The graphical method of jittering, used in Figures 1.1 and 1.2 and elsewhere in this book, is discussed in Chambers et al. (1983). For information on the statistical packages R and Bugs, see Becker, Chambers, and Wilks (1988), R Project (2002), Fox (2002), Venables and Ripley (2002), and Spiegelhalter et al. (1994, 2003).

Norvig (2007) describes the principles and details of the Bayesian spelling corrector.

1.12 Exercises

1. Conditional probability: suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $\Pr(\theta = 1) = 0.5$ and $\Pr(\theta = 2) = 0.5$.
   (a) For $\sigma = 2$, write the formula for the marginal probability density for y and sketch it.
   (b) What is $\Pr(\theta = 1|y = 1)$, again supposing $\sigma = 2$?
   (c) Describe how the posterior density of θ changes in shape as σ is increased and as it is decreased.
2. Conditional means and variances: show that (1.8) and (1.9) hold if u is a vector.
3. Probability calculation for genetics (from Lindley, 1965): suppose that in each individual of a large population there is a pair of genes, each of which can be either x or X, that controls eye color: those with xx have blue eyes, while heterozygotes (those with Xx or xX) and those with XX have brown eyes. The proportion of blue-eyed individuals is p^2 and of heterozygotes is $2p(1-p)$, where $0 < p < 1$. Each parent transmits one of its own genes to the child; if a parent is a heterozygote, the probability that it transmits the gene of type X is $\frac{1}{2}$. Assuming random mating, show that among brown-eyed children of brown-eyed parents, the expected proportion of heterozygotes is $2p/(1+2p)$. Suppose Judy, a brown-eyed child of brown-eyed parents, marries a heterozygote, and they have n children, all brown-eyed. Find the posterior probability that Judy is a heterozygote and the probability that her first grandchild has blue eyes.
4. Probability assignment: we will use the football dataset to estimate some conditional probabilities about professional football games. There were twelve games with point spreads of 8 points; the outcomes in those games were: -7, -5, -3, -3, 1, 6, 7, 13, 15, 16, 20, and 21, with positive values indicating wins by the favorite and negative values indicating wins by the underdog. Consider the following conditional probabilities:

$$\Pr(\text{favorite wins} \,|\, \text{point spread} = 8),$$
$$\Pr(\text{favorite wins by at least } 8 \,|\, \text{point spread} = 8),$$
$$\Pr(\text{favorite wins by at least } 8 \,|\, \text{point spread} = 8 \text{ and favorite wins}).$$

   (a) Estimate each of these using the relative frequencies of games with a point spread of 8.
   (b) Estimate each using the normal approximation for the distribution of (outcome − point spread).
5. Probability assignment: the 435 U.S. Congressmembers are elected to two-year terms; the number of voters in an individual congressional election varies from about 50,000 to 350,000. We will use various sources of information to estimate roughly the probability that at least one congressional election is tied in the next national election.
   (a) Use any knowledge you have about U.S. politics. Specify clearly what information you are using to construct this conditional probability, even if your answer is just a guess.
   (b) Use the following information: in the period 1900–1992, there were 20,597 congressional elections, out of which 6 were decided by fewer than 10 votes and 49 decided by fewer than 100 votes.

See Gelman, King, and Boscardin (1998), Mulligan and Hunter (2001), and Gelman, Katz, and Tuerlinckx (2002) for more on this topic.

6. Conditional probability: approximately 1/125 of all births are fraternal twins and 1/300 of births are identical twins. Elvis Presley had a twin brother (who died at birth). What is the probability that Elvis was an identical twin? (You may approximate the probability of a boy or girl birth as $\frac{1}{2}$.)

7. Conditional probability: the following problem is loosely based on the television game show *Let's Make a Deal*. At the end of the show, a contestant is asked to choose one of three large boxes, where one box contains a fabulous prize and the other two boxes contain lesser prizes. After the contestant chooses a box, Monty Hall, the host of the show, opens one of the two boxes containing smaller prizes. (In order to keep the conclusion suspenseful, Monty does not open the box selected by the contestant.) Monty offers the contestant the opportunity to switch from the chosen box to the remaining unopened box. Should the contestant switch or stay with the original choice? Calculate the probability that the contestant wins under each strategy. This is an exercise in being clear about the information that should be conditioned on when constructing a probability judgment. See Selvin (1975) and Morgan et al. (1991) for further discussion of this problem.

8. Subjective probability: discuss the following statement. 'The probability of event E is considered "subjective" if two rational persons A and B can assign unequal probabilities to E, $P_A(E)$ and $P_B(E)$. These probabilities can also be interpreted as "conditional": $P_A(E) = P(E|I_A)$ and $P_B(E) = P(E|I_B)$, where I_A and I_B represent the knowledge available to persons A and B, respectively.' Apply this idea to the following examples.
   (a) The probability that a '6' appears when a fair die is rolled, where A observes the outcome of the die roll and B does not.
   (b) The probability that Brazil wins the next World Cup, where A is ignorant of soccer and B is a knowledgeable sports fan.

9. Simulation of a queuing problem: a clinic has three doctors. Patients come into the clinic at random, starting at 9 a.m., according to a Poisson process with time parameter 10 minutes: that is, the time after opening at which the first patient appears follows an exponential distribution with expectation 10 minutes and then, after each patient arrives, the waiting time until the next patient is independently exponentially distributed, also with expectation 10 minutes. When a patient arrives, he or she waits until a doctor is available. The amount of time spent by each doctor with each patient is a random variable, uniformly distributed between 5 and 20 minutes. The office stops admitting new patients at 4 p.m. and closes when the last patient is through with the doctor.
   (a) Simulate this process once. How many patients came to the office? How many had to wait for a doctor? What was their average wait? When did the office close?
   (b) Simulate the process 100 times and estimate the median and 50% interval for each of the summaries in (a).