

Chapter 9

Binomial population distribution $X \sim B(n, p)$

9.1 Definition of a Binomial distribution, $X \sim B(n, p)$

If the random variable X has a Binomial population distribution,

i.e., $X \sim B(n, p)$ then its probability function is given by

$$p(X = x) = C_x^n p^x (1 - p)^{n-x} \quad \text{for } x = 0, 1, 2, 3, \dots, n$$

A Binomial distribution is often a good model for discrete random variables which are counts of the number of times an event A occurs out of a total number of n trials or repetitions of an experiment, e.g.

X = number of ‘successes’ from n trials

X = number of heads from n tosses of a coin

X = number of sixes from n tosses of a dice

X = number of people of type A from a random sample of size n

X = number of Labour supporters in a random sample of n people

X = number of girls born from a random sample of n babies

X = number of people cured by a treatment from a random sample of size n

X = number of insects killed by a dose of insecticide from a random sample of size n

NOTE the range of possible values of the variable X is $x = 0, 1, 2, 3, \dots, n$,

i.e. X is a discrete variable which takes non-negative integer values from 0 to n inclusive.

Examples of binomial distributions

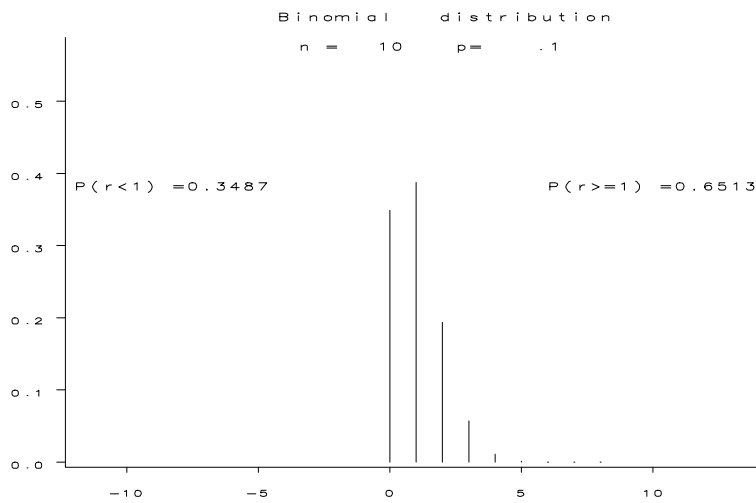


Figure 2.1: Binomial distribution with $n = 10$ and $p = 0.1$

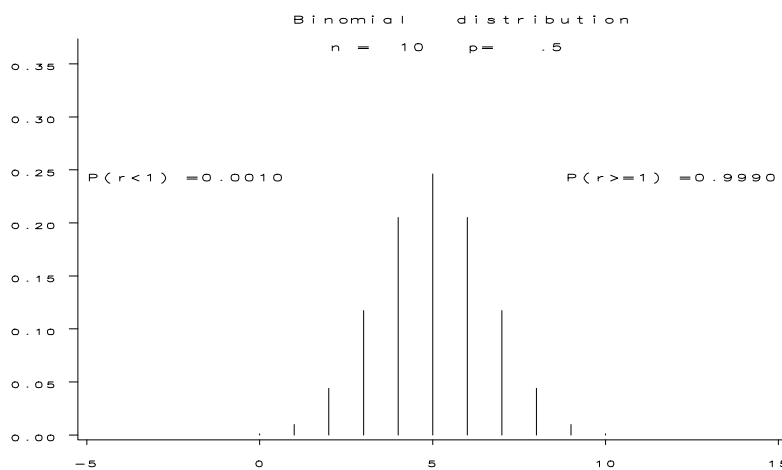


Figure 2.2: Binomial distribution with $n = 10$ and $p = 0.5$

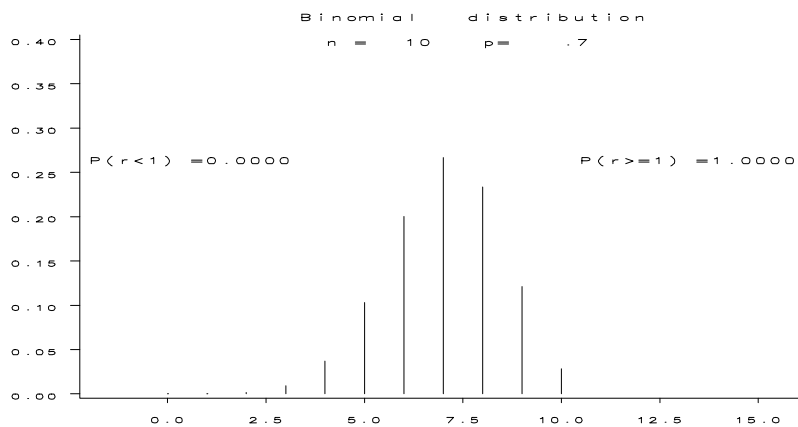


Figure 2.3: Binomial distribution with $n = 10$ and $p = 0.7$.

9.2 Theoretical conditions leading to a Binomial variable $X \sim B(n, p)$

Theorem 1

If X counts the number of times event A occurs out of n trials (i.e. repetitions of an experiment), then X has a Binomial distribution,

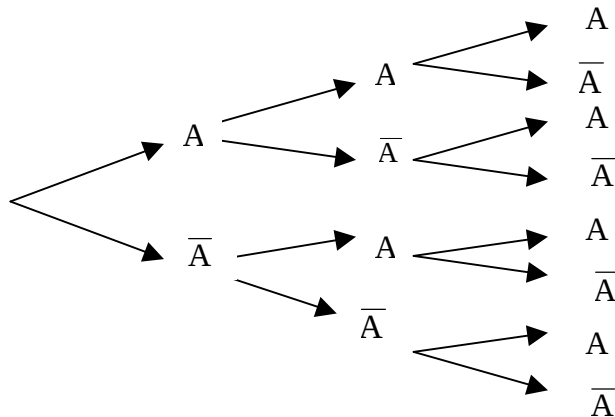
$$X \sim B(n, p)$$

with $p(X = x) = C_x^n p^x (1 - p)^{n-x}$ for $x = 0, 1, 2, 3, \dots, n$

provided

- i) the outcomes of the trials are independent of each other
- ii) the probability of event A occurring is the same value p for each of the n trials

Proof 1 First consider $n=3$



X = number of As from 3 repetitions of the experiment

e.g. X = number of heads from 3 tosses of a coin

where p = probability of heads on each toss of the coin

and $q = (1-p)$ = probability of tails on each toss of the coin

sample outcomes	x	x	$p(X=x)$
HHH	3	→ 3	p^3
HHT	2	→ 2	$3p^2q$
HTH	2	→ 2	
HTT	1	→ 2	
THH	2	→ 1	$3pq^2$
THT	1	→ 1	
TTH	1	→ 1	
TTT	0	→ 0	q^3

Proof 1 Now consider the general case of n trials (e.g. n tosses of a coin)

$p(X = x) =$ (number of different ways to get x heads from n tosses)*(probability of each way)

(number of different ways to get x heads from n tosses)

= (number of ways to choose x from n with NO repetition and order NOT important)

= C_x^n

(probability of each way)

$$= \underbrace{p(\text{HH} \dots \text{H})}_x \underbrace{p(\text{TT} \dots \text{T})}_{n-x} = pp \dots pqq \dots q = p^x q^{n-x} = p^x (1-p)^{n-x}$$

Hence

$$p(X = x) = C_x^n p^x (1-p)^{n-x}$$

9.3 Population summary measures for a Binomial variable $X \sim B(n, p)$

mean $\mu_X = np$

variance $\sigma_X^2 = npq = np(1-p)$

standard deviation $\sigma_X = \sqrt{npq}$

9.4 Procedure for finding a Binomial probability

- 1) identify the event A, 'success'
- 2) find the probability p of event A occurring in each of the trials
- 3) identify the total number n of trials

$$p(X = x) = C_x^n p^x (1-p)^{n-x}$$

9.5 Examples

Example 1

Toss a fair coin 10 times and count X the number of heads from the 10 tosses.

What is the distribution of X ?

$$X \sim B(n, p) = B(10, 0.5)$$

What is the probability of getting 8 heads from 10 tosses?

$$p(X = 8) = C_8^{10} p^8 (1-p)^2 = 45 * 0.5^8 * 0.5^2 = 45 * 0.5^{10} = 0.0439 \text{ i.e. } 4.39\%$$

What is the probability of getting 10 heads from 10 tosses?

$$p(X = 10) = C_{10}^{10} p^{10} (1-p)^0 = 1 * 0.5^{10} * 1 = 0.5^{10} = 0.000977\% \text{ i.e. } 0.0977\%$$

Example 2

Toss a fair dice 20 times and count X the number of sixes from the 20 tosses.

What is the distribution of X ?

$$X \sim B(n, p) = B(20, \frac{1}{6})$$

What is the probability of getting 5 sixes from 20 tosses?

$$p(X = 5) = C_5^{20} p^5 (1-p)^{15} = 15504 * \left(\frac{1}{6}\right)^5 * \left(\frac{5}{6}\right)^{15} = 0.1294 \quad \text{i.e. 12.94\%}$$

What is the probability of getting 0 sixes from 20 tosses?

$$p(X = 0) = C_0^{20} p^0 (1-p)^{20} = 1 * \left(\frac{1}{6}\right)^0 * \left(\frac{5}{6}\right)^{20} = 0.0261 \quad \text{i.e. 2.61\%}$$

Example 3

The probability a newborn baby is a girl is 0.487 (*Statistical Abstract of the United States*)
Let X count the number of girls out of 8 babies born in a hospital in the United States.

What is the distribution of X ?

$$X \sim B(n, p) = B(8, 0.487)$$

What is the probability of getting 6 girls from 8 babies?

$$p(X = 6) = C_6^8 p^6 (1-p)^2 = 28 * 0.487^6 * (1-0.487)^2 = 28 * 0.487^6 * 0.513^2 = 0.0983$$

What is the probability of getting 6 or more girls from the 8 babies?

$$p(X \geq 6) = p(X = 6) + p(X = 7) + p(X = 8)$$

$$p(X = 6) = 0.0983$$

$$p(X = 7) = C_7^8 p^7 (1-p)^1 = 8 * 0.487^7 * 0.513^1 = 0.0267$$

$$p(X = 8) = C_8^8 p^8 (1-p)^0 = 1 * 0.487^8 * 0.513^0 = 0.0032$$

$$p(X \geq 6) = p(X = 6) + p(X = 7) + p(X = 8) = 0.0983 + 0.0276 + 0.0032 = 0.1291$$

9.6 Using MINITAB to calculate Binomial probabilities

In example 3 above, X counts the number of girls from 8 babies born in a hospital in the USA

$$X \sim B(n, p) = B(8, 0.487)$$

Use MINITAB to output the values of the pf and cdf for the $X \sim B(n, p) = B(8, 0.487)$ model

Type values 0, 1, 2, 3, ..., 8 into C1

Probability Function

> Calc > Random Distributions > Binomial

| ☒ Probability | Number of trials '8' | Probability '0.487' | Input Variable 'C1' | OK

Binomial with n = 8 and p = 0.487000

x	P(X = x)
0.00	0.0048
1.00	0.0364
2.00	0.1210
3.00	0.2298
4.00	0.2727
5.00	0.2071
6.00	0.0983
7.00	0.0267
8.00	0.0032

Cumulative Distribution Function

> Calc > Random Distributions > Binomial

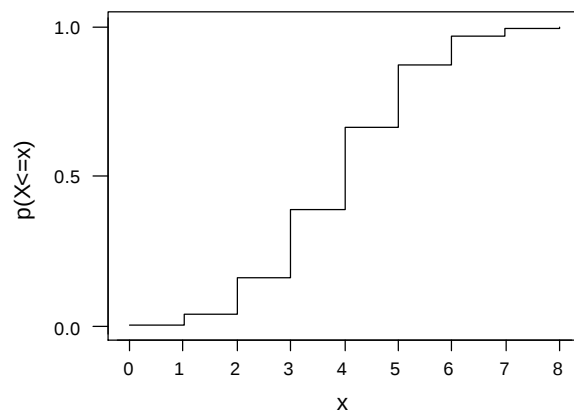
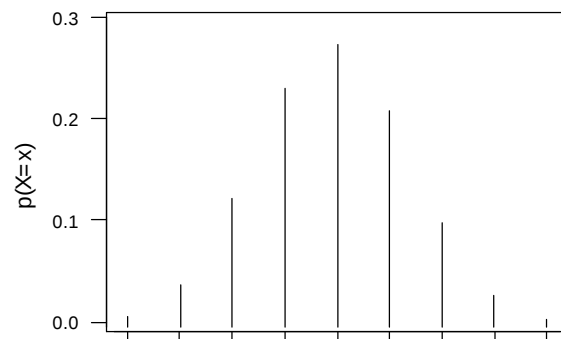
| ☒ Cumulative Prob | Number of trials '8' | Probability '0.487' | Input Variable 'C1' | OK

Binomial with n = 8 and p = 0.487000

x	P(X <= x)
0.00	0.0048
1.00	0.0412
2.00	0.1623
3.00	0.3921
4.00	0.6648
5.00	0.8719
6.00	0.9702
7.00	0.9968
8.00	1.0000

9.7 Using MINITAB to calculate and plot Binomial probabilities and cumulative probabilities

C1	C2	C3
x	$p(X=x)$	$p(X \leq x)$
0	0.004797	0.00480
1	0.036428	0.04123
2	0.121038	0.16226
3	0.229806	0.39207
4	0.272699	0.66477
5	0.207102	0.87187
6	0.098303	0.97017
7	0.026663	0.99684
8	0.003164	1.00000



9.8 Normal approximation to the Binomial

If X has a Binomial distribution,

$$X \sim B(n, p)$$

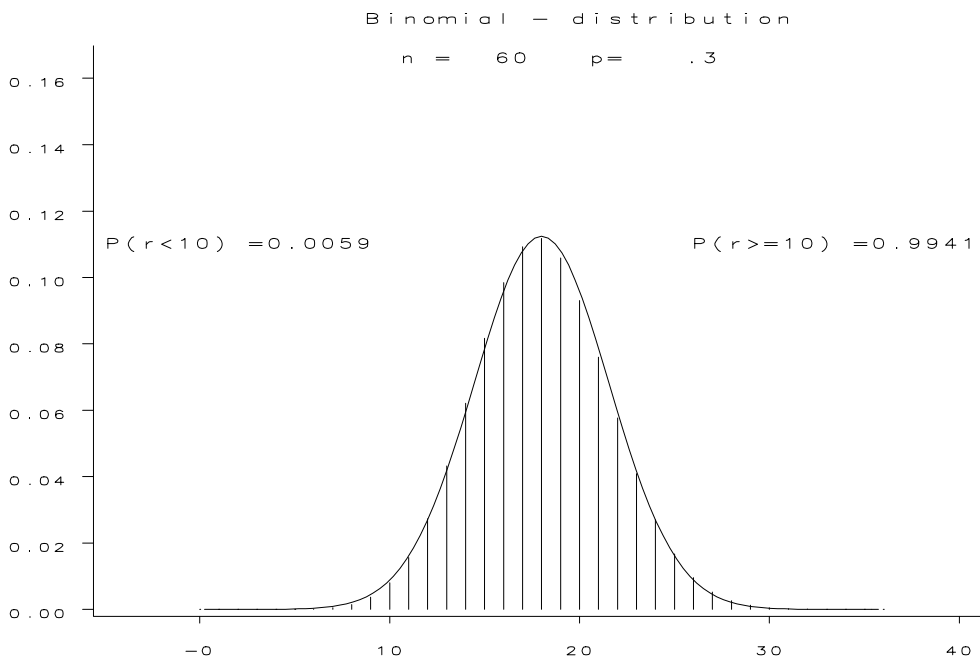
then X can be approximated by a Normal distribution with the same mean and variance as the Binomial, i.e.

$$X \sim N(\mu, \sigma^2) = N(np, npq) \quad \text{where } q = 1 - p$$

provided

n is large ($n \geq 20$) and p is not too extreme i.e. not too close to 0 or 1 ($0.1 \leq p \leq 0.9$)

Example $X \sim B(60, 0.3) \sim N(60 * 0.3, 60 * 0.3 * 0.7) = N(18, 12.6)$



PRACTICAL 6 Discrete Distributions

Q1 Binomial random variables $X \sim B(n, p)$

According to the US National Institute of Mental Health 20% of adult Americans suffer from a psychiatric disorder. [Weiss, p309]

Let X count the number of adult Americans suffering from a psychiatric disorder from a random sample of 20.

- i) state the distribution of X
- ii) state the mean and standard deviation of X
- iii) use MINITAB to calculate in C2 and plot the probability function for X by

typing the values 0, 1, 2, 3,, 20 into C1 and then

> Calc > Probability Distributions > Binomial
| ☒ Probability | Number of trials 'n' | Probability 'p' | Input Variable 'C1' | OK

> Calc > Probability Distributions > Binomial
| Output Variable 'C2' | OK

> Graph > Scatterplot > Simple
| Y 'C2' | X 'C1' |
> Data View | Data Display ☒ Project Lines | OK
> OK

- iv) use MINITAB to calculate in C3 and then plot the cumulative distribution function for X , using for the plot
> Graph > Scatterplot > With Connect Line
| Y 'C3' | X 'C1' |
> Data View | Data Display ☐ Symbols ☒ Connect Line | OK
> OK

Edit the graph by double clicking on the curve
> Options | ☒ Step > Left | OK

- v) use the output from the cdf of X from iv) to find
 $p(X \leq 3)$, $p(X < 3)$, $p(X \geq 8)$, $p(X > 8)$, $p(4 \leq X \leq 8)$, $p(4 < X < 8)$,
 $p(X = 3)$

Q2 Binomial random variables $X \sim B(n, p)$

In the UK population 6% of males but only 0.4% of females suffer from colour blindness. Random samples of males and females each of size 100 were obtained and the numbers suffering colour blindness was recorded as variables X_1 and X_2 respectively

- i) state the distributions of X_1 and X_2
- ii) state the mean and standard deviation of X_1 and X_2
- iii) use MINITAB to calculate the following
 $p(X_1 \leq 10)$, $p(X_2 \leq 10)$
- i) use MINITAB to find the median and quartiles of X_1

Q3 Poisson random variables $X \sim P(\lambda)$

The incidence of major earthquakes follows a Poisson process with mean number of earthquakes per year equal to $\lambda = 0.835$

[and hence times between earthquakes have an Exponential distribution with mean

$$\mu = \frac{1}{\lambda} = \frac{1}{0.835} = 1.1976 \text{ yrs (= 437 days)} \quad \text{Hand et al., p203]}$$

Let X count the number of earthquakes in a year.

- i) state the distribution of X
- ii) state the mean and standard deviation of X
- iii) use MINITAB to calculate the following
 $p(X \leq 2)$, $p(X < 2)$, $p(X = 0)$

Let X_2 count the number of earthquakes in a five-year period, then it can be proved that $X_2 \sim P(5\lambda)$

- i) state the mean and standard deviation of X_2
- ii) use MINITAB to calculate the following
 $p(X_2 \leq 5)$, $p(X_2 = 0)$