STAT 2857A – Lecture 24 Examples and Exercises

Solutions

Example 23.1

a) The description will depend on the exact values generated by the students in the class. However, it should be unimodal with a peak just left of the centre and slightly left (negatively) skewed so that students end up with balances less than FUN 5000 (i.e., less than their starting balance) more often than not. The most common outcome should be FUN 4000 with an observed proportion near .32. The next most likely outcome should be FUN 6000 with an observed proportion near .30. The least likely outcomes should be FUN 0 with an observed proportion near .04 and FUN 10000 with an observed proportion near .03.

b) Statistics that could be computed include:

- the sample mean,
- the sample variance,
- the sample median,
- the maximum,
- the minimum,
- the number, proportion, or percentage of times any specific value occurs,
- the ratio or difference of any of the number, proportion, or percentage for two values,
- etc.
- The number

Example 23.2

a) Suppose that we have n trials (the number of students who played the game). Then the average final balance is

$$\bar{W} = \frac{\sum_{i=1}^{n} W_i}{n} = \frac{2000 \sum_{i=1}^{n} X_i}{n} = \frac{2000 X_.}{n}$$

where $X_. = \sum_{i=1}^{n} X_i$ represents the total number of wins over all n students. Consider that over all students we have $5n$ spins of the roulette wheel (trials), the trials still result in one of two outcomes (win or lose), the probability is still constant, and all $5n$ trials are independent. This implies that if the simulator is realistic then the total number of wins over all trials is still binomial:

$$X_. \sim \text{Binomial}(5n, 18/37).$$

Given that $X_.$ can take the values $0, 1, 2, \dots, 5n$ the possible values (the support) of $\bar{W}$ are

$$0, \frac{2000}{n}, \frac{4000}{n}, \dots, 10000.$$

Then for any $\bar{w}$ in the support

$$P(\bar{W} = \bar{w}) = P\left(\frac{2000 X_.}{n} = \bar{w}\right)$$

$$= P\left(X_. = \frac{n\bar{w}}{2000}\right)$$

$$= \binom{5n}{n\bar{w}/2000} \left(\frac{18}{37}\right)^{n\bar{w}/2000} \left(1 - \frac{18}{37}\right)^{5n - n\bar{w}/2000}.$$

E.g., assume that n is even so that the total number of games played, $5n$, is also even. Then the probability that the average balance is FUN 5000 (meaning that exactly half of all games are won) is

$$P(\bar{W} = 5000) = \binom{5n}{2.5n} \left(\frac{18}{37}\right)^{2.5n} \left(1 - \frac{18}{37}\right)^{2.5n}.$$

The exact probability can then be computed by plugging in the true value of n.

b) Note that $\bar{W}$ is a linear transformation of $X_.$. Then

$$E(\bar{W}) = E\left(\frac{2000 X_.}{n}\right)$$

$$= \frac{2000}{n} E(X_.)$$

$$= \frac{2000}{n} \cdot (5n) \frac{18}{37}$$

$$= 4864.86$$

2

and

$$V(\bar{W}) = V\left(\frac{2000X_.}{n}\right)$$
$$= \left(\frac{2000}{n}\right)^2 V\left(X_.\right)$$
$$= \frac{2000^2}{n^2} \cdot n \left(\frac{18}{37}\right)\left(1 - \frac{18}{37}\right)$$
$$= \frac{4996348}{n}.$$

Alternatively, the standard deviation is

$$SD(\bar{W}) = \frac{999.635}{\sqrt{n}}.$$

Note that the expected value does not depend on n. However, the variance decreases with n. This makes intuitive sense, the more data we collect (i.e., the bigger our sample) the more we should learn so the variance should decrease.

c) One way to assess whether or not the simulator is realistic is to compare the average we observed with the sampling distribution computed in part b). There are a couple of ways we can do this.

   i) My belief is that the simulator produces more wins than it should in free mode. We can assess this by computing the probability of getting an average as big as what we observed **or higher**. I.e., we can compute

   $$P(\bar{W} \geq \bar{w}_{\text{obs}})$$

   where $\bar{w}_{\text{obs}}$ represents the value of the average we observed. If this probability is small, then it tells us that value we observed is far out in the tail of the distribution (i.e., far from the centre). This is unlikely to occur by chance, and so it is more likely to believe that the simulator is generating too many wins.

   ii) Alternatively, we could compute how far the observed value is above the mean in standard deviations. Given the mean and variance from part b), the deviation (distance between the observation and its mean) measured in standard deviations is

   $$\frac{\bar{w}_{\text{obs}} - 18000/37}{\sqrt{4996348/n}}.$$

   If this is large, much more than 2, then it again indicates that the observed mean is unlikely to have occured if the simulator is realistic. That is, it is more reasonable to believe that the simulator is producing too many wins.

Exercise 24.1

a) If the simulator is realistic, then counting the total number of spins that it takes for n students to win 5 games each would be the same as counting the number of spins it takes one student to win $5n$ games. Let

$$X_{\cdot} = \sum_{i=1}^{n} X_i.$$

Then

$$X_{\cdot} \sim \text{Neg. Bin.}(5n, 18/37)$$

where we are using the form of the negative binomial that counts the number of trials so that $X_{\cdot}$ can take the values $5n, 5n+1, 5n+2, \dots$. Further

$$\bar{X} = X_{\cdot}/n.$$

Hence, the support of $\bar{X}$ is

$$\{5, 5+1/n, 5+2/n, \dots, \}$$

and for any value $\bar{x}$ in this set

$$P(\bar{X} = \bar{x}) = P\left(X_{\cdot} = n\bar{x}\right) = \binom{n\bar{x}}{5n-1} \left(\frac{18}{37}\right)^{5n} \left(1 - \frac{18}{37}\right)^{n(\bar{x}-5)}.$$

b) Using the formulas for the negative binomial

$$E(X_{\cdot}) = \frac{5n}{18/37} = \frac{185n}{18} = 10.28n$$

and

$$V(X_{\cdot}) = \frac{5n(1 - 18/37)}{(18/37)^2} = 10.849n.$$

Then

$$E(\bar{X}) = 10.28$$

and

$$V(\bar{X}) = \frac{10.849}{n}.$$

c) We can use the observed sample to determine if the simulator is realistic by comparing the observed value of $\bar{X}$, call it $\bar{x}_{\text{obs}}$ with the sampling distribution. If $\bar{x}_{\text{obs}}$ falls far in the tail of the distribution meaning that it is far above or below the mean or that the probability above or below this value is very small, then this provides evidence that the simulator is not realistic. If the observed value is close to the mean or the probability above/below the observed value is close to .50, then this provides no evidence that the simulator is not realistic.