UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 6

The Normal Approximation to the Binomial.

For $Y \sim Binomial(n,p)$ we know that its probability mass function is:

$$p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \dots, n$$

Moreover, its expectation and its variance are

$$\mathbb{E}[Y] = np \quad \text{and} \quad Var[Y] = np(1-p).$$

Now, consider a sequence of binomial random variables $Y_n \sim Binomial(n,p)$. Note that, while the number of trials n varies, the probability of success in every trial p remains the same for all n. The *normal approximation to the binomial* is a theorem which states that

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \stackrel{\mathcal{D}}{\Rightarrow} N(0,1)$$

Practically, this means that Y_n is "approximately" normal with mean np and variance $np(1-p)$ for "large" n. The usual rule of thumb is that both $np > 10$ and $n(1-p) > 10$.

Another practical adjustment needs to be made due to the fact that discrete distributions of Y_n are approximated by a continuous (normal) distribution. This adjustment is usually referred to as the **continuity correction**. More specifically, provided that the conditions above are satisfied, for every integer a and b, we have that

$$\begin{aligned}\mathbb{P}[a \le Y_n \le b] &= \mathbb{P}\left[a - \tfrac{1}{2} < Y_n < b + \tfrac{1}{2}\right] \\ &= \mathbb{P}\left[\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}} < \frac{Y_n - np}{\sqrt{np(1-p)}} < \frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right] \approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)\end{aligned}$$

where Φ, as usual, stands for the cummulative distribution function of the standard normal distribution.

For more about the history of the theorem and ideas for its proof, go to: Wikipedia: de Moivre-Laplace.

Problem 6.1. A student takes an exam with 200 TRUE/FALSE questions. Shirley knows the correct answer to exactly 100 questions. For the remaining questions, she guesses at random. The passing mark is 136 correct answers. What is the (approximate) probability she passes the exam?

Solution: Let Y be the random variable representing the number of questions Shirley answers correctly by guessing at random. Then,

$$Y \sim Binomial(n = 100, p = 0.5).$$

We need to find the probability $\mathbb{P}[Y \ge 36]$. The conditions to use the *normal approximation to the binomial* are satisfied since $np = 50 > 10$ and $n(1-p) = 50 > 10$. The mean and the variance of Y are

$$\mathbb{E}[Y] = np = 50 \quad \text{and} \quad Var[Y] = np(1-p) = 25.$$

So, we have that

$$\mathbb{P}[Y \ge 36] = 1 - \Phi\left(\frac{35.5 - 50}{5}\right) = 1 - \Phi(-2.9).$$

Using **pnorm** in **R**, we get 0.9981342. For comparison, using **pbinom** gives us 0.9982412.

Problem 6.2. A (hypothetical) 13-sided die is thrown 169 times (each of the numbers $1, 2, \dots, 13$ is equally likely on each throw). Every time 5 or a larger number is obtained, the player wins a candy bar. What is the probability that the player will receive at least 100 candy bars? (Use the normal approximation and note that $169 = 13 \times 13$).

Solution: The probability of winning a candy bar on a given throw is $\frac{9}{13}$. Therefore, the number of candy bars won has the binomial distribution $Binomial(n,p)$ with $n = 169$ and $p = \frac{9}{13}$. We approximate $Binomial(n,p)$ by the normal distribution with mean $\mu = np = 9 \times 13 = 117$ and standard deviation $\sigma = \sqrt{npq} = 6$. We are looking for the probability of getting at least 100 candy bars, so the required probability is given by

$$1 - \Phi(\frac{99.5-117}{6}) = 1 - \Phi(-2.92) = \Phi(2.92) = 0.9982.$$