say that G is a family of **conjugate priors** to H. Thus, the beta distribution is a conjugate prior to the binomial, and the normal is self-conjugate. Conjugate priors may not exist; when they do, selecting a member of the conjugate family as a prior is done mostly for mathematical convenience, since the posterior can be evaluated very simply. More generally, numerical methods of integration would have to be used to evaluate the posterior.

15.4 Concluding Remarks

This chapter has presented a brief introduction to two areas of mathematical statistics, decision theory and Bayesian inference. The analysis has been somewhat more abstract and theoretical than that of earlier chapters.

In order for decision theory to be relevant to a practical problem, the problem must be such that a choice is to be made from a set of specified actions and a definite measure of loss is associated with each possible action. Critics of decision theory claim that most scientific investigations do not fit this paradigm; some have gone so far as to call such a point of view totalitarian! In fact, experimental scientists very rarely use decision theoretic methods in analyzing their data, but decision theory has been used more often in business and economics, where it seems to be easier to specify the relevant loss functions.

The subjectivist, or Bayesian, point of view toward statistics is somewhat controversial. Arguments between frequentists and Bayesians have often been quite heated, and it is probably fair to say that the Bayesian approach has not been adopted by most practicing statisticians. Critics of the Bayesian paradigm look with disfavor on the introduction of a prior distribution. Some dislike the subjective element thereby introduced, maintaining that statistics should be "objective." Others, although not totally unsympathetic, question how prior distributions can be determined in practice.

This book has been relatively unconcerned with philosophical or foundational matters and has made no attempt to develop a consistent "theory" of statistics. The subject matter under investigation and the role that statistics plays in the investigation should effectively determine whether it is more appropriate for the user of statistics to take a Bayesian, decision-theoretic, frequentist, or purely data-analytic point of view, or some combination of these.

15.5 Problems

1 The losses of actions a_1, a_2, and a_3, depending on θ_1, θ_2, and θ_3, are given in the following table:

	θ_1	θ_2	θ_3
a_1	2	1	5
a_2	0	4	6
a_3	4	0	2

An observation X has the following distributions depending on θ:

	θ_1	θ_2	θ_3
x_1	.4	.5	.3
x_2	.6	.5	.7

a Pick three decision rules and evaluate their risks.

b Determine the minimax rule among those chosen in part (a).

c Assuming a uniform prior distribution on θ, determine the Bayes rule.

d Suppose that $X = x_2$. Use posterior analysis to determine the Bayes rule among all possible decision rules.

2 Suppose that for Example B in Section 15.2.1 the prior distribution of k is taken to be uniform. How does the analysis change?

3 Prove Theorem A of Section 15.2.2 for the discrete case.

4 Your investment advisor tells you that he believes a certain stock is going up and that you should buy. If he is correct and you buy, you will gain \$1000; if he is incorrect and you buy, you will lose \$1000. If you do nothing, you lose nothing. You might also choose to sell. If you sell and he is correct, you lose \$1000; if you sell and he is incorrect, you gain \$1000. On the basis of experience, you believe that his prognosis is correct about $\frac{2}{3}$ of the time. Your prior probability that the stock will go up or down is $\frac{2}{5}$ or $\frac{3}{5}$, respectively. What are the posterior risks of buying, selling, and doing nothing? What is the Bayes rule?

5 A biased coin has an unknown probability θ of heads. It is thrown once, and you observe the outcome. You then have a choice whether or not to play a certain game. In the game, you predict the next toss. If you predict correctly, you are paid \$2; if you predict incorrectly, you lose \$1. If you don't play the game, you lose nothing and gain nothing. You thus have three possible actions: (1) play and predict the same as on the first toss, (2) play and predict the opposite of the first toss, (3) don't play.

a What are the risk functions for the three actions?

b What is the minimax rule?

c If you have a uniform prior on θ, what is the Bayes rule?

d Suppose that you are paid \$$D$ if you predict correctly and you lose \$1 if you predict incorrectly. How does the analysis of parts (a) through (c) change?

6 Consider the following classification problem. Classes A and B are equally likely, so that the prior distribution is $P(A) = P(B) = \frac{1}{2}$. A random variable X is observed; if X comes from class A, its distribution is $N(0, 1)$, and if X comes from class B, its distribution is $N(2, 1)$.

a Under 0–1 loss, what is the Bayes rule for classification?

b Suppose that $X = 1.5$. How does the Bayes rule classify this observation? What is the probability of making a classification error in this case?

c What is the probability of making a classification error when using the Bayes rule?

d Suppose that the loss function is such that you lose twice as much by misclassifying A as you do by misclassifying B. Redo parts (a) through (c) under this assumption.

e Redo parts (a) through (c) under the prior probabilities $P(A) = \frac{2}{3}$ and $P(B) = \frac{1}{3}$.

7 Suppose that as in the previous problem, $P(A) = P(B)$, but that if X comes from A, its distribution is $N(0, 1)$ whereas if X comes from B, its distribution is $N(1, 25)$. Under 0–1 loss, what is the Bayes rule for classification?

8 Consider a two-class discrimination problem in which X is either $N(0, \sigma^2)$ under A or $N(1, \sigma^2)$ under B and suppose that $P(B) = 2P(A)$. For $\sigma^2 = .1, 1, 25, 100$:

a What is the Bayes rule with 0–1 loss?

b In the long run, what proportion of observations will be classified as A?

9 Suppose that $X \sim \text{bin}(3, p)$. Consider choosing between H: $p \leq \frac{1}{2}$ and K: $p > \frac{1}{2}$. Compare the risks of the following decision rules as functions of p if you lose \$1 for incorrectly choosing H and \$2 for incorrectly choosing K:

d_1: choose K if $X > 0$
d_2: choose H if $X \leq 1$
d_3: choose H if $X \leq 2$

Assuming a uniform prior on p, which rule has the smallest Bayes risk?

10 Prove Theorem B of Section 15.2.4 for the discrete case. Why is the assumption that $\pi(\theta) > 0$ for all θ necessary?

11 Suppose that $X \sim \text{bin}(2, p)$. Compare the risk functions for the following estimates of p using squared error loss:

$$\hat{p} = \frac{X}{2}$$

$$\tilde{p} = \frac{X+1}{3}$$

$$\bar{p} = \frac{X+1}{4}$$

Are any of these estimates dominated by the others?

12 Suppose that $X_1, \ldots, X_n$ are independent and that each follows an exponential distribution

$$f(x \mid \theta) = \frac{1}{\theta} e^{-x/\theta}$$

Assuming squared error loss, graph the risk function of each of the following estimates of θ:

$$d_1 = \overline{X}$$
$$d_2 = nX_{(1)}$$
$$d_3 = \frac{n\overline{X}}{n+1}$$

Does any estimate dominate another? (*Hint:* Show that $X_{(1)}$ follows an exponential distribution.)

13 A lot consisting of n items contains k defective items. Assume that k has a prior distribution, which is $\text{bin}(n, p)$. One item is selected at random and tested. Determine the posterior distribution of k if (a) the item is defective and (b) the item is not defective.

14 Suppose that X is a geometric random variable:

$$f(x \mid p) = (1-p)^{x-1}p$$

Let p have a prior distribution that is uniform on [0, 1].

a What is the posterior distribution of p?

b What is the Bayes estimate of p under squared error loss?

c What is the mle?

15 Suppose that a parameter, Θ, takes on values $\theta_1 = 1$, $\theta_2 = 10$, and $\theta_3 = 20$. The distribution of X is discrete and depends on Θ as shown in the following table:

	θ_1	θ_2	θ_3
x_1	.1	.2	.4
x_2	.1	.2	.2
x_3	.2	.2	.2
x_4	.6	.4	.2

Assume a prior distribution of Θ: $P(\theta_1) = .5$, $P(\theta_2) = .25$, and $P(\theta_3) = .25$.

a Suppose that x_2 is observed. What is the posterior distribution of Θ?

b What is the Bayes estimate under squared error loss in this case?

c What is the Bayes estimate for the loss function $l(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$?

d Suppose that a second independent observation, x_1, is made. What does the posterior distribution become?

16 Suppose that the prior distribution of θ is uniform on some interval. Show that the posterior distribution, $h(\theta \mid \mathbf{x})$ is exactly proportional to the likelihood function $f(\mathbf{x} \mid \theta)$.

17 Consider two simple hypotheses H and K with prior probabilities $P(H)$ and $P(K)$, giving prior odds for H equal to $P(H)/P(K)$. Data $\mathbf{X}$ are observed that have density $f(\mathbf{x} \mid H)$ and $f(\mathbf{x} \mid K)$ under the two hypotheses. Show that the posterior odds of H to K are equal to the prior odds multiplied by the likelihood ratio.

18 For an estimation problem, suppose that the loss function is $l(\theta, \hat{\theta}) = w(\theta)(\theta - \hat{\theta})^2$. Show that the Bayes estimate of θ is

$$\hat{\theta} = \frac{E[w(\Theta)\Theta \mid \mathbf{x}]}{E[w(\Theta) \mid \mathbf{x}]}$$

19 Suppose that θ is a parameter that takes on values on the real line. Consider the loss function

$$l(\theta, \hat{\theta}) = \begin{cases} 0, & |\theta - \hat{\theta}| \leq c \\ 1, & |\theta - \hat{\theta}| > c \end{cases}$$

Show that the Bayes estimate of θ is the midpoint of the interval I of length $2c$ that maximizes $P(\Theta \in I \mid \mathbf{x})$.

20 Let $X \sim N(\mu, 1)$ and consider estimating μ under squared error loss. Is the estimate $\hat{\mu} = 1944$ (my favorite number) admissible?

21 In an estimation problem, suppose that the loss function is $l(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$. Show that the Bayes estimate of θ is any median of the posterior distribution, i.e., any number $m(\mathbf{x})$ such that $P(\theta \leq m(\mathbf{x}) \mid \mathbf{X} = \mathbf{x}) \geq .5$ and $P(\theta \geq m(\mathbf{x}) \mid \mathbf{X} = \mathbf{x}) \geq .5$.

22 (Laplace's rule of succession) Laplace claimed that having observed an event happen n times in a row and never failing to happen, the probability that the event will occur the next time is $(n+1)/(n+2)$. Can you suggest a rationale for this claim?

23 Suppose that μ has a prior distribution that is $N(0, 1)$. Nine independent observations are taken and are $N(\mu, 9)$.

a Calculate the posterior mean and the precision.

b Find the Bayes risk of the estimate.

c Find a 95% credibility interval for μ.

24 Suppose that μ has a prior distribution that is $N(0, 1)$ and that $X \sim N(\mu, .25)$. What is the Bayes estimate of μ under squared error loss? Compare its risk to that of the mle.

25 Let X have a beta distribution with parameters a and b.

a Show that

$$E(X^r) = \frac{\Gamma(a+r)}{\Gamma(a)} \left[\frac{\Gamma(a+b)}{\Gamma(a+b+r)} \right]$$

[*Hint:* Use the recurrence relation for the gamma function: $\Gamma(x+1) = x\Gamma(x)$.]

b Prove Theorem A of Section 15.3.2.

26 A beta distribution has mean .3 and variance .01. What are a and b?

27 If a thumbtack is tossed in the air, it can land either on its base (with the point sticking up) or on its side. Let p denote the probability that it lands on its base.

a Without doing any experiments, what is your best guess for p?

b Select a and b to form a beta prior distribution on p that reasonably reflects your prior opinion about possible values of p. Graph this prior.

c Toss a thumbtack 10 times. Evaluate the posterior distribution of p, and plot it.

d Toss the thumbtack 40 more times. Evaluate the posterior, and plot it.

e Compare your results with those of other students.

28 Suppose that $X_1, \ldots, X_n$ are independent Poisson random variables with mean θ and that θ has a gamma prior distribution.

a Show that the posterior distribution of θ is also gamma. (Thus, the gamma is a conjugate prior of the Poisson.)

b Determine the Bayes estimate of θ under squared error loss, and show that it is a weighted average of the prior mean and $\overline{X}$.

c Compare the risk of the Bayes estimate to that of $\overline{X}$.

29 Show that the gamma is a conjugate prior of the exponential distribution.

30 Suppose that X is normally distributed with mean 0 and unknown precision ξ. Find a family of conjugate priors.