Chapter 1

Introduction

1.1 What Is Probability?

Probability as defined by Webster's dictionary is "the chance that a given event will occur". Examples that we are familiar with are the probability that it will rain the next day or the probability that you will win the lottery. In the first example, there are many factors that affect the weather—so many, in fact, that we cannot be certain that it will or will not rain the following day. Hence, as a predictive tool we usually assign a number between 0 and 1 (or between 0% and 100%) indicating our degree of certainty that the event, rain, will occur. If we say that there is a 30% chance of rain, we believe that if identical conditions prevail, then 3 times out of 10, rain will occur the next day. Alternatively, we believe that the *relative frequency* of rain is 3/10. Note that if the science of meteorology had accurate enough models, then it is conceivable that we could determine exactly whether rain would or would not occur. Or we could say that the *probability* is either 0 or 1. Unfortunately, we have not progressed that far. In the second example, winning the lottery, our chance of success, assuming a fair drawing, is just one out of the number of possible lottery number sequences. In this case, we are uncertain of the outcome, not because of the inaccuracy of our model, but because the experiment has been designed to produce uncertain results.

The common thread of these two examples is the presence of a *random experiment*, a *set of outcomes*, and the *probabilities* assigned to these outcomes. We will see later that these attributes are common to all probabilistic descriptions. In the lottery example, the experiment is the drawing, the outcomes are the lottery number sequences, and the probabilities assigned are $1/N$, where N = total number of lottery number sequences. Another common thread, which justifies the use of probabilistic methods, is the concept of *statistical regularity*. Although we may never be able to predict with certainty the outcome of an experiment, we are, nonetheless, able to predict "averages". For example, the average rainfall in the summer in Rhode Island is 9.76 inches, as shown in Figure 1.1, while in Arizona it is only 4.40

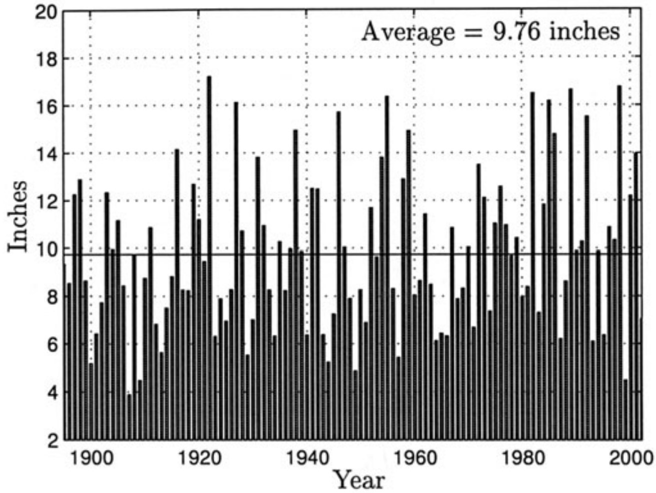

Figure 1.1: Annual summer rainfall in Rhode Island from 1895 to 2002 [NOAA/NCDC 2003].

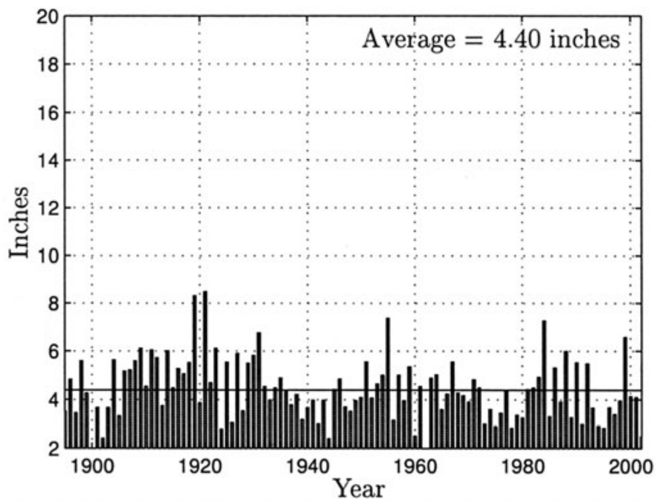

Figure 1.2: Annual summer rainfall in Arizona from 1895 to 2002 [NOAA/NCDC 2003].

inches, as shown in Figure 1.2. It is clear that the decision to plant certain types of crops could be made based on these averages. This is not to say, however, that we can predict the rainfall amounts for any given summer. For instance, in 1999 the summer rainfall in Rhode Island was only 4.5 inches while in 1984 the summer rainfall in Arizona was 7.3 inches. A somewhat more controlled experiment is the repeated tossing of a fair coin (one that is equally likely to come up heads or tails). We would expect about 50 heads out of 100 tosses, but of course, we could not predict the outcome of any one particular toss. An illustration of this is shown in Figure 1.3. Note that 53 heads were obtained in this particular experiment. This

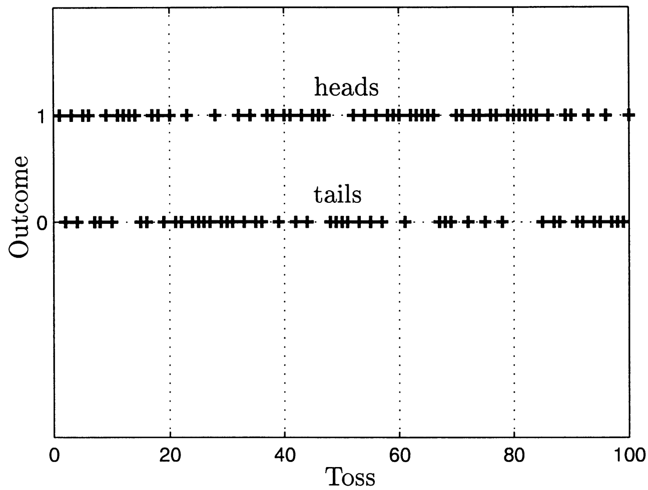

Figure 1.3: Outcomes for repeated fair coin tossings.

example, which is of seemingly little relevance to physical reality, actually serves as a good *model* for a variety of random phenomena. We will explore one example in the next section.

In summary, probability theory provides us with the ability to predict the behavior of random phenomena in the "long run." To the extent that this information is useful, probability can serve as a valuable tool for assessment and decision making. Its application is widespread, encountering use in all fields of scientific endeavor such as engineering, medicine, economics, physics, and others (see references at end of chapter).

1.2 Types of Probability Problems

Because of the mathematics required to determine probabilities, probabilistic methods are divided into two distinct types, *discrete* and *continuous*. A discrete approach is used when the number of experimental outcomes is finite (or infinite but countable as illustrated in Problem 1.7). For example, consider the number of persons at a business location that are talking on their respective phones anytime between 9:00 AM and 9:10 AM. Clearly, the possible outcomes are $0, 1, \ldots, N$, where N is the number of persons in the office. On the other hand, if we are interested in the length of time a particular caller is on the phone during that time period, then the outcomes may be anywhere from 0 to T minutes, where $T = 10$. Now the outcomes are infinite in number since they lie within the interval $[0, T]$. In the first case, since the outcomes are discrete (and finite), we can assign probabilities to the outcomes $\{0, 1, \ldots, N\}$. An equiprobable assignment would be to assign each outcome a probability of $1/(N+1)$. In the second case, the outcomes are continuous (and therefore infinite) and so it is not possible to assign a nonzero probability *to each outcome* (see Problem 1.6).

We will henceforth delineate between probabilities assigned to discrete outcomes and those assigned to continuous outcomes, with the discrete case always discussed first. The discrete case is easier to conceptualize and to describe mathematically. It will be important to keep in mind which case is under consideration since otherwise, certain paradoxes may result (as well as much confusion on the part of the student!).

1.3 Probabilistic Modeling

Probability models are simplified approximations to reality. They should be detailed enough to capture important characteristics of the random phenomenon so as to be useful as a prediction device, but not so detailed so as to produce an unwieldy model that is difficult to use in practice. The example of the number of telephone callers can be modeled by assigning a probability p to each person being on the phone anytime in the given 10-minute interval and *assuming* that whether one or more persons are on the phone does not affect the probability of others being on the phone. One can thus liken the event of being on the phone to a coin toss—if heads, a person is on the phone and if tails, a person is not on the phone. If there are $N = 4$ persons in the office, then the experimental outcome is likened to 4 coin tosses (either in succession or simultaneously—it makes no difference in the modeling). We can then ask for the probability that 3 persons are on the phone by determining the probability of 3 heads out of 4 coin tosses. The solution to this problem will be discussed in Chapter 3, where it is shown that the probability of k heads out of N coin tosses is given by

$$P[k] = \binom{N}{k} p^k (1-p)^{N-k} \tag{1.1}$$

where

$$\binom{N}{k} = \frac{N!}{(N-k)!k!}$$

for $k = 0, 1, \ldots, N$, and where $M! = 1 \cdot 2 \cdot 3 \cdots M$ for M a positive integer and by definition $0! = 1$. For our example, if $p = 0.75$ (we have a group of telemarketers) and $N = 4$ a compilation of the probabilities is shown in Figure 1.4. It is seen that the probability that three persons are on the phone is 0.42. Generally, the coin toss

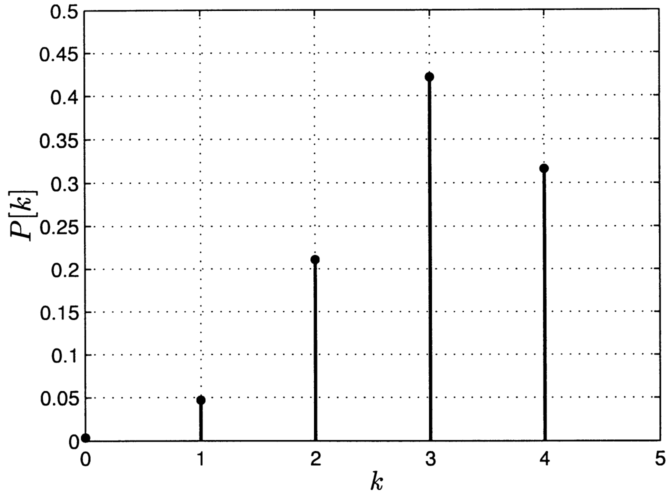

Figure 1.4: Probabilities for $N = 4$ coin tossings with $p = 0.75$.

model is a reasonable one for this type of situation. It will be poor, however, if the *assumptions are invalid.* Some practical objections to the model might be:

1. Different persons have different probabilities p (an eager telemarketer versus a not so eager one).

2. The probability of one person being on the phone is affected by whether his neighbor is on the phone (the two neighbors tend to talk about their planned weekends), i.e., the events are not "independent".

3. The probability p changes over time (later in the day there is less phone activity due to fatigue).

To accommodate these objections the model can be made more complex. In the end, however, the "more accurate" model may become a poorer predictor if the additional information used is not correct. It is generally accepted that a model should exhibit the property of "parsimony"—in other words, it should be as simple as possible.

The previous example had discrete outcomes. For continuous outcomes a frequently used probabilistic model is the *Gaussian* or "bell"-shaped curve. For the modeling of the length of time T a caller is on the phone it is not appropriate to ask for the probability that T will be *exactly*, for example, 5 minutes. This is because this probability will be zero (see Problem 1.6). Instead, we inquire as to the probability that T will be between 5 and 6 minutes. This question is answered by determining the area under the Gaussian curve shown in Figure 1.5. The form of the curve is given by

$$p_T(t) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(t-7)^2\right] \qquad -\infty < t < \infty \tag{1.2}$$

and although defined for all t, it is physically meaningful only for $0 \leq t \leq T_{\max}$,

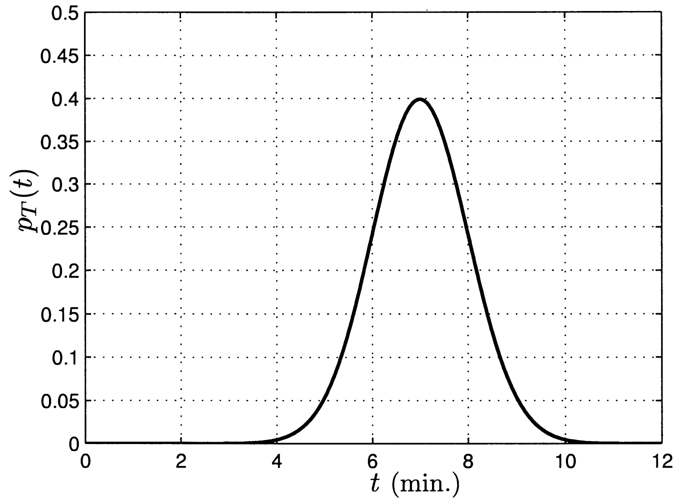

Figure 1.5: Gaussian or "bell"-shaped curve.

where $T_{\max} = 10$ for the current example. Since the area under the curve for times less than zero or greater than $T_{\max} = 10$ is nearly zero, this model is a reasonable approximation to physical reality. The curve has been chosen to be centered about $t = 7$ to relect an "average" time on the phone of 7 minutes for a given caller. Also, note that we let t denote the actual *value* of the *random* time T. Now, to determine the probability that the caller will be on the phone for between 5 and 6 minutes we integrate $p_T(t)$ over this interval to yield

$$P[5 \leq T \leq 6] = \int_5^6 p_T(t)dt = 0.1359. \tag{1.3}$$

The value of the integral must be numerically determined. Knowing the function $p_T(t)$ allows us to determine the probability for any interval. (It is called the probability density function (PDF) and is the probability per unit length. The PDF will be discussed in Chapter 10.) Also, it is apparent from Figure 1.5 that phone usage of duration less than 4 minutes or greater than 10 minutes is highly unlikely. Phone usage in the range of 7 minutes, on the other hand, is most probable. As before, some objections might be raised as to the accuracy of this model. A particularly lazy worker could be on the phone for only 3 minutes, as an example.

In this book we will henceforth assume that the models, which are mathematical in nature, are perfect and thus can be used to determine probabilities. In practice, the user must ultimately choose a model that is a reasonable one for the application of interest.

1.4 Analysis versus Computer Simulation

In the previous section we saw how to compute probabilities once we were given certain probability functions such as (1.1) for the discrete case and (1.2) for the continuous case. For many practical problems it is not possible to determine these functions. However, if we have a model for the random phenomenon, then we may carry out the experiment a large number of times to obtain an approximate probability. For example, to determine the probability of 3 heads in 4 tosses of a coin with probability of heads being $p = 0.75$, we toss the coin four times and count the number of heads, say $x_1 = 2$. Then, we repeat the experiment by tossing the coin four more times, yielding $x_2 = 1$ head. Continuing in this manner we execute a succession of 1000 experiments to produce the sequence of number of heads as $\{x_1, x_2, \ldots, x_{1000}\}$. Then, to determine the probability of 3 heads we use a *relative frequency* interpretation of probability to yield

$$P[3 \text{ heads}] = \frac{\text{Number of times 3 heads observed}}{1000}. \tag{1.4}$$

Indeed, early on probabilists did exactly this, although it was extremely tedious. *It is therefore of utmost importance to be able to simulate this procedure.* With the advent of the modern digital computer this is now possible. A digital computer has no problem performing a calculation once, 100 times, or 1,000,000 times. What is needed to implement this approach is a means to simulate the toss of a coin. Fortunately, this is quite easy as most scientific software packages have built-in *random number generators.* In MATLAB, for example, a number in the interval $(0,1)$ can be produced with the simple statement `x=rand(1,1)`. The number is chosen "at random" so that it is equally likely to be anywhere in the $(0,1)$ interval. As a result, a number in the interval $(0,1/2]$ will be observed with probability 1/2 and a number in the remaining part of the interval $(1/2,1)$ also with probability 1/2. Likewise, a number in the interval $(0,0.75]$ will be observed with probability $p = 0.75$. A computer simulation of the number of persons in the office on the telephone can thus be implemented with the MATLAB code (see Appendix 2A for a brief introduction to MATLAB):

```
number=0;
for i=1:4 % set up simulation for 4 coin tosses
  if rand(1,1)<0.75 % toss coin with p=0.75
    x(i,1)=1; % head
  else
```

```
        x(i,1)=0; % tail
      end
    number=number+x(i,1); % count number of heads
    end
```

Repeating this code segment 1000 times will result in a simulation of the previous experiment.

Similarly, for a continuous outcome experiment we require a means to generate a continuum of outcomes on a digital computer. Of course, strictly speaking this is not possible since digital computers can only provide a finite set of numbers, which is determined by the number of bits in each word. But if the number of bits is large enough, then the approximation is adequate. For example, with 64 bits we could represent 2^{64} numbers between 0 and 1, so that neighboring numbers would be $2^{-64} = 5 \times 10^{-20}$ apart. With this ability MATLAB can produce numbers that follow a Gaussian curve by invoking the statement `x=randn(1,1)`.

Throughout the text we will use MATLAB for examples and also exercises. However, any modern scientific software package can be used.

1.5 Some Notes to the Reader

The notation used in this text is summarized in Appendix A. Note that boldface type is reserved for vectors and matrices while regular face type will denote scalar quantities. All other symbolism is defined within the context of the discussion. Also, the reader will frequently be warned of potential "pitfalls". Common misconceptions leading to student errors will be described and noted. The pitfall or caution symbol shown below should be heeded.

The problems are of four types: computational or formula applications, word problems, computer exercises, and theoretical exercises. Computational or formula (denoted by **f**) problems are straightforward applications of the various formulas of the chapter, while word problems (denoted by **w**) require a more complete assimilation of the material to solve the problem. Computer exercises (denoted by **c**) will require the student to either use a computer to solve a problem or to simulate the analytical results. This will enhance understanding and can be based on MATLAB, although equivalent software may be used. Finally, theoretical exercises (denoted by **t**) will serve to test the student's analytical skills as well as to provide extensions to the material of the chapter. They are more challenging. Answers to selected problems are given in Appendix E. Those problems for which the answers are provided are noted in the problem section with the symbol (☺).

The version of MATLAB used in this book is 5.2, although newer versions should provide identical results. Many MATLAB outputs that are used for the text figures and for the problem solutions rely on random number generation. To match your results against those shown in the figures and the problem solutions, the same set of random numbers can be generated by using the MATLAB statements `rand('state',0)` and `randn('state',0)` at the beginning of each program. These statements will initialize the random number generators to produce the same set of random numbers. Finally, the MATLAB programs and code segments given in the book are indicated by the "typewriter" font, for example, `x=randn(1,1)`.

There are a number of other textbooks that the reader may wish to consult. They are listed in the following reference list, along with some comments on their contents.

Davenport, W.B., *Probability and Random Processes*, McGraw-Hill, New York, 1970. (Excellent introductory text.)

Feller, W., *An Introduction to Probability Theory and its Applications*, Vols. 1, 2, John Wiley, New York, 1950. (Definitive work on probability—requires mature mathematical knowledge.)

Hoel, P.G., S.C. Port, C.J. Stone, *Introduction to Probability Theory*, Houghton Mifflin Co., Boston, 1971. (Excellent introductory text but limited to probability.)

Leon-Garcia, A., *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, MA, 1994. (Excellent introductory text.)

Parzen, E., *Modern Probability Theory and Its Applications*, John Wiley, New York, 1960. (Classic text in probability—useful for all disciplines).

Parzen, E., *Stochastic Processes*, Holden-Day, San Francisco, 1962. (Most useful for Markov process descriptions.)

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965. (Classic but somewhat difficult text. Best used as a reference.)

Ross, S., *A First Course in Probability*, Prentice-Hall, Upper Saddle River, NJ, 2002. (Excellent introductory text covering only probability.)

Stark, H., J.W. Woods, *Probability and Random Processes with Applications to Signal Processing*, Third Ed., Prentice Hall, Upper Saddle River, NJ, 2002. (Excellent introductory text but at a somewhat more advanced level.)

References

Burdic, W.S., *Underwater Acoustic Systems Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

Ellenberg, S.S., D.M. Finklestein, D.A. Schoenfeld, "Statistical Issues Arising in AIDS Clinical Trials," *J. Am. Statist. Asssoc.*, Vol. 87, pp. 562–569, 1992.

Gustafson, D.E., A.S. Willsky, J.Y. Wang, M.C. Lancaster, J.H. Triebwasser, "ECG/VCG Rhythm Diagnosis Using Statistical Signal Analysis, Part II: Identification of Transient Rhythms," *IEEE Trans. Biomed. Eng.*, Vol. BME-25, pp. 353–361, 1978.

Ives, R.B., "The Applications of Advanced Pattern Recognition Techniques for the Discrimination Between Earthquakes and Nuclear Detonations," *Pattern Recognition*, Vol. 14, pp. 155–161, 1981.

Justice, J.H., "Array Processing in Exploration Seismology," in *Array Signal Processing*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, NJ, 1985.

Knight, W.S., R.G. Pridham, S.M. Kay, "Digital Signal Processing for Sonar," *Proc. IEEE*, Vol. 69, pp. 1451–1506, Nov. 1981.

NOAA/NCDC, Lawrimore, J., "Climate at a Glance," National Oceanic and Atmospheric Administration, http://www.ncdc.noaa.gov/oa/climate/resarch/cag3/NA.html, 2003.

Proakis, J., *Digital Communications*, Second Ed., McGraw-Hill, New York, 1989.

Skolnik, M.I., *Introduction to Radar Systems*, McGraw-Hill, New York, 1980.

Taylor, S., *Modelling Financial Time Series*, John Wiley, New York, 1986.

Problems

1.1 (⌣̈) **(w)** A fair coin is tossed. Identify the random experiment, the set of outcomes, and the probabilities of each possible outcome.

1.2 (w) A card is chosen at random from a deck of 52 cards. Identify the random experiment, the set of outcomes, and the probabilities of each possible outcome.

1.3 (w) A fair die is tossed and the number of dots on the face noted. Identify the random experiment, the set of outcomes, and the probabilities of each possible outcome.

1.4 (w) It is desired to predict the annual summer rainfall in Rhode Island for 2010. If we use 9.76 inches as our prediction, how much in error might we be, based on the past data shown in Figure 1.1? Repeat the problem for Arizona by using 4.40 inches as the prediction.

1.5 (⌣̈) **(w)** Determine whether the following experiments have discrete or continuous outcomes:

a. Throw a dart with a point tip at a dartboard.

b. Toss a die.

c. Choose a lottery number.

d. Observe the outdoor temperature using an analog thermometer.

e. Determine the current time in hours, minutes, seconds, and AM or PM.

1.6 (w) An experiment has $N = 10$ outcomes that are equally probable. What is the probability of each outcome? Now let $N = 1000$ and also $N = 1,000,000$ and repeat. What happens as $N \to \infty$?

1.7 (⌣̈) **(f)** Consider an experiment with possible outcomes $\{1, 2, 3, \ldots\}$. If we assign probabilities

$$P[k] = \frac{1}{2^k} \qquad k = 1, 2, 3, \ldots$$

to the outcomes, will these probabilties sum to one? Can you have an infinite number of outcomes but still assign nonzero probabilities to each outcome? Reconcile these results with that of Problem 1.6.

1.8 (w) An experiment consists of tossing a fair coin four times in succession. What are the possible outcomes? Now count up the number of outcomes with three heads. If the outcomes are equally probable, what is the probability of three heads? Compare your results to that obtained using (1.1).

1.9 (w) Perform the following experiment by *actually tossing* a coin of your choice. Flip the coin four times and observe the number of heads. Then, repeat this experiment 10 times. Using (1.1) determine the probability for $k = 0, 1, 2, 3, 4$ heads. Next use (1.1) to determine the number of heads that is most probable for a single experiment? In your 10 experiments which number of heads appeared most often?

1.10 (⌣̈) **(w)** A coin is tossed 12 times. The sequence observed is the 12-tuple $(H, H, T, H, H, T, H, H, H, H, T, H)$. Is this a fair coin? Hint: Determine $P[k = 9]$ using (1.1) assuming a probability of heads of $p = 1/2$.

1.11 (t) Prove that $\sum_{k=0}^{N} P[k] = 1$, where $P[k]$ is given by (1.1). Hint: First prove the binomial theorem

$$(a + b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$$

by induction (see Appendix B). Use Pascal's "triangle" rule

$$\binom{M}{k} = \binom{M-1}{k} + \binom{M-1}{k-1}$$

where

$$\binom{M}{k} = 0 \qquad k < 0 \text{ and } k > M.$$

1.12 (t) If $\int_a^b p_T(t)dt$ is the probability of observing T in the interval $[a, b]$, what is $\int_{-\infty}^{\infty} p_T(t)dt$?

1.13 (⌣) (f) Using (1.2) what is the probability of $T > 7$? Hint: Observe that $p_T(t)$ is symmetric about $t = 7$.

1.14 (⌣) (c) Evaluate the integral

$$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}t^2\right] dt$$

by using the approximation

$$\sum_{n=-L}^{L} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(n\Delta)^2\right] \Delta$$

where L is the integer closest to $3/\Delta$ (the rounded value), for $\Delta = 0.1$, $\Delta = 0.01$, $\Delta = 0.001$.

1.15 (c) Simulate a fair coin tossing experiment by modifying the code given in Section 1.4. Using 1000 repetitions of the experiment, count the number of times three heads occur. What is the simulated probability of obtaining three heads in four coin tosses? Compare your result to that obtained using (1.1).

1.16 (c) Repeat Problem 1.15 but instead consider a biased coin with $p = 0.75$. Compare your result to Figure 1.4.