Chapter 5 Sampling distributions

1. Binomial distribution

a) Definition

The Binomial setting:

- A fixed n of observations
- All observations are independent
- Each observations falls into one of just two categories: "success" and "failure".
- For each observation, P(success) = p.

Let X = no. of success out of n observations, then

- X is a discrete random variable.
- Possible values of X are 0, 1, 2, …, n.

The probability distribution of X is the so called Binomial distributions with parameter n and p. We say that X is B(n, p) or **X ~ B(n, p).**

b) The mean and standard deviation of Binomial distribution

If X ~ B(n, p), then

$$\mu_X = np$$
$$\sigma_X^2 = np(1-p)$$
$$\sigma_X = \sqrt{np(1-p)}$$

c) Probability calculation for Binomial distribution

i) Binomial formulas

If X ~ B(n, p), then

Value of X	0	1	2	…	n
Probability	P(X=0)	P(X=1)	P(X=2)	…	P(X=n)

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

where $n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1$

$0! = 1$

ii) Using Table C (at the back of the text book)

Example: A university that is better known for its basketball program than for its academic strength claims that 80% of its basketball players get degrees. An investigation examines the fate of all 20 players who entered the program over a period of several years that ended 6 years ago. Of these players, 11 graduated and the remaining 9 are no longer in school.

(a) Find the probability that exactly 11 of the 20 players graduate.

(b) Find the probability that 11 or fewer players graduate. Do you think that the university's claim is reliable?

(c) Find the mean number of graduates out of 20 players if the university's claim is true.

(d) Find the standard deviation σ of the count X.

2. Statistical Inference (most material here from section 3.4 of Chapter 3)

a) Parameter and statistics

Parameter: a number that describes the population. It is an unknown constant.

Statistic: a number that describes a sample. It is a random variable. It varies from sample to sample.

b) Statistical inference

Having population ⇒ Draw a sample ⇒ construct a statistic ⇒ get the value of that statistic ⇒ conclusion

The above process is called statistical inference. I.e. statistical inference is the process of producing data in order to draw conclusions.

c) Sampling distribution

Since a statistic is a random variable, its distribution of all possible values is called sampling distribution of that statistic.

d) Bias and variability

If the mean of a statistic is equal to the parameter itself, we say that this statistic is **unbiased**, otherwise it is biased.

Variability of a statistic is described by the spread of its sampling distribution.