Lecture 19

Recall from the last lecture we discussed the concept of a sampling distribution. This referred to the probability distribution of a statistic computed from a random sample.

Today we shall discuss the sampling distribution of the sample proportion. First some definitions

Count number of occurences of some outcome in a fixed number of observations n. Random Variable X

Sample proportion $\hat{p} = \dfrac{X}{n}$

The Binomial Setting

1. Fixed number n observations (sometimes called trials)

2. Each observation is independent of all other observations
3. Each observation falls into one of two categories often referred to as "Success" (S) and "Failure" (F)

4. The probability of success, referred to using p, is same for every observation.

Some examples

1. Tossing a fair coin many times (say n)

$S \equiv$ Heads $\quad P(\text{Heads}) = p$
$F =$ Tails

2. Poll question "Do you agree with ____?"
Ask n people choosen using SRS
$S \equiv$ "Yes" $\quad P(\text{Yes}) = p$
$F =$ "No or don't have opinion"

The distribution of X (the number of Successes) in the Binomial setting is given by the binomial distribution with parameters n, p.

The Formula

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

gives the probability distribution. Note this is true for $k = 0, 1, \ldots, n$.

Combinatorial Notation

$\binom{n}{k}$ is read "n choose k" and

may be expanded to

$$\frac{n!}{k!(n-k)!}$$

It gives the number of ways to choose k objects ~~distinct~~ from n distinct objects.

Note that the "!" symbol is read as the word "factorial" and that $n! = n(n-1)(n-2)\ldots(3)(2)(1)$

eg

$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

$7! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5040$

Your calculator likely has a factorial button. Make sure you know how to use it.

A Combinatorial button is also on many calculators. Make sure you know how to use it if it is. If you don't have such a button it is very easy to use the formula above to work it out using factorials

eg $\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{120}{6(2)} = 10$

$\binom{7}{4} = \frac{7!}{4!(7-4)!} = \frac{5040}{24(6)} = 35$

<u>Note</u> $1! = 1$ and $0! = 1$

Also $\binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!0!} = 1$

$\binom{n}{0} = \frac{n!}{0!(n-0)!} = 1$

Example

Suppose $n=20$ $p=0.8$ and that X has the binomial dist. what is the probability

$P(X=11)$?

$$P(X=11) = \binom{20}{11} .8^{11} .2^{20-11} = \frac{20!}{11!\,9!} .8^{11} .2^{20-11}$$

$$= .0073 \text{ (4dp)}$$

what is $P(X=18)$?

$$P(X=18) = \binom{20}{18} .8^{18} .2^{20-18} = \frac{20!}{18!\,2!} .8^{18} .2^{2}$$

$$= .1369 \text{ (4dp)}$$

The mean of a binomial R.V. is given by

$$\mu_X = E[X] = np$$

The Variance of a binomial R.V is given by

$$\sigma^2 = Var(X) = np(1-p)$$

$\Rightarrow$ stddev $\sigma = \sqrt{np(1-p)}$

For SRS situations with sample size n and population proportion p. The count of success X in the sample has binomial distribution $B(n, p)$

notation for binomial; parameters

Recall from beginning of lecture that

$$\hat{p} = \frac{X}{n}$$

since X is a random variable with distribution $B(n,p)$ so is $\hat{p}$ a random variable. It can be shown that

$$\mu_{\hat{p}} = E[\hat{p}] = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

When the sample size is very large it can be shown that

X is approximately distributed $N(np, \sqrt{np(1-p)})$

and $\hat{p}$ is approximately distributed $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

Reasonable rules of thumb for the approximation to be accurate $np \geq 10$ $n(1-p) \geq 10$

Your textbook talks about a "continuity correction".
You do not need to know about this.

What does the above mean? It means that we can use the normal distribution to get approximate probabilities for questions about binomial random variables and sample proportions.

eg suppose $n = 200$, $p = 0.4$ and X is Binomial(n, p)

What is probability

$$P(75 \leq X \leq 85).$$

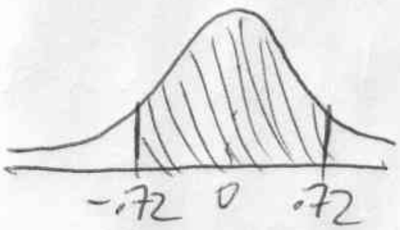
since $np = 200(0.4) = 80$ $n(1-p) = 200(.6) = 120$
are both larger than 10 the approximation is reasonable.

$$\mu_X = np = 200(0.4) = 80$$

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{200(0.4)(0.6)} = 6.9282$$

So $P(75 \le X \le 85)$

$$= P\left(\frac{75-80}{6.9282} \le \frac{X-80}{6.9282} \le \frac{85-80}{6.9282}\right)$$



$$= P(-0.72 \le z \le 0.72)$$

$$= P(z < +.72) - P(z < -.72)$$

$$= .7642 - .2358$$

from table

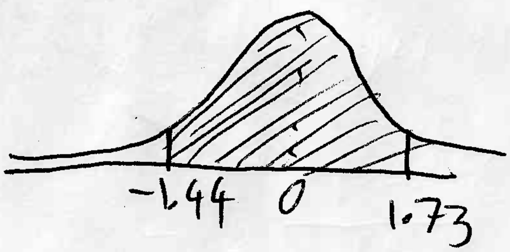
$$= .5284$$

What is probability $P(.35 \le \hat{p} \le .46)$

$\mu_{\hat{p}} = .4 \quad \sigma_{\hat{p}} = \sqrt{\frac{.4(.6)}{200}} = .0346$

So $P(.35 \le \hat{p} \le .46) = P\left(\frac{.35-.4}{.0346} \le \frac{\hat{p}-.4}{.0346} \le \frac{.46-.4}{.0346}\right)$

$$= P(-1.44 \le z \le 1.73)$$



$$= \cancel{P(1.73)}$$

$$P(z < 1.73) - P(z < -1.44)$$

$$= .9582 - .0749$$

$$= .8833$$