Math 373  Lecture 9

A bent coin is tossed 5 times. The probability of heads in each toss is \( p = .2 \). Is the probability distribution for the number \( x \) of heads in 5 tosses skewed to the left? symmetric? skewed to the right?

If \( p = .2, .5, \) and .8 respectively, the expected numbers of heads are \( E(x) = np = 5(.2) = 1, 5(.5) = 2.5, \) and \( 5(.8) = 4 \) respectively. When the expected number of successes (heads) is low \((np) \leq 5\) then the left tail is chopped off, the right tail is long and thus the curve is skewed right. When the expected number of failures is low \((nq) \leq 5\) the opposite is true and the curve is skewed to the left.

Otherwise a binomial distribution is approximately a normal distribution.

**Binomial Normal Approximation Theorem.** When the expected number of successes and failures both exceed \( 5, (np>5 \) and \( nq>5) \), an area under the binomial distribution is approximately the continuity corrected area under the normal distribution with the same mean \( \mu = np \) and std. dev. \( \sigma = \sqrt{npq} \). Let \( P_b \) and \( P_n \) be the binomial and normal probabilities, then

\[
P_b(x = k) \approx P_n(k - .5 \leq x \leq k + .5)
\]

\[
P_b(x \leq k) \approx P_n(x \leq k + .5)
\]

\[
P_b(x < k) \approx P_n(x \leq k - .5)
\]

\[
P_b(k \leq x) \approx P_n(k - .5 \leq x)
\]

\[
P_b(k < x) \approx P_n(k + .5 \leq x)
\]

The continuity correction is due to the fact that the bar above \( x = k \) has width 1 for the binomial distribution but width 0 for the continuous normal distribution. For the latter, \( P_b(x = k) = 0 \neq P_n(x = k) \). The correct approximation:

\[
P_b(x = k) \approx P_n(k - .5 \leq x \leq k + .5)
\]

Let \( x \) be a binomial variable with \( n = 100 \) and \( p = .3 \). Is the normal approximation appropriate? Yes since \( \mu = np = 30, nq = 70, \sigma = \sqrt{npq} = \sqrt{21} = 4.58 \).

Find \( P_b(x \geq 40) \) using the continuity-corrected normal approximation.

\[
P_b(x \geq 40) \approx P_n(x \geq 39.5).
\]

To convert to \( z \), subtract \( \mu \) and divide by \( \sigma \).

\[
P_N(x \geq 39.5) = P_N(\frac{x - \mu}{\sigma} \geq \frac{39.5 - 30}{4.58}) = P_N(z \geq 2.07) = P_N(2.07 \leq z) = .5 - P_N(0 \leq z \leq 2.07) = .5 - .4808 = .019
\]

A potter must fill an order for 180 glazed pots. The glazing process cracks 20% of the pots. What is the probability that he can produce 180 crack-free items from 200 unglazed pots?

\[
m = \mu = \sigma = 10
\]

\[
P(x \geq 180) = P_b(x \geq 180) \approx P_n(x \geq 179.5) \approx 0%
\]

**Choosing Statistical Samples**

Suppose you wish to estimate the average height of UH seniors. Selecting a sample from the UH basketball team would not be a representative sample. It would bias the height estimate upward.

There are many sampling plans or experimental designs for choosing samples so as to avoid bias. A standard way to avoid such biases is to pick a *random* sample. Suppose we are choosing \( n \)-element samples from a given population. An experimental design selects *random samples* if all \( n \)-element samples have the same probability of being selected. You can select random sample by assigning numbers to the population elements and randomly selecting \( n \) of these numbers. This can be done with a table of random numbers (Table 10, Appendix I). For example if we sorted UH seniors by their ID number, then we could use a table of random numbers to select a random sample. An easier way which we use in the Math Department is to sort students by the last 4 digits of their ID number, then pick the first \( n \) from this list.

A variant of this is a *1-in-k systematic sample*. As long as the population can be ordered randomly (last 4 soc. sec. digits), pick every \( k \)-th item from the list. Traffic data might be obtained by interviewing every 100th driver passing through a tollbooth.

If the population is divided into *strata* which affect the variable being measured, one should select a *stratified sample* in which a subsample is randomly selected from each strata with the size of the subsample proportional to the size of the strata. If half of UH seniors are male and half are female, then a stratified sample of 20 seniors would be selected by randomly choosing 10 male seniors and then choosing 10 female seniors. Such a sample is more likely to be representative than an unstratified sample which might end up with too many males.