**MDM4U Normal Distribution Assignment**
**Jin Hyung Park**

**1. Look at the sketches of continuous probability distributions below.**

For each sketch, give an example of a situation which might give rise to such a probability distribution, fully explaining your reasoning.

a) The right-skewed distribution has a long right tail and is also called a positive-skewed distribution. That's because the number line has a long tail in both directions, and the mean is also on the right side of the beak. This probability distribution curve can occur if there are more people below the median. To explain these inferences and apply them to situations, this may be a probability difference for scores of hard pop quizzes in class. This may also indicate the probability distribution of a randomly chosen person with a given income. Because income distributions generally have quantitative distortions, the probability distributions would look like the chart shown in the question.

b) This is called a constant probability distribution. This probability distribution curve can occur when all events are the same probability. For example, rolling a fair six-sided die produces a probability distribution curve for obtaining a given plane. This is because for a given probability distribution, all possibilities are equal. This is one of the features of rolling fair dice, all possibilities as often as others.

c) This shows that the probability of failure decreases with a typical curve. This sketch shows an extremely large probability at a low value. To explain it, there needs to be a situation in which numbers lower the probability. For example, on the first try, there is the biggest drawback for a penny bang, as the long tail line before the head comes out is increasingly unlikely. My inference is that the most likely outcome is the smallest, each of which is less likely than the previous one.

**2. In many situations, the normal distribution can be used to approximate the binomial distribution.**

a. Explain the conditions in which this can be done, and explain why we might want to take advantage of this property.
To use a normal distribution for an approximate situation, you must consider the number of trials and probabilities. To use a normal distribution for an approximate situation, you must consider the number of trials and probabilities. You can do this if you have enough trials to provide more than five bars that provide the required look.

The expected value must also be at least 5 less than the number of trials. Therefore, np >= 5 is required to determine whether it is appropriate to use a normal distribution, where n is the number of trials, p is the probability, and n (1-p) >= 5. This is useful because in many situations it is much faster and easier to use a normal distribution, especially when the number of trials is high.

b. Give an example of a situation in which we could do this.
For example, we can do this when we want to know the probability of a flight being delayed 20 times in August, when the probability of a flight being delayed is 0.6.

c. Give an example of a situation in which we would not be able to make this approximation and explain why.
We could not use a normal distribution to estimate the probability that the flight would be delayed three times a week at 0.3 probability. This means that the expected value is less than 5, meaning that the expected curve is not equal to the curve of the normal distribution.

**3. A species of alien has a mean height of 23 cm and a standard deviation of 3.6 cm. What is the probability that an alien chosen at random has a height of more than 20cm?**

The Z-score tells you how far you are from the mean measured by the standard deviation. An alien with a height of 20 cm is -0.5 (Z-score) away from its mean. If you look at the z-scores table, the z-score corresponds to a probability of 0.83. Therefore, a randomly selected alien has a 83% chance of being less than 20cm high. Since we are now calculating the probability that an alien at random has a height greater than 20 cm, we must subtract the value from 1. ( $1 - 0.83 = about\ 0.17$ )

Thus, the probability that an alien selected at random has a height greater than 20cm is about 17%.

- To calculate the Z-Score:
  - $Z - Score\ Formula:\ z(x) = \frac{x-\mu}{\alpha}$
  - $z(x) = \frac{20-23}{3.6} =- 0.83$

**4. Researchers have observed that regular smokers have an average lifespan that is normally distributed and is 68 years with a standard deviation of 10 years. What percent of smokers will live beyond age 76?**

The 76-year-old average smoker is 0.80 z-score away from the average. If you look at the z-scores table, the z-score corresponds to a probability of 0.7881. Therefore, there is a 78.81 percent chance that ordinary smokers will live under the age of 76. However, since we are calculating the probability of a smoker living beyond the age of 76, we need to subtract the value from 1.

- $1 - 0.7881 = 0.2119$

Thus, the probability that a regular smoker will live longer than age 76 is about 0.21%.

**5. The lifespan of a particular species of turtle is normally distributed with a mean of 180 years and a standard deviation of 40 years. What is the probability that one of these turtles will live more than a century?**

A 100-year-old turtle is -2.0 z-score from the average. If you look at the z-scores table, the z-score corresponds to a 0.228 probability. So the probability that a randomly chosen turtle will live less than 100 years is 2.28%.

However, you have to subtract the value from 1 because a randomly chosen turtle is calculating the probability of living for more than 100 years. ( $1 - 0.228 = 0.9772$ )

Therefore, there is a 97.72% chance that turtles will live more than 100 years.

- Calculating the Z-score:
  - $Z - Score\ Formula:\ z(x) = \frac{x-\mu}{\alpha}$
  - $z(x) = \frac{100-180}{40} =- 2.0$

**6. A second species of alien has a mean height of 71 cm and a standard deviation of 5.3 cm. An alientologist discovers that 30% of the aliens bump their heads into their spaceship. What is the height of the spaceship door?**

Because 30 percent of aliens hit their heads when they board a spacecraft, we can estimate that 30 percent are among the highest groups of very tall aliens. We can also assume that the remaining 70 percent passes.

If you look at the Z-scores table, the Z-score, which has the closest probability of 0.70, is 0.52, with a corresponding probability of 0.6895 and an error range of 0.0015. This Z-score can be used to resolve 'x' in the Z-score formula, an observation that determines the height of the spacecraft door. Therefore, the door is 74 cm high and rounded to the nearest integer.

- Calculating the height of the door:
  - $Z - Score\ Formula:\ z(x) = \frac{x-\mu}{\alpha}$
  - $x = (\sigma * z(x)) + \mu$
  - $x = (5.3 * 0.52) + 71$
  - $x = 73.756$

**7. In Bayfield, 65% of residents read the Bayfield Breeze, a local online blog. Dennis wants to know what people think of the blog, so he stops 40 people on the street to ask them if they read it.**

a. Verify that the normal distribution can be used to approximate this situation.

- In order to use normal distribution, we need to do the following:
  - $np \geq 5\ ,\ n(1-p) \geq 5$

- Thus,
  - $40(0.65) \geq 5$
  - $26 \geq 5$
  - $50(1-0.65) \geq 5$
  - $17.5$

Therefore, since the conditions have been met, one can use a normal distribution to give an overview of this situation.

b. What is the mean and standard deviation of the number of people he finds that read the Breeze?

- Mean
  - $Formula\ for\ Mean = \mu = np$
  - To switch from binomial to normal distribution
    - $\mu = (40)(0.65)$
    - $\mu = 26$

- Standard Deviation
  - $Formula\ for\ Standard\ Deviation\ = \sigma = \sqrt{np(1-p)}$
  - To switch from binomial to normal distribution
    - $\sigma = \sqrt{(40)(0.65)(1-0.65)}$
    - $\sigma = \frac{\sqrt{910}}{10} or\ 3.016$

- Thus, the mean and standard deviation is 26 and 3.016 respectively.

c. What is the probability that at least 25 of the people he asks read the blog?

- Calculating the Z-Score is the following:
  - $Z - Score\ Formula : z(x) = \frac{x-\mu}{\alpha}$
  - $z(x) = \frac{25-26}{3.016} = \frac{-1}{3.016} =- 0.3315$
- If you look at the z-scores table, the z-score -0.3315 corresponds to the 0.37 probability. So less than 25 out of 40 people surveyed by Dennis have a 37% chance of reading Breeze.
- However, we are calculating the probability that at least 25 of the people who read the blog, we need to subtract the value from 1.
  - $1 - 0.37 = 0.63\%$
- Therefore, the probability that at least 25 people who read the Breeze is 63%.

**8. Yuen Zhi is running a ring toss event at a school fair. There is a 15% chance that each attempt wins a prize. She has 45 prizes and believes 250 people attempt the event. She is worried she won't have enough prizes. Can you re-assure her she will probably be ok?**

In order to decide whether Yuen Zhi will be able to receive enough prize money for the event, we have to figure out the odds of winning more than 45 people. Because there are more than 45 awards. Before using a normal distribution instead of a binomial distribution, make sure you have enough space to create a bell-shaped curve of the normal distribution.

A minimum of 5 bars are required to the left and right of the peak for the binomial distribution to resemble the normal distribution. Therefore, because the conditions have been met, you can use a normal distribution to get an overview of this situation.

- To get the qualification for exploiting the normal distribution:
  - $np \geq 5$
  - $(250)(0.15) \geq 5$
  - $37.5 \geq 5 \;==\; TRUE$

  - $n(1-p) \geq 5$
  - $(250)(1-0.15) \geq 5$
  - $212.5 == TRUE$
- Calculate the mean number of people who will win prizes under the 15% chances:
  - $Formula\,for\,Mean: \mu = np$
  - $\mu = (250)(0.15) = 37.5$
- Calculate the Z-Score:
  - $Z - Score\,Formula: \; z(x) = \frac{x-\mu}{\sigma}$
  - $z(x) = \frac{45.5-37.5}{5.6457} = 1.416$

When checking the z-score table, a Z-Score of 1.42 matches with the probability of 0.9222. Thus, the probability that at least 45 people will win prizes is 92.22%. However, we are calculating the probability of more than 45 people winning prizes, we need to subtract the value from 1. To do so, $(1-0.9222 = 0.0778)$ is the answer.

Therefore, the probability that more than 45 people will win prizes is 7.78%, which means that Yuen Zin does not need to worry about prizes for her event.

**9. We have been using the normal distribution to approximate situations that are in fact binomial events.**

a. Demonstrate how accurate the approximation is by using both approaches to find the probability of the same event.

Let Jin - me - flip the coin in 50 times. To get the probability that the coin lands on its head at least 15 times but not exceeds 20 times. Let us use two approaches, binomial distribution and normal distribution.

- **Using binomial distribution**: $P(x\;times\;of\;success) = {}_nC_x * p^x * (1-p)^{(n-x)}$

- $P(15\ times\ for\ heads\ to\ be\ landed) = {}_{50}C_{15} * (1/2)^{15} * (1-(1/2))^{(50-15)} = 0.001999...$
- $P(16\ times\ for\ heads\ to\ be\ landed) = {}_{50}C_{16} * (1/2)^{16} * (1-(1/2))^{(50-16)} = 0.004373...$
- $P(17\ times\ for\ heads\ to\ be\ landed) = {}_{50}C_{17} * (1/2)^{17} * (1-(1/2))^{(50-17)} = 0.008746...$
- $P(18\ times\ for\ heads\ to\ be\ landed) = {}_{50}C_{18} * (1/2)^{18} * (1-(1/2))^{(50-18)} = 0.016034...$
- $P(19\ times\ for\ heads\ to\ be\ landed) = {}_{50}C_{19} * (1/2)^{19} * (1-(1/2))^{(50-19)} = 0.027005...$
- $P(15 \leq x < 20) = 0.001999... + 0.004373 + 0.008746 + 0.016034 + 0.027005 = 0.058159$

- Thus, the probability that the coin flipped 50 times will land on its head at least 50 times but not exceed 20 times is about 0.0582 which is 5.82%.

- **Using normal distribution**
  - In order to utilise the normal distribution, required to check the following:
    - $np \geq 5$
      - $(50)(0.5) \geq 5$
    - $n(1-p) \geq 5$
      - $(50)(1-0.5) \geq 5$
  - Calculating the mean:
    - $Formula\ for\ Mean : \mu = np$
    - $\mu = (50)(0.5) = 25$
  - Calculating the standard deviation:
    - $Formula\ for\ Standard\ Deviation\ = \sigma = \sqrt{np(1-p)}$
    - $\sigma = \sqrt{25(1-0.5)} = \frac{5\sqrt{2}}{2} = 3.5355...$
  - Calculating the Z-Score:
    - $Z - Score\ Formula :\ z(x) = \frac{x-\mu}{\sigma}$
    - To get the probability for 15 times of flipping coin landing on its head is the following:
      - $z(15) = \frac{15-20}{3.5355} = \frac{-5}{3.5355} = -0.7071$
      - When checking the z-score table, a z-score of -0.70 corresponds with a probability of 0.2419.
    - To get the probability for 20 times of flipping coin landing on its head is the following:
      - $z(20) = \frac{20-20}{3.5355} = \frac{0}{3.5355} = 0$
      - When checking the z-score table, a z-score of 0.02 corresponds with a probability of 0.4920.
    - $0.2419 - 0 = about\ 0.2419$
    - Therefore, the probability that the coin flipped 50 times which lands on its heads at least 15 times but not exceeds 20 times is about 24%.
- Conclusion: In ideally speaking, the binomial distribution would work but for observing the rough scheme of an experiment, normal distribution can become one of the candidates for saving the time and cost.

b. Describe the conditions under which the normal would give a less accurate approximation.

Theoretically, assuming that a normal distribution can be used instead of a binomial distribution, a binomial condition should use a binomial distribution if the result values are not that much greater than 5 even if np and n (1-p) are still greater than or equal to 5. Although both $np$ and $n(1-p)$ are greater than 5, they are not that much greater than 5 to be qualified for approximating a binomial distribution with a normal distribution much better.

<u>c. Explain a situation in which the criteria for using the approximation would be met, ie. $np \geq 5$ and $n(1 - p) \geq 5$, and yet you would decide not to use the normal distribution.</u>

Theoretically, suppose a normal distribution can be used instead of a binomial distribution. Even if np and n (1-p) are still greater than or equal to 5, a binomial distribution must be used if the result value is not significantly greater than 5. For example, if you can calculate the probability with just 10 calculations, it would be much better to use a binomial distribution to obtain accurate information about the probabilities. If you are calculating the probability of throwing a coin with 40 of the 50 tosses on the head, you only need to calculate the probability in binomial distribution twice.