1

A Wide Range of Applications

This is a book about testing hypotheses; more accurately, it is a book about making decisions. In this chapter we introduce some basic concepts in statistics related to decision theory, including events, random variables, samples, variation, and hypothesis. We consider a simple example of decision-making under uncertainty and review the history of statistics in decision-making.

1.1 Basic Concepts

1.1.1 Stochastic Phenomena

The two factors that distinguish the statistical from the deterministic approach are variation and the possibility of error. The effect of this variation is that a distribution of values takes the place of a single, unique outcome.

I found freshman physics extremely satisfying. Boyle's Law, $V = kT/P$, with its tidy relationship between the volume, temperature, and pressure of a perfect gas is just one example of the perfection I encountered there. The problem was I could never quite duplicate this (or any other) law in the freshman physics' laboratory. Maybe it was the measuring instruments, my lack of familiarity with the equipment, or simple measurement error, but I kept getting different values for the constant k.

By now I know that variation is the norm—particularly in the clinical and biological areas. Instead of getting a fixed, reproducible V to correspond to a specific T and P, one ends up, due to errors in measurement, with a distribution F of values instead. But I also know that with a large enough sample the mean and shape of this distribution are reproducible.

Figure 1.1a and 1.1b depict two such distributions. The first is a *normal* or *Gaussian distribution.* Examining the distribution curve, we see that the normally distributed variable can take all possible values between $-\infty$ and $+\infty$, but most of the time it takes values that are close to its median (and mean).

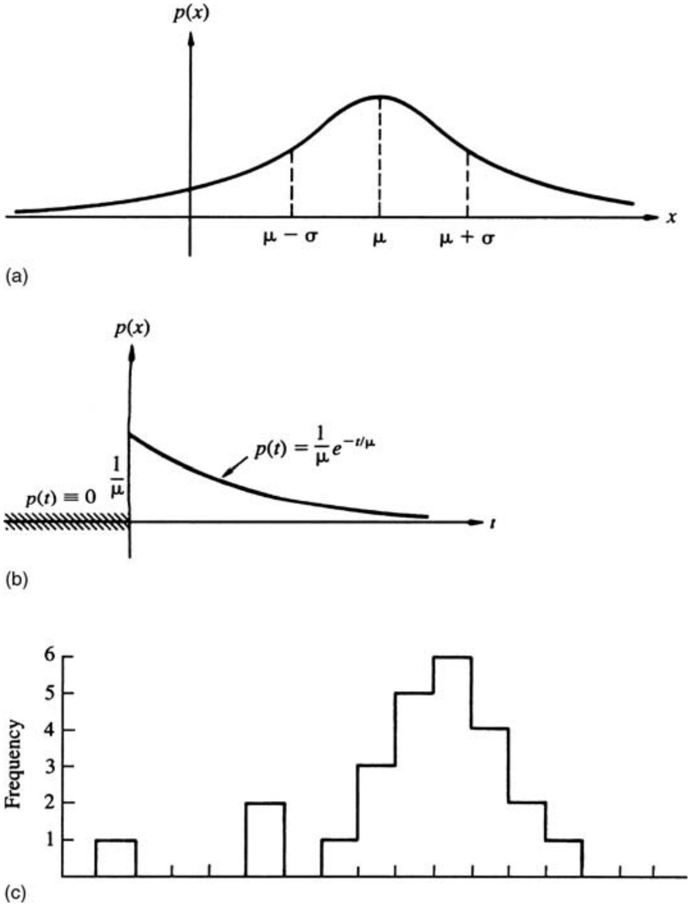

Fig. 1.1. Distributions: a) normal distribution, b) exponential distribution, c) distribution of values in a sample taken from a normal distribution.

The second is an *exponential distribution*: The exponentially distributed variable only takes positive values; half of the observations are small, crowded together in the left half of the distribution, but the balance is stretched out across a far wider range.

These distributions are both limiting cases: they represent the aggregate result of an infinite number of observations; thus, the distribution curves are smooth. The choppy histogram in Figure 1.1c is typical of what one sees with a small, finite sample of observations—in this case, a sample of 25 observations taken from a normal distribution. Still more typical of real-world data is the histogram of Figure 7.2, based on a sample taken from a mixture of two normal distributions.

The sample is not the population. To take an extreme, almost offensive example, suppose that every member of a Los Angeles, California jury were