Deep Thoughts

Unit 4: Probability

1. Each of 5 balls in tossed independently and at random into one of 5 bins. Use 3 simulations to estimate:
   (a) The expected number of balls that fall into bin one.

   (b) The probability, p, that all bins end up with exactly 1 ball.

   (c) The probability, q that one bin ends up with 2 balls, another bins ends up with 3 balls, and the remaining bins end up with 0 balls

2. Determine the theoretical value for p/q.

Deep Thoughts
Unit 4: Probability

1. The ELISA tests whether a patient has contracted HIV. The ELISA is said to be positive if it indicates that HIV is present in a blood sample, and the ELISA is said to be negative if it does not indicate that HIV is present in a blood sample. Instead of directly measuring the presence of HIV, the ELISA measures levels of antibodies in the blood that should be elevated if HIV is present. Because of variability in antibody levels among human patients, the ELISA does not always indicate the correct result.

   As part of a training program, staff at a testing lab applied the ELISA to 500 blood samples known to contain HIV. The ELISA was positive for 489 of those blood samples and negative for the other 11 samples. As part of the same training program, the staff also applied the ELISA to 500 other blood samples known to not contain HIV. The ELISA was positive for 37 of those blood samples and negative for the other 463 samples.

   (a) When a new blood sample arrives at the lab, it will be tested to determine whether HIV is present. Using the data from the training program, estimate the probability that the ELISA would be positive when it is applied to a blood sample that does not contain HIV.

   Solution: Essentially correct (E) if the student writes the correct fraction for the estimated probability.

   The estimated probability of a positive ELISA if the blood sample does not have HIV present is

   $$\frac{37}{500} = 0.074$$

   (b) Among the blood samples examined in the training program that provided positive ELISA results for HIV, what proportion actually contained HIV?

   Solution: Essentially correct (E) if the student writes the correct fraction for the proportion or gives a decimal approximation with justification.

   A total of $489 + 37 = 526$ blood samples resulted in a positive ELISA. Of these, 489 samples actually contained HIV. Therefore the proportion of samples that resulted in a positive ELISA that actually contained HIV is

   $$\frac{489}{526} = 0.9297$$

(c) When a blood sample yields a positive ELISA result, two more ELISAs are performed on the same blood sample. If at least one of the two additional ELISAs is positive, the blood sample is subjected to a more expensive and more accurate test to make a definitive determination of whether HIV is present in the sample. Repeated ELISAs on the same sample are generally assumed to be independent. Under the assumption of independence, what is the probability that a new blood sample that comes into the lab will be subjected to the more expensive test if that sample does not contain HIV?

Solution: Essentially correct (E) if the student computes the correct probability, showing work.

$$0.015617552$$

2. Every Monday a local radio station gives coupons away to 50 people who correctly answer a question about a news fact from the previous day's newspaper. The coupons given away are numbered from 1 to 50, with the first person receiving coupon 1, the second person receiving coupon 2, and so on, until all 50 coupons are given away. On the following Saturday, the radio station randomly draws numbers from 1 to 50 and awards cash prizes to the holders of the coupons with these numbers. Numbers continue to be drawn without replacement until the total amount awarded first equals or exceed \$300. If selected, the coupons 1 through 5 each have a cash value of \$200, coupons 6 through 20 each have a cash value of \$1000, and coupons 21 through 50 each have a cash value of \$50.

(a) Explain how you would conduct a simulation using the random number table provided below to estimate the distribution of the number of prize winners each week.

72749	13347	65030	26128	49067	02904	49953	74674	94617	13317
81638	36566	42709	33717	59943	12027	46547	61303	46699	76423
38449	46438	91579	01907	72146	05764	22400	94490	49833	09258

Solution:

1. **Scheme:** Obtain a two-digit random number from the random number table. If it is between 01 and 50, use it to represent the selected ticket. Ignore numbers 00 and 51 – 99.
2. **Stopping Rule:** Determine the amount of the prize associated with the chosen ticket, and add this amount to the total amount awarded so far. If the total amount awarded so far is less than \$300, repeat this process.
3. **Count:** Note the total number of winners.
4. **Non-Replacement:** Ignore any ticket number that has already been awarded a prize in this trial.

Repeats steps 1 – 4 above a large number of times.

Note: It is OK to also devise a scheme that uses 2 two-digit numbers to represent each ticket (for example, 01 and 51 both representing ticket 1; 02 and 52 both representing ticket 2; etc.) that also addresses the issue of assigning 2 two-digit numbers to each coupon correctly.

(b) Perform your simulation 3 times. (That is, run 3 trials of your simulation.) Start at the leftmost digit in the first row of the table and move across. Make your procedure clear so that someone can follow what you did. You must do this by marking directly on or above the table. Report the number of winners in each of your 3 trials.

Solution will depend on answer to part (a).

For example, using scheme above:

Trial 1			Trial 2			Trial 3		
					Total so far			Total so far
72	ignore	0	02	200	200	06	100	100
74	ignore	0	61	ignore	200	70	ignore	100
91	ignore	0	28	50	250	29	50	150
33	50	50	48	50	300	04	200	350
47	50	100						
65	ignore	100	Total number of winners: 3			Total number of winners: 3		
03	200	300						
Total number of winners: 3								

Solution:

Students should perform 3 trials. You will have to look at each student response carefully. Some will continue on 1st row, some will use 2nd row for second trial, etc.

3. A grocery store purchases melons from two distributors, J and K. Distributor J provides melons from organic farms. The distribution of the diameters of the melons from Distributor J is approximately normal with mean 133 millimetres (mm) and standard deviation 5 mm.

   (a) For a melon selected at random from Distributor J, what is the probability that the melon will have a diameter greater than 137 mm?

   Solution: Let X denote the randomly selected melon from Distributor J. X had as approximately normal distribution with mean 133 mm and standard deviation 5 mm.

   $$z = \frac{137 - 133}{5} = 45 = 0.8$$

   $$P(X > 137) = P(Z > 0.8) = 1 - 0.7881 = .2119$$

   Essentially correct if:

   1. Normality and parameters: indicated normal distribution and correctly indicates parameters. Correct components of a z score and labelling mean and standard deviation in calculator picture, or drawing picture clearly indicating mean and standard deviation.
   2. Uses the correct boundary value of 137, or 0.8 and the correct direction.
   3. Reports the correct probability

   (b) Distributor K provides melons from non organic farms the probability is 0.8413 that a melon selected at random from Distributor K will have a diameter greater than 137 mm. For all the melons at the grocery store, 70 percent of the melons are provided by Distributor J and 30 percent are provided by Distributor K. For a melon selected at random from the grocery store, what is the probability that the melon will have a diameter greater than 137 mm?

   Solution:
   Essentially correct if probability is computed correctly AND work is shown that includes correct numerical values using a formula, end results from a tree diagram, or some other appropriate method.

Define events:

J: melon is from Distributor J

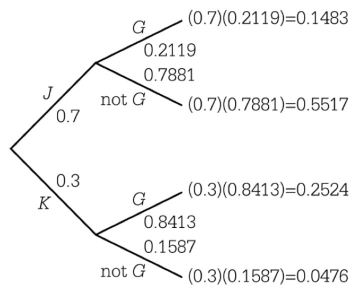
K: melon is from Distributor K

G: melon diameter is greater than 137 mm

For a randomly selected melon from the grocery store,

$$\begin{aligned} P(G) &= P(G \mid J) \times P(J) + P(G \mid K) \times P(K) \\ &= (0.2119)(0.7) + (0.8413)(0.3) \\ &= 0.1483 + 0.2524 \\ &= 0.4007 \end{aligned}$$

OR



From the tree diagram, $P(G) = P(G \text{ and } J) + P(G \text{ and } K) = 0.1483 + 0.2524 = 0.4007$.

(c) Given that a melon selected at random from the grocery store has a diameter greater than 137 mm, what is the probability that the melon will be from Distributor J?

Solution:

Essentially correct (E) if the probability is computed correctly AND work is shown that illustrated how the probability was found.

Using the events defined in part (b), the requested probability is

$$P(J \mid G) = \frac{P(J \text{ and } G)}{P(G)} = \frac{P(G \mid J)\,P(J)}{P(G)} = \frac{(0.2119)(0.7)}{0.4007} = \frac{0.1483}{0.4007} = 0.3701.$$

4. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

   (a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

   Solution: Essentially correct (E) if the response correctly computes the probability of selecting the three women, and shows how the probability was computed.

   $P(\text{All three are women}) = P(\text{First is a woman}) \times P(\text{Second is a women} | \text{First is a woman}) \times P(\text{third is a woman} | \text{first two are women})$

   $$\frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} \approx 0.012$$

   (b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

   Solution: Essentially correct (E) if the response states that the probability from part (a) is small (or insufficiently small), makes an appropriate decision consistent with the probability being small (or insufficiently small), and does so in the context of this situation.

   The probability calculated in part (a) does provide a reason to doubt the manager's claim that the selections were made at random. The calculation shows that there is only about a 1.2% chance that random selection would have resulted in three women being selected. The probability is small enough that it may cast doubt on the manager's claim that the selections were made at random.

   (c) An alternative to calculating the exact probability is to conduct a simulation to estimate the probability. A proposed simulation process is described below.

   Each trial in the simulation consists of rolling three fair, six-sided dice, one die for each of the convention attendees. For each die, rolling a 1, 2, 3, or 4 represents selecting a man; rolling a 5 or 6 represents selecting a woman. After 1,000 trials, the number of times the dice indicate selecting 3 women is recorded.

   Does the proposed process correctly simulate the random selection of 3 women from a group of 9 people consisting of 6 men and 3 women? Explain why or why not.

   Solution: Essentially correct (E) if the response answers no AND states that the dice outcomes in the proposed simulation are independent AND states that the genders of the selected convention attendees are dependent.

   No, the process does not correctly simulate the random selection of three women from a group of nine people of whom six are men and three are women. The random selection of three people among nine is done without replacement. However, in the simulation with the dice, the three dice rolls in any given trial are independent of one another, indicating a selection process that is done with replacement.