

## 1 Section 7.7: Normal approximation to the Binomial Distribution

Let's start with some more examples of normal distribution.

Example: Copperhead snakes have lengths which are normally distributed with mean  $\mu = 70$  cm and standard deviation of 8 cm. Suppose a random copperhead is captured and measured. Find the following probabilities:

1. The snake is less than 70 cm
2. The snake is more than 76 cm
3. the snake is between 60 and 72 cm
4. Find the 80th percentile for copperhead sizes

When we look at the probability histograms for a binomial histogram we note that many of these are bell shaped. From last section we know that bell shaped means normal distribution. It turns out we have the following result relating the two.

If  $X$  is a binomial distribution with parameters  $n$  and  $p$  then the probability histogram for  $X$  can be approximated by the normal curve with  $\mu = np$  and  $\sigma = \sqrt{npq}$ . This approximation is very good when  $np > 5$  and  $nq > 5$ .

How does this help us? Binomial probabilities can be a pain to compute because of the high powers and binomial coefficients. But normal probabilities are easy because we can look them up in a table. There is one difference in how we look at the probabilities.

Suppose  $X$  is a binomial distribution,  $Y$  is the associated normal distribution, and we want to approximate  $P(X = 4)$ . How can we do this with a normal distribution? We can't just say  $P(Y = 4)$  because that will always be 0. So instead we have to change our event slightly.

Think in terms of height. How tall are you? Are you exactly that tall? What range of values could you actually be in to say that height? We have to do the same thing since the binomial distribution is discrete and the normal distribution is continuous. If  $X$  is a binomial distribution and  $Y$  is the associated normal distribution, then we turn  $P(X = 4)$  into  $P(3.5 < Y < 4.5)$ . This is called continuity correction. We also need to be more careful about strict and non-strict inequalities. Let's do some examples.

Example: We know that 10% of UVA students take some sort of math class. We decide to select 100 student at random and count how many are taking a math class. Approximate the following probabilities with a normal distribution:

1. Probability exactly 11 are taking a math class
2. Probability more than 15 are taking a math class
3. Probability between 5 and 10 (inclusive) are taking a math class
4. Probability less than 7 are taking a math class.

Before we do any approximations we need to find our normal distribution and make sure it is valid. The normal distribution we use to approximate should have mean  $\mu = np = 10$  and  $\sigma = \sqrt{npq} = \sqrt{100 \cdot 0.1 \cdot 0.9} = 3$ . We can also check  $np = 10$  and  $nq = 90$  so our approximation is valid. Now we can begin with our approximation.

For our first approximation is the simple continuity correction. If  $Y$  is distributed normally with mean 10 and standard deviation 3 then we get

$$P(X = 11) \approx P(10.5 < Y < 11.5) = P\left(\frac{10.5 - 10}{3} < Z < \frac{11.5 - 10}{3}\right) = P(0.17 < Z < 0.5) = 0.6915 - 0.5675 = 0.124$$

The next example is where we need to start being a little careful. We want more than 15. Thus we should not include the number 15 in our calculation. This gives

$$P(X > 15) \approx P(Y > 15.5) = 1 - P(Z \leq 1.83) = 1 - 0.9664 = 0.0336$$

We have to be careful again with the next example along the same lines. This one is inclusive so we need 5 and 10 to be in our computation.

$$P(5 \leq X \leq 10) \approx P(4.5 \leq Y \leq 10.5) = P(-2.17 \leq Z \leq 0.17) = 0.5675 - 0.0150 = 0.5525$$

The last example we do the same thing but we should not include 7 in our range

$$P(X < 7) \approx P(Y < 6.5) = P(Z < -1.5) = 0.0668$$

Let's just check we aren't crazy. For the first example we could compute exactly

$$P(X = 11) = \binom{100}{11} (0.1)^{11} (0.9)^{89} = 0.1198$$

which is within 5 hundredths with very little work.

The other nice thing is that this gets more accurate the bigger  $n$  is which is exactly when exact probabilities are harder to compute.