Statistics Sector 2: Normal Approximation to Binomial

Aims:

- recognise when the symmetry of a binomial distribution may permit the use of a normal distribution as a model and hence deduce approximate values of the mean and standard deviation of the population.

The Normal distribution can be used to approximate the Binomial distribution.

> If $X \sim B(n,p)$ then $X \approx N(np, np(1-p))$ if :
>
> n is large $(n \geq 20)$ and $p \approx 0.5$
> OR
> $np \geq 5$ and $n(1-p) \geq 5$

Example 1

Let X be a random variable such that $X \sim \mathrm{B}(150{,}0.46)$. Find a suitable normal approximation for X, giving μ and σ^2.

n is large and $p \approx 0.5$

$np = 150 \times 0.46 = 69 = \mu$

$np(1-p) = 69(1-0.46) = 37.26 = \sigma^2$

$X \sim N(69, 37.26)$

Example 2

The random variable $X \sim (800{,}0.52)$ is to be approximated using the normally distributed random variable $Y \sim N(\mu, \sigma^2)$.

a) Verify that a normal approximation is appropriate, and specify the distribution of Y.
b) Find an approximate value for $P(400 < X < 440)$

a) n is large, $p \approx 0.5$ $\therefore$ Y is Normally Dist.

$np = 800 \times 0.52 = 416$

$np(1-p) = 416(1-0.52) = 199.68$

$Y \sim N(416, 199.68)$

b) $P(400 < Y < 440) = 0.82652..$

$= 0.827$ (3sf)

Example 3

A certain variety of flower seed is sold in packets containing about 1000 seeds. The packets claim that 45% will bloom white and 55% red. This may be assumed to be accurate.

If 100 seeds are planted, use a suitable approximation to find the probability that at most 30 will bloom white.

$np = 1000 \times 0.45 = 45\cancel{0}$

$np(1-p) = 450(1-0.45) = 24.75$

$X \sim N(45, 24.75)$

$P(X \leq 30) = 0.00128$ (3sf)

Example 4

It is estimated that 8% of people are carriers of a certain disease. A random sample of people are tested for the disease.

a) Amrit suggests taking a sample of 40 people. Explain whether or not a normal approximation would be appropriate to find the probability of more than 10% of them carrying the disease.
b) What is the minimum sample size that should be used for a normal approximation in this case?
c) Use this approximation to find the probability of more than 10% of them carrying the disease.

a) $np = 3.2 < 5$ $\quad n(1-p) = 36.8 > 5$

too small $\therefore$ normal approximation should not be used.

b) $np = 5$

$n \times 0.08 = 5$

$n = 62.5$ $\quad \therefore$ sample size of at least 63.

c) $np = 0.08 \times 63 = 5.04$

$np(1-p) = 5.04(1-0.08) = 4.6368$

$X \sim N(5.04, 4.6368)$ $\quad$ 10% = 6.3 people

$P(X > 6) = 0.328$ $\quad$ $P(X > 7) = 0.181$