UNIVERSITY OF TEXAS AT AUSTIN

Quiz 3

The normal approximation to the binomial.

Please, provide your complete solution to the following problems. There is no need to use the continuity correction.

Problem 3.1. (5 points) According to the Pew research center, 64% of Americans say that social media have a mostly negative effect on things (see https://pewrsr.ch/3dsV7uR). You take a sample of 1000 randomly chosen Americans. What is the approximate probability that at least 650 of them say that social media have a mostly negative effect on things?

Solution: The number X of Americans in the sample who say that social media have a mostly negative effect on things is a binomial random variable with 1000 trials and 64% probability of success in every trial. Formally,

$$X \sim Binomial(n = 1000, p = 0.64).$$

We have $\mu_X = \mathbb{E}[X] = 1000(0.64) = 640$ and $\sigma_X = SD[X] = \sqrt{1000(0.64)(0.36)} = 15.17893$.

Since $1000(0.64) = 640 \geq 10$ and $1000(0.36) = 360 \geq 10$, we can safely use the normal approximation to the binomial. The problem says to ignore the continuity correction, so we will do so. We have

$$\mathbb{P}[X \geq 650] = \mathbb{P}\left[\frac{X - 640}{15.17893} \geq \frac{650 - 640}{15.17893} = 0.658808\right] \approx 1 - N(0.66)$$

where N denotes the standard normal cumulative distribution function. From the IFM tables, we obtain $1 - N(0.66) = 1 - 0.7454 = 0.2546$.

Note: If you were to calculate the above **binomial** probability in **R**, you would get 0.2663. Not too bad! Also, using the normal approximation **with** the continuity correction, we get 0.2657009. This is a better approximation.

Problem 3.2. (5 points) According to a Gallup survey, only 22% of American young adults rate their mental health as *excellent*:

https://www.gallup.com/education/328961/why-higher-education-lead-wellbeing-revolution.aspx.

You sample 6000 randomly chosen American young adults. What is the approximate probability that at most 1400 of them rate their mental health as *excellent*?

Solution: Let the random variable X denote the number of American young adults in the sample who deem their mental health *excellent*. It is a binomial random variable with 6000 trials and 22% probability of success in every trial. Formally,

$$X \sim Binomial(n = 6000, p = 0.22).$$

We have $\mu_X = \mathbb{E}[X] = 6000(0.22) = 1320$ and $\sigma_X = SD[X] = \sqrt{6000(0.22)(0.78)} = 32.08738$.

Since $6000(0.22) = 1320 \geq 10$ and $6000(0.78) = 4680 \geq 10$, we can safely use the normal approximation to the binomial. The problem says to ignore the continuity correction, so we will do so. We have

$$\mathbb{P}[X \leq 1400] = \mathbb{P}\left[\frac{X - 1320}{32.08738} \geq \frac{1400 - 1320}{32.08738} = 2.493192\right] \approx N(2.49)$$

where N denotes the standard normal cumulative distribution function. From the IFM tables, we obtain $N(2.49) = 0.9936$.

Note: If you were to calculate the above **binomial** probability in **R**, you would get 0.9936818.

Problem 3.3. (5 points) You toss a fair coin 10,000 times. What is the approximate probability that the number of *Heads* exceeds the number of *Tails* by between 200 and 500 (inclusive)?

Solution: The number of *Heads* in the $10,000$ tosses of a fair coin can be modelled using a binomial random variable X with parameters $n = 10,000$ (for the number of trials) and $p = 1/2$ (for the probability of success in every trial). Since X denotes the number of *Heads*, we have that $10,000 - X$ stands for the number of *Tails* in the same string of cointosses. We are looking for the probability

$$\begin{aligned}\mathbb{P}[200 \leq X - (10000 - X) \leq 500] &= \mathbb{P}[200 \leq 2X - 10000 \leq 500] \\ &= \mathbb{P}[10200 \leq 2X \leq 10500] = \mathbb{P}[5100 \leq X \leq 5250].\end{aligned}$$

The mean and the standard deviation of the random variable X are

$$\mu_X = \mathbb{E}[X] = 10000(1/2) = 5000 \quad \text{and} \quad \sigma_X = SD[X] = \sqrt{10000(1/2)(1/2)} = 50.$$

Evidently, we can use the normal approximation to the binomial. We get

$$\mathbb{P}[5100 \leq X \leq 5250] = \mathbb{P}\left[\frac{5100 - 5000}{50} \leq \frac{X - \mu_X}{\sigma_X} \leq \frac{5250 - 5000}{50}\right] \approx \mathbb{P}\left[2 \leq Z \leq 5\right]$$

with $Z \sim N(0,1)$. Our final answer will be

$$\mathbb{P}\left[2 \leq Z \leq 5\right] = N(5) - N(2) = 1 - N(2) = 1 - 0.9772 = 0.0228.$$

Note: Calculating the **binomial** probability directly in **R**, we get 0.02329249.