{Bin}\left(10, \frac{1}{6}\right)$$

$$\begin{aligned} \mu &= np \\ &= 10 \times \frac{1}{6} \\ &= 1.67 \text{ (to 2 dp)} \end{aligned}$$

$$\begin{aligned} \sigma^2 &= np(1-p) \\ &= 10 \times \frac{1}{6} \times \frac{5}{6} \\ &= 1.3889 \text{ (to 4 dp)} \\ \sigma &= 1.18 \text{ (to 2 dp)} \end{aligned}$$

b) The experiment is simulated 1000 times, and the results are in the table below.

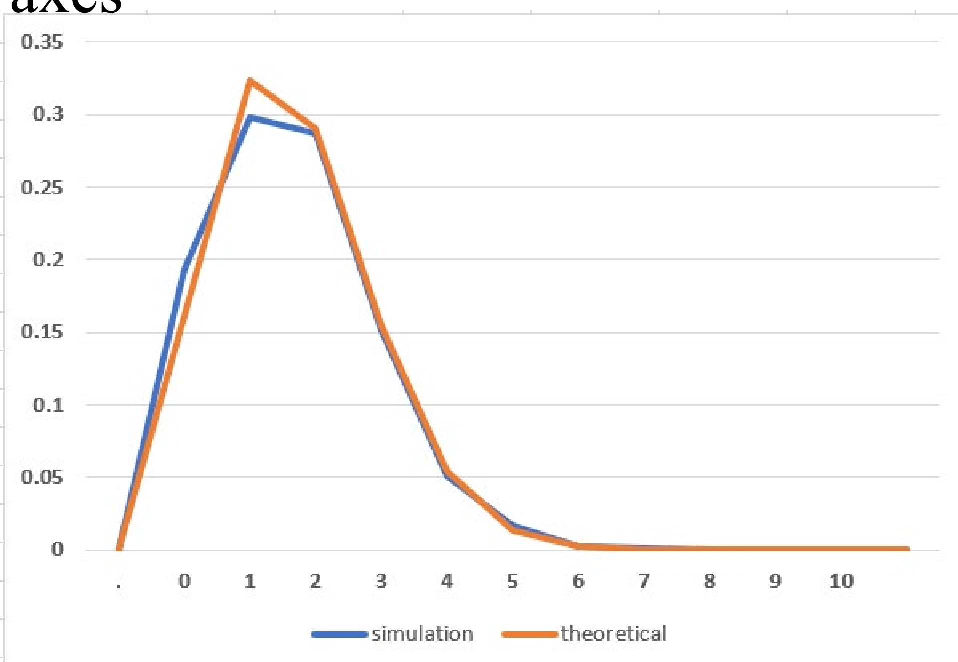
x	0	1	2	3	4	5	6	7	8	9	10	Σ
f	193	298	287	151	51	17	2	1	0	0	0	1000
f_r	0.193	0.298	0.287	0.151	0.051	0.017	0.002	0.001	0	0	0	1
xf_r	0	0.298	0.574	0.453	0.204	0.085	0.012	0.007	0	0	0	1.633
x^2f_r	0	0.298	1.148	1.359	0.816	0.425	0.072	0.049	0	0	0	4.167

c) Find the mean and standard deviation of this simulation

$$\begin{aligned}\bar{x} &= E(X)\\ &= \Sigma x f_r\\ &= 1.633\end{aligned}$$

$$\begin{aligned}s^2 &= E(X^2) - \bar{x}^2\\ &= \Sigma x^2 f_r - \bar{x}^2\\ &= (4.167) - (1.633)^2\\ &= 1.500311\\ s &= 1.22 \text{ (to 2 dp)}\end{aligned}$$

d) Draw both the simulation and experimental frequency polygons on the same set of axes



Exercise 17B; 1, 3, 4, 5, 6, 8, 10, 11,14, 15