Connections to Discrete Random Variables

Lesson objectives

- I can make connections between a normal distribution and a binomial distribution
- I can make connections between a normal distribution and a hypergeometric distribution
- I can recognize the role of the number of trials in these connections

1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 370 #s 1 - 6 & 8 - 11

Warm up

Recall that tossing a coin several times and recording the number of heads obtained is an example of a binomial distribution.

- How does the shape of the distribution depend on the number of times the experiment is tried?
- Predict the form of the graph of the number of heads possible when a coin is flipped five times. Make a sketch of your prediction.

Definition

Continuity Correction

- A correction applied when using the normal approximation to correct for the difference between a discrete and continuous distribution

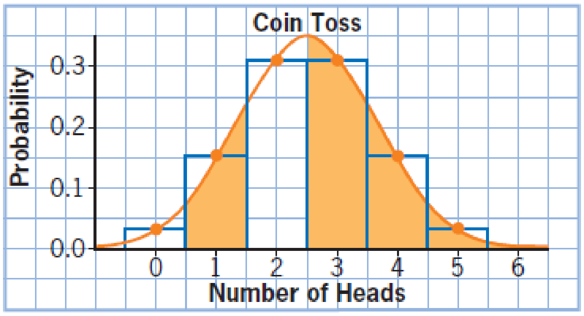
You can calculate a binomial probability using the formula from chapter 4. However, this formula requires the use of factorials. If the number of successes being considered is large, this can lead to a large number of computations. You can avoid this by using the approximation.

The normal approximation for a binomial distribution is usually considered reasonable if the values of np and nq are both greater than 5: $np > 5$ and $nq > 5$.

The normal approximation for a hypergeometric distribution is usually considered reasonable if the sample size is small compared to the size of the population, typically less than one-tenth: $n < \frac{1}{10}NP$, where n is the sample size and NP is the size of the population.

Since the binomial and hypergeometric distributions are discrete but the normal distribution is continuous, you must apply a **continuity correction** when using the approximation.

For example, suppose you want to determine the probability that, in 5 tosses of a coin, there is exactly 1 head. Refer to the graph shown.



The area under the normal distribution that represents this probability runs from 0.5 to 1.5. Therefore, you must calculate $P(0.5 \le X \le 1.5)$.

Similarly, if you want the probability of getting 3 or more heads, you need the area under the graph from 2.5 to infinity on the right, or $P(X \ge 2.5)$.

On the other hand, if you want to determine the probability of getting more than 3 heads, you need to calculate $P(X \ge 3.5)$.

Example 1

Normal Approximation for a Binomial Distribution

Suzette rolls a die 36 times. She records the number of times the die shows a 6.

a) Is it reasonable to approximate this distribution with a normal distribution? Give a reason for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) Determine the probability that the die will show a 6 at least 10 times.

d) What is the probability that the die will show a 6 more than 10 times, fewer than 10 times, and at most 10 times?

a) Calculate np and nq

np = 36(1/6)   nq = 36(5/6)

np = 6   nq = 30

As both are greater than 5, it is reasonable to use a normal approximation.

b) $\mu = np$

$= 36 \times \frac{1}{6}$

≈ 6

$\sigma = \sqrt{npq}$

$= \sqrt{36 \times \frac{1}{6} \times \frac{5}{6}}$

≈ 2.236

The mean is 6 and the standard deviation is 2.236.

c) Since we want the die to show 6 at least 10 times, we must apply the continuity correction (because this is discrete data) and find P(X ≥ 9.5).

First find the z-score for 9.5, then find P(X ≥ 9.5)

$z = \frac{X - \mu}{\sigma}$

$\approx \frac{9.5 - 6}{2.236}$

≈ 1.565

P(X ≥ 9.5) = 1 - P(x < 9.5)

= 1 - P(z < 1.57)

= 1 - 0.9418

= 0.0582

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418

d) We need to find P(X ≥ 10.5), P(X ≤ 9.5), and P(X ≤ 10.5).

Find the z-scores for each and then use the table to calculate the probabilities.

z = $\frac{10.5 - 6}{2.236}$

= 2.01

z = $\frac{9.5 - 6}{2.236}$

= 1.565

z = $\frac{10.5 - 6}{2.236}$

= 2.01

P(X ≥ 10.5) = 1 - P(x < 10.5)

= 1 - P(z < 2.01)

= 1 - 0.9778

= 0.0222

P(X ≤ 9.5) = P(x < 9.5)

= P(z < 1.565)

= 0.9418

P(X ≤ 10.5) = P(x < 10.5)

= P(z < 2.01)

= 0.9778

Your Turn

The probability of rolling a 6 on a weighted die is 0.25. The die is rolled 25 times.

a) Is it reasonable to approximate this distribution with a normal distribution? Give reasons for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) Determine the probability that the die will show a 6 fewer than 8 times.

n = 25, p = 0.25, q = 0.75

a) Calculate np and nq

np = 25(0.25)   nq = 25(0.75)

np = 6.25   nq = 18.75

As both are greater than 5 it is reasonable to use a normal approximation.

b) $\mu = np = 25(0.25) = 6.25$ $\sigma = \sqrt{npq} = \sqrt{25(0.25)(0.75)} \approx 2.165$

The mean is 6.25 and the standard deviation is about 2.165.

c) For fewer than 8 successes we need P(X ≤ 7.5). Find the z-score and then use the table.

$$z = \frac{7.5 - 6.25}{2.165} = 0.577$$

P(X ≤ 7.5) = P(x < 7.5)
= P(z < 0.577)
= 0.7190

The probability of a 6 fewer than 8 times is about 0.7190

Example 2

Normal Approximation for a Hypergeometric Distribution

Chris works at a local daycare on a co-op work term. Chris plays a game with the children that involves pulling marbles from a bag. The bag contains 24 black marbles and 36 red marbles, well mixed. One of the children reaches in and takes out 5 marbles without looking. Chris records the number of black marbles.

a) Is it reasonable to approximate this distribution with a normal distribution? Give a reason for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) What is the probability that exactly 3 of the marbles are black?

d) What is the probability that a child pulls out at least 3 black marbles? more than 3 black marbles? fewer than 3 black marbles? at most 3 black marbles?

e) How does the answer to part c) compare with the probability calculated from the hypergeometric distribution?

These marbles are not replaced so this is a hypergeometric distribution.

60 marbles in the bag, 5 are chosen.

a) For a normal approximation to be reasonable the sample should be less than 10% of the population. As 5 < 0.1(60) this would be reasonable.

b) $\mu = np = 5 \times \frac{24}{60} = 2$

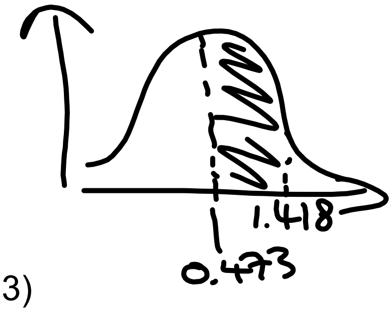
$$\sigma = \sqrt{npq\left(\frac{NP - n}{NP - 1}\right)} = \sqrt{5 \times \frac{24}{60} \times \frac{36}{60}\left(\frac{60 - 5}{60 - 1}\right)} \approx 1.058$$

The mean is 2, and the standard deviation is 1.058.

c) For exactly 3 black marbles we need P(2.5 ≤ X ≤ 3.5). Find the z-score for each value and then use the table.

$z = \frac{3.5 - 2}{1.058} = 1.418$

$z = \frac{2.5 - 2}{1.058} = 0.473$



P(X ≤ 3.5) = P(x ≤ 3.5)
= P(z ≤ 1.418)
= 0.9222

P(X ≤ 2.5) = P(x ≤ 2.5)
= P(z ≤ 0.473)
= 0.6808

P(2.5 ≤ X ≤ 3.5) = P(x ≤ 3.5) - P(x ≤ 2.5)
= 0.9222 - 0.6808
= 0.2414

The probability of getting exactly 3 black marbles is about 0.2414.

e) The normal approximation gives an answer of about 0.2414 where as the combinatoric method gives an answer of about 0.2335. They are similar in value.

$$P(X = 3) = \frac{{}_{24}C_3 \times {}_{36}C_2}{{}_{60}C_5} \approx 0.2335$$

d) We need to find P(X ≥ 2.5), P(X ≥ 3.5) P(X ≤ 2.5), and P(X ≤ 3.5).

Find the z-scores for each value and then use the table.

$z = \frac{2.5 - 2}{1.058} = 0.473$

P(X ≥ 2.5) = 1 - P(x ≤ 2.5)
= 1 - P(z ≤ 0.473)
= 1 - 0.6808 = 0.3192

The probability of getting at least 3 blacks is about 0.3192

$z = \frac{3.5 - 2}{1.058} = 1.418$

P(X ≥ 3.5) = 1 - P(x ≤ 3.5)
= 1 - P(z ≤ 1.418)
= 1 - 0.9222 = 0.0778

The probability of getting more than 3 blacks is about 0.0778

$z = \frac{2.5 - 2}{1.058} = 0.473$

P(X ≤ 2.5) = P(x ≤ 2.5)
= P(z ≤ 0.473)
= 0.6808

The probability of getting fewer than 3 blacks is about 0.6808

$z = \frac{3.5 - 2}{1.058} = 1.418$

P(X ≤ 3.5) = P(x ≤ 3.5)
= P(z ≤ 1.418)
= 0.9222

The probability of getting at most 3 blacks is about 0.9222

Your Turn

Allison has a drawer full of unmatched socks. The drawer contains 30 blue socks, 30 green socks, and 30 yellow socks. She pulls seven socks from the drawer and records the number of blue socks in the sample.

a) Is it reasonable to approximate this distribution with a normal distribution? Give a reason for your answer.

b) Determine the mean and standard deviation of the normal approximation.

c) What is the probability that 3, 4, or 5 of the socks are blue?

These socks are not replaced so this is a hypergeometric distribution.

90 socks in the draw, 7 are chosen.

a) For a normal approximation to be reasonable the sample should be less than 10% of the population. As 7 < 0.1(90) this would be reasonable.

b) $\mu = np$

$= 7\left(\frac{30}{90}\right)$

≈ 2.333

$\sigma = \sqrt{npq\left(\frac{NP-n}{NP-1}\right)}$

$= \sqrt{7\left(\frac{30}{90}\right)\left(\frac{60}{90}\right)\left(\frac{90-7}{90-1}\right)}$

≈ 1.204

The mean is about 2.333 and the standard deviation is about 1.204

c) For there to be 3, 4, or 5 blue socks find P(2.5 ≤ 5.5). First find the z-scores and then use the table.

$z = \frac{5.5 - 2.33}{1.204}$

$= 2.633$

$z = \frac{2.5 - 2.33}{1.204}$

$= 0.141$

P(X ≤ 5.5) = P(x ≤ 5.5)

= P(z ≤ 2.633)

= 0.9961

P(X ≤ 2.5) = P(x ≤ 2.5)

= P(z ≤ 0.141)

= 0.5557

P(2.5 ≤ X ≤ 5.5)

= 0.9961 - 0.5557

= 0.4404

The probability that 3, 4, or 5 of the socks are blue is about 0.4404

Key Concepts

- As the number of trials increases, a binomial distribution takes on the characteristics of a normal distribution.
- If the values of np and nq are both greater than 5, you can approximate the binomial distribution using a normal distribution.
- If the sample size is small compared to the population size, a hypergeometric distribution takes on the characteristics of a normal distribution.
- If the sample size is less than one-tenth of the population size, $n < \frac{1}{10} NP$, you can approximate the hypergeometric distribution using a normal distribution.
- You must use a continuity correction when approximating a discrete distribution with a normal distribution. For example, if you want the probability of rolling a 6 exactly 3 times, you must calculate $P(2.5 \le X \le 3.5)$. If you want the probability of rolling at least 3 sixes, you must calculate $P(X \ge 2.5)$. However, if you want the probability of rolling more than 3 sixes, you must calculate $P(X \ge 3.5)$.

Binomial distributions

$\mu = np \qquad \sigma = \sqrt{npq}$

Hypergeometric distributions

$\mu = np \qquad \sigma = \sqrt{npq\left(\frac{NP-n}{NP-1}\right)}$

n = sample size

NP = population size

p = probability of success

q = probability of failure

R1. Suggest another situation that is usually modelled with a binomial distribution but could reasonably be approximated with a normal distribution.

Any situation that has only two outcomes (success, p and failure, q). There also needs to be enough **INDEPENDENT** trials such $np > 5$ **AND** $nq > 5$. Eg Number of sixes rolled on a die.

R2. Suggest another situation that is usually modelled with a hypergeometric distribution but could reasonably be approximated with a normal distribution.

Any situation that has only two outcomes (success, p and failure, q). The number of **DEPENDENT** trials also needs to be less than 10% of the population. Eg Choosing players for a hockey team.

R3. Suggest possible reasons why you might prefer to use a normal approximation rather than a binomial or hypergeometric distribution.

It is easier to calculate the probabilities for a range of values using the approximation than it is using the binomial or hypergeometric formulas.

R4. Describe a situation where it is not possible to use a normal approximation for the binomial or hypergeometric distribution.

Any situation for a binomial distribution where $np \leq 5$ **OR** $nq \leq 5$. Any situation for a hypergeometric distribution where the number of dependent trials is greater than 10% of the population.