Chapter 5

A Tale of Two Distributions

Again, it is important to keep in mind where we are headed. We want to be able to compare data to a probabilistic model. There are two extremely useful distributions which we can use as prototypes for many, many probabilistic models: the binomial distribution and the normal distribution.

Two things make these two distributions very useful. The first is that they apply to so many situations. The second is that their properties are relatively easy to calculate. This second point is very important. Think about the effort involved in finding the mean and standard deviation of a random variable that has many possible values. (Consider Exercise 4.3, for example.) If we can get a couple distributions for which we can tabulate all the properties, then we are in great shape as long as we can describe the random variable we want in terms of these distributions.

And that is exactly what we can do. Sometimes, we can restate our question in such a way that the binomial distribution gives us the answer. Other times we can stretch (or squish) and shift our distribution to compare it to the known properties of the normal distribution. There are things to beware of, but the first step is to become very familiar with these two distributions.

5.1 Vocabulary

Make sure you understand how each of the following terms relate to binomial or normal distributions

- binomial random variable
- probability of success
- independence
- binomial coefficient
- smooth
- symmetrical
- bell-shaped
- z-transform

5.2 Equations, specific rules, etc.

The following formulas demonstrate how easy it is to calculate the properties of binomial and normal distributions:

- The mean and variance of the binomial distribution depend only on the number of trials n and the probability of success p.

$$\mu = np$$

$$\sigma = np(1-p)$$

- The probability of any particular value in a binomial distribution is also relatively easy:

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  This function is tabulated in many books and can be computed by many computer software packages.

- The mean and standard deviation of the standard normal distribution are even easier! They are just 0 and 1, respectively!
- To transform any normal distribution to the standard normal, we can just stretch and shift it by its own mean and standard deviation. This is called the z-transform where

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

- The probability of any particular range of values in a normal distribution is tabulated in many books and can be computed by many software packages.

5.3 Specific notes

- pp. 74-78: These pages provide some examples of situations where the binomial distribution can be used. The key here is to recognize that there are many questions that can be restated in terms of success and failure. If you can do this – that is, if you can state your situation in such a way that there are only two outcomes – then you can use the binomial distribution.
- p. 77: Pascal's triangle is a very useful thing. Once you get used to what it means, it is easy to use. And it is easy to create. As the page says, each number is just the sum of the two above it.
- pp. 79-83: Although the information in these pages about the standard normal distribution is important, what you really need to know is how to use the standard normal distribution and how to transform other distributions to the standard normal through the z-transform. Focus on pp. 84-85 for this.

- p. 87: The continuity correction is easier to see if you think about it as choosing the z that corresponds to the upper boundary of your bin. So

$$z = \frac{(x + 0.5) - \mu}{\sigma}$$

5.4 Exercises

Exercise 5.1: You roll one six-sided die 3 times. Sketch the frequency distribution (or present it as a table) for the situation where rolling a 'one' is considered success and everything else is considered failure.

Exercise 5.2: You roll one six-sided die 3 times. Sketch the frequency distribution (or present it as a table) for the situation where rolling either a 'one' or a 'two' is considered success and everything else is considered failure.

Exercise 5.3: You roll one six-sided die 10 times. What is the probability of rolling a 'one' exactly 2 times? at least 2 times?

Exercise 5.4: You take an HIV test twice. The test has a false positive rate of 1%. What is the probability that you test positive both times even though you are HIV negative? What is the probability that you test positive on one of the two tests even though you are HIV negative? Calculate these probabilities both by using the binomial distribution and by simply combining the known probabilities.

Exercise 5.5: Use the table on p. 84 to determine the probability of rolling two six-sided dice and having their sum be 6 or less. Hint: You will need to use the μ and σ for the sum of two six-sided dice that you calculated in Exercise 4.1. Do the calculation with and without the continuity correction, and compare it to the actual probability you calculate from the data on p. 55.
