Chapter 4

Random Variables

In both this chapter and the next, it is important to keep in mind where we are headed. What we eventually want to be able to do is to compare data to a probabilistic model. A probabilistic model doesn't make certain predictions but rather predicts that certain events occur with specific frequency or probability.

In order to be able to compare data to a probabilistic model, one thing we must do is describe the predictions of a probabilistic model. And we would like to use the same descriptors that we have already found to be useful in describing data – the mean and the standard deviation. These descriptors are easy to define for the predictions of a probabilistic model, but it is important to keep in mind that they are different in some sense from a regular mean and standard deviation. For data, the mean and standard deviation are "real" in the sense that the data are already set. For predictions, there are no data. There are only expected values. So the mean of a random variable is the average of a "typical" set of data. Keeping this distinction between "real" and "typical" is important. That is, in fact, one of the reasons why different symbols are used for the mean of real data ($\bar{x}$) and the mean of a random variable (μ).

The text emphasizes this distinction in its discussion of sample properties (the "real" data) and model or population properties (which are based on "typical" data).

4.1 Vocabulary

Make sure you understand each of the following terms well enough to use them appropriately when discussing situations which involve probability: For terms like mean, variance, and standard deviation, make sure you understand the difference between their use in describing models and their use in describing data.

- random experiment
- random variable
- sample properties
- model (or population) properties
- mean
- variance
- standard deviation
- expected (or expectation) value
- continuous random variable

4.2 Symbols

You should know that μ denotes the mean of a random variable and σ denotes the standard deviation of a random variable. This is in contrast to $\bar{x}$ and s which are the mean and standard deviation of sample data.

In addition, $p(x)$ is the probability of a random variable having the value x.

4.3 Equations, specific rules, etc.

This chapter has several formulas you should be able to use.

- To get the mean of a random variable, you just add up the products of each possible value of the variable times the probability of that value.

$$\mu = \sum_i x_i p(x_i)$$

.

- Similarly, the variance of a random variable is found by adding up the products of the squared distance of each possible value from the mean times the probability of that value.

$$\sigma^2 = \sum_i (x_i - \mu)p(x_i) =$$

.

- The standard deviation is just the square root of the variance.
- Stretching and shifting distributions stretches and shifts the mean and standard deviation in a predictable way:

$$\mu(aX + b) = a\mu(X) + b$$
$$\sigma(aX + b) = |a|\sigma(X)$$

where $|a|$ is the absolute value of a

4.4 Specific notes

- p. 53 Pay attention to the summaries of chapters 2 and 3. The summaries show what is important as a foundation for this and later chapters.

- pp. 53-58: These pages are just meant to provide examples of random variables and to make clear the distinction between "real" sample data and "typical" model data.

- p. 57: It may seem that the relative frequency histogram and the probability histogram are the same, but again they are distinct on the basis of one being the sample data (the relative frequency histogram) and the other being the model (the probability histogram). We want to compare one to the other.

- pp.59-62: These pages show you the formulas for the mean and variance of a random variable by using the sample mean and sample variance as an analogy. Another way to see it is to write out an actual "typical" data set for the random variable. For example, for the sum of two dice on p. 55, a sample set of 36 data points would be

  $\{2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 11, 11, 12\}$

  Can you see why this qualifies as a typical data set? Given this data set, it is easy to find the mean and standard deviation. But it is also clear that you can simplify the calculations by noting the frequency of each value. That is where $p(x)$ comes into the formula.

- pp. 63-67: If you have not had calculus, this section on continuous random variables will not make much sense. It is not crucial for understanding the concepts of this or future chapters, so feel free to skim it. However, you should know the difference between discrete and continuous random variables. Discrete variables can only take on one of a specific set of values. Continuous variables can take on an infinite number of different values.

- pp. 68-72: This section on adding random variables is not too difficult, but it may seem more abstract than it needs to be. The main point is that if you shift the distribution of a random variable by adding b to each possible outcome, the mean of the distribution shifts by b and the variance doesn't change. And if you stretch (or squish) the distribution by multiplying each each possible outcome by a, the variance is stretched (or squished) by a factor of a^2 and the mean is shifted by a factor of a. Putting these two together yields the formulas for the mean and variance of a stretched and shifted distributions in terms of the mean and variance of the original distribution.

4.5 Exercises

Exercise 4.1: Calculate μ and σ for the sum of two six-sided dice. (see p. 55).

Exercise 4.2: Here's a simple game with two six-sided dice. You pay me \$25 per roll to roll the dice with the agreement that I pay you \$3 for every dot that comes up. So if you roll double-sixes, I pay you \$36. (Does it make sense that your winnings are defined by the random variable $3Y - 25$ where Y is the sum of the two dice?)
Sketch the frequency distribution of your winnings (or fill out a table like that on p. 55), and report the expected mean and standard deviation of your winnings.

Exercise 4.3: Calculate μ and σ for the sum of three six-sided dice.
