Stat 321 - Day 24
Normal Approximation to Binomial Distribution

Recall that you have studied normal probability distributions and learned how to calculate probabilities using standardization and then either a normal probability table or Minitab or Java applet. Today you will see how the normal distribution can be used to approximate the binomial distribution in certain circumstances.

Example: Hospital Births
Suppose that hospital A has 20 births per day and hospital B has 100 births per day. Assume that each birth is equally likely to be a boy or girl, independently from birth to birth. Suppose that you observe these hospitals over a 365-day year and record the percentage of girls born each day.

(a) Before doing any calculations, make an educated guess as to which hospital would have more days on which 60% or more of the births are girls. Explain your reasoning.

Let X be the number of girls born in a day at hospital A, and let Y be the number of girls born in a day at hospital B.

(b) Is X discrete or continuous? What about Y? Explain.

(c) What probability distribution does X have? What about Y? [*Hint*: Specify the name and also the parameter values.]

Before we calculate exact probabilities, let us develop a sense for the probability distributions through *simulation*.

(d) Use Minitab to simulate 1000 years of 365 days for each hospital:

```
MTB> random 1000 c1;
SUBC> binomial 20 .5.
MTB> random 1000 c2;
SUBC> binomial 100 .5.
MTB> name c1 'A girl births' c2 'B girl births'
```

(Alternatively, you can choose `Calc> Random data> Binomial`.)
Create dotplots and/or histograms of these distributions. Comment on the shape.

(e) Count how many of the 1000 simulated years produced 60% or more girl births in hospital A:

```
MTB> let c3=(c1>=12)
MTB> tally c3
```

What proportion of the 1000 simulated years is this?

(f) Repeat (e) for hospital B.

(g) Based on your simulation results, which hospital is more likely to have a day with 60% or more girl births? How does this compare to your guess in (a)?

Now calculate these probabilities exactly from the binomial distribution.

(h) Use Minitab (`Calc> Probability distributions> Binomial`, with the "cumulative probability" option) to determine these exact probabilities. [*Hint*: Remember that with the binomial distribution, $P(X \geq k) = 1 - P(X \leq k-1)$.]

Hospital A: Hospital B:

(i) Are these probabilities close to the predictions of your simulations?

(j) Judging from the histograms or dotplots of your simulation results, would you say that a normal distribution would approximate the binomial distribution well in these cases? With which case (A or B) do you think the approximation will be closer? Explain.

(k) Determine the expected (mean) value and standard deviation of X and of Y. [*Hint*: Use what you know about the binomial distribution.]

E(X): E(Y):

SD(X): SD(Y):

(l) Let V be a normal random variable with mean and standard deviation equal to those of X. Determine P(V≥12). Is it reasonably close to P(X≥12)?

(m) Let W be a normal random variable with mean and standard deviation equal to those of Y. Determine P(W≥60). Is it reasonably close to P(Y≥60)?

To adjust for approximating a discrete distribution (binomial) with a continuous one (normal), we can employ a *continuity correction*. This correction recognizes that only integers are possible values of the binomial distribution and so adjusts by shifting the calculation to the appropriate midpoint between integers.

(n) Determine P(V≥11.5). Is this a more accurate approximation to P(X≥12) than you found in (l)? Then determine P(W≥59.5) and see if it provides a more accurate approximation to P(Y≥60) than you found in (m).

(o) Consider the probability that hospital B has between 45 and 55 girl births in a day. First approximate this probability through simulation. Then determine it exactly from the binomial distribution. Finally, approximate the probability using the normal distribution with continuity correction. Comment on how well the normal approximation performed in this case.

simulation:

binomial calculation:

normal approximation:

Now suppose that 20% of all births in a certain community are Hispanic babies and that hospital C has ten births per day. Let U be the number of Hispanic babies born in one day, and consider the probability of getting three or more Hispanic births in a day.

(p) First approximate this probability through simulation. Then determine it exactly from the binomial distribution. Finally, approximate the probability using the normal distribution with continuity correction.

simulation:

binomial calculation:

normal approximation:

(q) How well did the normal distribution approximate the binomial in this case? Explain why this happened, based on examination of a histogram of the simulation results.

> The *normal approximation to the binomial distribution* is more accurate for large values of n and for values of p close to .5. As a rule-of-thumb, we generally require $np \geq 10$ and $n(1-p) \geq 10$ in order to consider the approximation accurate. When this condition is satisfied, we can approximate the binomial distribution with a normal one having mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. The *continuity correction* makes the approximation more accurate.

(r) In which of the above three cases is this criterion handily satisfied? In which is it barely satisfied? In which is it not satisfied? Are the relative accuracies of the approximations as you would expect? Explain.