

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 6

The Normal Approximation to the Binomial.

For $Y \sim \text{Binomial}(n, p)$ 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

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

Now, consider a sequence of binomial random variables $Y_n \sim \text{Binomial}(n, p)$. 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

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

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 $a < b$, we have that

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

where Φ , as usual, stands for the cumulative 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?

→: Y ... # of correct guesses

$$Y \sim \text{Binomial}(n=200, p=0.5)$$

$$\text{Check: } n \cdot p = n \cdot (1-p) = 100 > 10 \quad \checkmark$$

We seek: $\mathbb{P}[Y \geq 136] = ?$

$$\mathbb{P}[Y \geq 136] = \mathbb{P}[Y > 135.5]$$

$$\mathbb{E}[Y] = n \cdot p = 100, \quad \text{Var}[Y] = n \cdot p(1-p) = 50 \Rightarrow \text{SD}[Y] = \sqrt{50}$$

$$\mathbb{P}[Y > 135.5] = \mathbb{P}\left[\frac{Y - 100}{\sqrt{50}} > \frac{135.5 - 100}{\sqrt{50}}\right] \approx 1 - \Phi(2.9)$$

$$\text{standard normal tables: } 1 - 0.0019 = 0.9981 \quad \square$$

OR

using 'R': 0.9981342

$Y \approx \text{Normal}(\text{mean}=50, \text{sd}=5)$

$$P[Y > 35.5] = 1 - \text{pnorm}(35.5, \text{mean}=50, \text{sd}=5) = 0.9981342$$

Just so: $1 - \text{pbinom}(35, \text{size}=100, \text{prob}=0.5) = 0.9982412$



Problem 6.2. A (hypothetical) 13-sided die is thrown 169 times (each of the numbers 1, 2, ..., 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 $169 = 13 \times 13$).

→: Y ...# of candy bars won.

$$Y \sim \text{Binomial}(n = 169, p = \frac{9}{13})$$

$$P[Y \geq 100] = ?$$

$$P[Y > 99.5] = ?$$

Check: $n \cdot p = 169 \cdot \frac{9}{13} = 117 > 10 \quad \checkmark$

$$n(1-p) = 169 \cdot \frac{4}{13} = 52 > 10 \quad \checkmark$$

$$\underline{E[Y] = 117}, \quad \text{Var}[Y] = \cancel{169} \cdot \cancel{\frac{9}{13}} \cdot \cancel{\frac{4}{13}} = 36 \Rightarrow \underline{SD[Y] = 6}$$

$$1 - \text{pnorm}(99.5, \text{mean} = 117, \text{sd} = 6) = 0.998231 \quad \square$$

Just so: $1 - \text{pbinom}(99, \text{size} = 169, \text{prob} = 9/13) = 0.9978254$