where $m = \max(x_1, \dots, x_n)$, $c(b,t) = [\sum_{j=t}^{\infty} j^{-b}]^{-1}$, $b > 1$.

(b) Suppose that $\max(x_1, \dots, x_n) = x_1 = m$ for all n. Show that $\pi(m \mid x_1, \dots, x_n) \to 1$ as $n \to \infty$ whatever be a. Interpret this result.

5. In Example 1.2.1 suppose n is large and $(1/n)\sum_{i=1}^n x_i = \bar{x}$ is not close to 0 or 1 and the prior distribution is beta, $\beta(r,s)$. Justify the following approximation to the posterior distribution

$$P[\theta \le t \mid X_1 = x_1, \dots, X_n = x_n] \approx \Phi\left(\frac{t - \widetilde{\mu}}{\widetilde{\sigma}}\right)$$

where Φ is the standard normal distribution function and

$$\widetilde{\mu} = \frac{n}{n+r+s}\bar{x} + \frac{r}{n+r+s}, \quad \widetilde{\sigma}^2 = \frac{\widetilde{\mu}(1-\widetilde{\mu})}{n+r+s}.$$

Hint: Let $\beta(a,b)$ denote the posterior distribution. If a and b are integers, then $\beta(a,b)$ is the distribution of $(a\bar{V}/b\bar{W})[1 + (a\bar{V}/b\bar{W})]^{-1}$, where $V_1, \dots, V_a, W_1, \dots, W_b$ are independent standard exponential. Next use the central limit theorem and Slutsky's theorem.

6. Show that a conjugate family of distributions for the Poisson family is the gamma family.

7. Show rigorously using (1.2.8) that if in Example 1.1.1, $D = N\theta$ has a $\mathcal{B}(N, \pi_0)$ distribution, then the posterior distribution of D given $X = k$ is that of $k + \mathbf{Z}$ where $\mathbf{Z}$ has a $\mathcal{B}(N - n, \pi_0)$ distribution.

8. Let $(X_1, \dots, X_{n+k})$ be a sample from a population with density $f(x \mid \theta)$, $\theta \in \Theta$. Let $\boldsymbol{\theta}$ have prior density π. Show that the conditional distribution of $(\boldsymbol{\theta}, X_{n+1}, \dots, X_{n+k})$ given $X_1 = x_1, \dots, X_n = x_n$ is that of $(Y, \mathbf{Z}_1, \dots, \mathbf{Z}_k)$ where the marginal distribution of Y equals the posterior distribution of $\boldsymbol{\theta}$ given $X_1 = x_1, \dots, X_n = x_n$, and the conditional distribution of the $\mathbf{Z}_i$'s given $Y = t$ is that of sample from the population with density $f(x \mid t)$.

9. Show in Example 1.2.1 that the conditional distribution of $\boldsymbol{\theta}$ given $\sum_{i=1}^n X_i = k$ agrees with the posterior distribution of $\boldsymbol{\theta}$ given $X_1 = x_1, \dots, X_n = x_n$, where $\sum_{i=1}^n x_i = k$.

10. Suppose $X_1, \dots, X_n$ is a sample with $X_i \sim p(x \mid \theta)$, a regular model and integrable as a function of θ. Assume that $A = \{x : p(x \mid \theta) > 0\}$ does not involve θ.

(a) Show that the family of priors

$$\pi(\theta) = \prod_{i=1}^{N} p(\xi_i \mid \theta) \Big/ \int_\Theta \prod_{i=1}^{N} p(\xi_i \mid \theta) d\theta$$

where $\xi_i \in A$ and $N \in \{1, 2, \dots\}$ is a conjugate family of prior distributions for $p(\mathbf{x} \mid \theta)$ and that the posterior distribution of θ given $\mathbf{X} = \mathbf{x}$ is

$$\pi(\theta \mid \mathbf{x}) = \prod_{i=1}^{N'} p(\xi_i' \mid \theta) \Big/ \int_\Theta \prod_{i=1}^{N'} p(\xi_i' \mid \theta) d\theta$$

where $N' = N + n$ and $(\xi_i', \dots, \xi_{N'}') = (\xi_1, \dots, \xi_N, x_1, \dots, x_n)$.

(b) Use the result (a) to give $\pi(\theta)$ and $\pi(\theta \mid \mathbf{x})$ when

$$\begin{aligned} p(x \mid \theta) &= \theta \exp\{-\theta x\},\ x > 0,\ \theta > 0 \\ &= 0 \text{ otherwise.} \end{aligned}$$

11. Let $p(x \mid \theta) = \exp\{-(x - \theta)\}$, $0 < \theta < x$ and let $\pi(\theta) = 2\exp\{-2\theta\}$, $\theta > 0$. Find the posterior density $\pi(\theta \mid x)$.

12. Suppose $p(\mathbf{x} \mid \theta)$ is the density of i.i.d. $X_1, \dots, X_n$, where $X_i \sim \mathcal{N}\left(\mu_0, \frac{1}{\theta}\right)$, μ_0 is known, and $\theta = \sigma^{-2}$ is (called) the *precision* of the distribution of X_i.

(a) Show that $p(\mathbf{x} \mid \theta) \propto \theta^{\frac{1}{2}n} \exp\left(-\frac{1}{2}t\theta\right)$ where $t = \sum_{i=1}^n (X_i - \mu_0)^2$ and $\propto$ denotes "proportional to" as a function of θ.

(b) Let $\pi(\theta) \propto \theta^{\frac{1}{2}(\lambda-2)} \exp\left\{-\frac{1}{2}\nu\theta\right\}$, $\nu > 0$, $\lambda > 0$; $\theta > 0$. Find the posterior distribution $\pi(\theta \mid \mathbf{x})$ and show that if λ is an integer, given $\mathbf{x}$, $\theta(t + \nu)$ has a $\chi^2_{\lambda+n}$ distribution. Note that, unconditionally, $\nu\theta$ has a χ^2_λ distribution.

(c) Find the posterior distribution of σ.

13. Show that if $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ and we formally put $\pi(\mu, \sigma) = \frac{1}{\sigma}$, then the posterior density $\pi(\mu \mid \bar{x}, s^2)$ of μ given $(\bar{x}, s^2)$ is such that $\sqrt{n}\frac{(\mu - \bar{X})}{s} \sim t_{n-1}$. Here $s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$.
Hint: Given μ and σ, $\bar{X}$ and s^2 are independent with $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ and $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$. This leads to $p(\bar{x}, s^2 \mid \mu, \sigma^2)$. Next use Bayes rule.

14. In a Bayesian model where $X_1, \dots, X_n, X_{n+1}$ are i.i.d. $f(x \mid \theta)$, $\boldsymbol{\theta} \sim \pi$, the *predictive distribution* is the marginal distribution of X_{n+1}. The *posterior predictive distribution* is the conditional distribution of X_{n+1} given $X_1, \dots, X_n$.

(a) If f and π are the $\mathcal{N}(\theta, \sigma_0^2)$ and $\mathcal{N}(\theta_0, \tau_0^2)$ densities, compute the predictive and posterior predictive distribution.

(b) Discuss the behavior of the two predictive distributions as $n \to \infty$.

15. *The Dirichlet distribution is a conjugate prior for the multinomial.* The Dirichlet distribution, $\mathcal{D}(\boldsymbol{\alpha})$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_r)^T$, $\alpha_j > 0$, $1 \le j \le r$, has density

$$f_{\boldsymbol{\alpha}}(\mathbf{u}) = \frac{\Gamma\left(\sum_{j=1}^r \alpha_j\right)}{\prod_{j=1}^r \Gamma(\alpha_j)} \prod_{j=1}^r u_j^{\alpha_j - 1},\ 0 < u_j < 1,\ \sum_{j=1}^r u_j = 1.$$

Let $\mathbf{N} = (N_1, \dots, N_r)$ be multinomial

$$\mathcal{M}(n, \boldsymbol{\theta}),\ \boldsymbol{\theta} = (\theta_1, \dots, \theta_r)^T,\ 0 < \theta_j < 1,\ \sum_{j=1}^r \theta_j = 1.$$

Show that if the prior $\pi(\theta)$ for θ is $\mathcal{D}(\alpha)$, then the posterior $\pi(\theta \mid \mathbf{N} = \mathbf{n})$ is $\mathcal{D}(\alpha + \mathbf{n})$, where $\mathbf{n} = (n_1, \dots, n_r)$.

Problems for Section 1.3

1. Suppose the possible states of nature are θ_1, θ_2, the possible actions are a_1, a_2, a_3, and the loss function $l(\theta, a)$ is given by

$\theta \backslash a$	a_1	a_2	a_3
θ_1	0	1	2
θ_2	2	0	1

Let X be a random variable with frequency function $p(x, \theta)$ given by

$\theta \backslash x$	0	1
θ_1	p	$(1-p)$
θ_2	q	$(1-q)$

and let $\delta_1, \dots, \delta_9$ be the decision rules of Table 1.3.3. Compute and plot the risk points when

(a) $p = q = .1$,

(b) $p = 1 - q = .1$.

(c) Find the minimax rule among $\delta_1, \dots, \delta_9$ for the preceding case (a).

(d) Suppose that θ has prior $\pi(\theta_1) = 0.5$, $\pi(\theta_2) = 0.5$. Find the Bayes rule for case (a).

2. Suppose that in Example 1.3.5, a new buyer makes a bid and the loss function is changed to

$\theta \backslash a$	a_1	a_2	a_3
θ_1	0	7	4
θ_2	12	1	6

(a) Compute and plot the risk points in this case for each rule $\delta_1, \dots, \delta_9$ of Table 1.3.3.

(b) Find the minimax rule among $\{\delta_1, \dots, \delta_9\}$.

(c) Find the minimax rule among the randomized rules.

(d) Suppose θ has prior $\pi(\theta_1) = \gamma$, $\pi(\theta_2) = 1 - \gamma$. Find the Bayes rule when (i) $\gamma = 0.5$ and (ii) $\gamma = 0.1$.

3. The problem of selecting the better of two treatments or of deciding whether the effect of one treatment is beneficial or not often reduces to the problem of deciding whether $\theta < 0$, $\theta = 0$ or $\theta > 0$ for some parameter θ. See Example 1.1.3. Let the actions corresponding to deciding whether $\theta < 0$, $\theta = 0$ or $\theta > 0$ be denoted by $-1, 0, 1$, respectively and suppose the loss function is given by (from Lehmann, 1957)

$\theta \backslash a$	-1	0	1
< 0	0	c	$b+c$
$= 0$	b	0	b
> 0	$b+c$	c	0

where b and c are positive. Suppose $\mathbf{X}$ is a $\mathcal{N}(\theta, 1)$ sample and consider the decision rule

$$\begin{aligned} \delta_{r,s}(\mathbf{X}) = \quad & -1 \quad \text{if } \bar{X} < r \\ & 0 \quad \text{if } r \leq \bar{X} \leq s \\ & 1 \quad \text{if } \bar{X} > s. \end{aligned}$$

(a) Show that the risk function is given by

$$\begin{aligned} R(\theta, \delta_{r,s}) &= c\bar{\Phi}(\sqrt{n}(r-\theta)) + b\bar{\Phi}(\sqrt{n}(s-\theta)), & \theta < 0 \\ &= b\bar{\Phi}(\sqrt{n}s) + b\Phi(\sqrt{n}r), & \theta = 0 \\ &= c\Phi(\sqrt{n}(s-\theta)) + b\Phi(\sqrt{n}(r-\theta)), & \theta > 0 \end{aligned}$$

where $\bar{\Phi} = 1 - \Phi$, and Φ is the $\mathcal{N}(0,1)$ distribution function.

(b) Plot the risk function when $b = c = 1$, $n = 1$ and

$$\text{(i) } r = -s = -1, \text{ (ii) } r = -\frac{1}{2}s = -1.$$

For what values of θ does the procedure with $r = -s = -1$ have smaller risk than the procedure with $r = -\frac{1}{2}s = -1$?

4. *Stratified sampling.* We want to estimate the mean $\mu = E(X)$ of a population that has been divided (stratified) into s mutually exclusive parts (strata) (e.g., geographic locations or age groups). Within the jth stratum we have a sample of i.i.d. random variables $X_{1j}, \ldots, X_{n_j j}$; $j = 1, \ldots, s$, and a stratum sample mean $\bar{X}_j$; $j = 1, \ldots, s$. We assume that the s samples from different strata are independent. Suppose that the jth stratum has $100p_j\%$ of the population and that the jth stratum population mean and variances are μ_j and σ_j^2. Let $N = \sum_{j=1}^{s} n_j$ and consider the two estimators

$$\widehat{\mu}_1 = N^{-1} \sum_{j=1}^{s} \sum_{i=1}^{n_j} X_{ij}, \; \widehat{\mu}_2 = \sum_{j=1}^{s} p_j \bar{X}_j$$

where we assume that p_j, $1 \leq j \leq s$, are known.

(a) Compute the biases, variances, and MSEs of $\widehat{\mu}_1$ and $\widehat{\mu}_2$. How should n_j, $0 \leq j \leq s$, be chosen to make $\widehat{\mu}_1$ unbiased?

(b) *Neyman allocation.* Assume that $0 < \sigma_j^2 < \infty$, $1 \leq j \leq s$, are known (estimates will be used in a later chapter). Show that the strata sample sizes that minimize $MSE(\widehat{\mu}_2)$ are given by

$$n_k = N \frac{p_k \sigma_k}{\sum_{j=1}^{s} p_j \sigma_j}, \; k = 1, \ldots, s. \tag{1.7.3}$$

Hint: You may use a Lagrange multiplier.

(c) Show that $MSE(\widehat{\mu}_1)$ with $n_k = p_k N$ minus $MSE(\widehat{\mu}_2)$ with n_k given by (1.7.3) is $N^{-1}\sum_{j=1}^{s} p_j(\sigma_j - \bar{\sigma})^2$, where $\bar{\sigma} = \sum_{j=1}^{s} p_j \sigma_j$.

5. Let $\bar{X}_b$ and $\widehat{X}_b$ denote the sample mean and the sample median of the sample $X_1 - b, \dots, X_n - b$. If the parameters of interest are the population mean and median of $X_i - b$, respectively, show that $MSE(\bar{X}_b)$ and $MSE(\widehat{X}_b)$ are the same for all values of b (the MSEs of the sample mean and sample median are *invariant* with respect to shift).

6. Suppose that $X_1, \dots, X_n$ are i.i.d. as $X \sim F$, that $\widehat{X}$ is the median of the sample, and that n is odd. We want to estimate "the" median ν of F, where ν is defined as a value satisfying $P(X \le \nu) \ge \frac{1}{2}$ and $P(X \ge \nu) \ge \frac{1}{2}$.

(a) Find the MSE of $\widehat{X}$ when

(i) F is discrete with $P(X = a) = P(X = c) = p$, $P(X = b) = 1 - 2p$, $0 < p < 1$, $a < b < c$.

Hint: Use Problem 1.3.5. The answer is $MSE(\widehat{X}) = [(a-b)^2 + (c-b)^2]P(S \ge k)$ where $k = .5(n+1)$ and $S \sim \mathcal{B}(n, p)$.

(ii) F is uniform, $\mathcal{U}(0, 1)$.

Hint: See Problem B.2.9.

(iii) F is normal, $\mathcal{N}(0, 1)$, $n = 1, 5, 25, 75$.

Hint: See Problem B.2.13. Use a numerical integration package.

(b) Compute the relative risk $RR = MSE(\widehat{X})/MSE(\bar{X})$ in question (i) when $b = 0$, $a = -\Delta$, $b = \Delta$, $p = .20, .40$, and $n = 1, 5, 15$.

(c) Same as (b) except when $n = 15$, plot RR for $p = .1, .2, .3, .4, .45$.

(d) Find $E|\widehat{X} - b|$ for the situation in (i). Also find $E|\bar{X} - b|$ when $n = 1$, and 2 and compare it to $E|\widehat{X} - b|$.

(e) Compute the relative risks $MSE(\widehat{X})/MSE(\bar{X})$ in questions (ii) and (iii).

7. Let $X_1, \dots, X_n$ be a sample from a population with values

$$\theta - 2\Delta, \theta - \Delta, \theta, \theta + \Delta, \theta + 2\Delta;\ \Delta > 0.$$

Each value has probability .2. Let $\bar{X}$ and $\widehat{X}$ denote the sample mean and median. Suppose that n is odd.

(a) Find $MSE(\widehat{X})$ and the relative risk $RR = MSE(\widehat{X})/MSE(\bar{X})$.

(b) Evaluate RR when $n = 1, 3, 5$.

Hint: By Problem 1.3.5, set $\theta = 0$ without loss of generality. Next note that the distribution of $\widehat{X}$ involves Bernoulli and multinomial trials.

8. Let $X_1, \dots, X_n$ be a sample from a population with variance σ^2, $0 < \sigma^2 < \infty$.

(a) Show that $s^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is an unbiased estimator of σ^2.
Hint: Write $(X_i - \bar{X})^2 = ([X_i - \mu] - [\bar{X} - \mu])^2$, then expand $(X_i - \bar{X})^2$ keeping the square brackets intact.

(b) Suppose $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

(i) Show that $MSE(s^2) = 2(n-1)^{-1}\sigma^4$.

(ii) Let $\widehat{\sigma}_0^2 = c\sum_{i=1}^n (X_i - \bar{X})^2$. Show that the value of c that minimizes $MSE(\widehat{\sigma}_c^2)$ is $c = (n+1)^{-1}$.

Hint for question (b): Recall (Theorem B.3.3) that $\sigma^{-2}\sum_{i=1}^n (X_i - \bar{X})^2$ has a χ_{n-1}^2 distribution. You may use the fact that $E(X_i - \mu)^4 = 3\sigma^2$.

9. Let θ denote the proportion of people working in a company who have a certain characteristic (e.g., being left-handed). It is known that in the state where the company is located, 10% have the characteristic. A person in charge of ordering equipment needs to estimate θ and uses

$$\widehat{\theta} = (.2)(.10) + (.8)\widehat{p}$$

where $\widehat{p} = X/n$ is the proportion with the characteristic in a sample of size n from the company. Find $MSE(\widehat{\theta})$ and $MSE(\widehat{p})$. If the true θ is θ_0, for what θ_0 is

$$MSE(\widehat{\theta})/MSE(\widehat{p}) < 1?$$

Give the answer for $n = 25$ and $n = 100$.

10. In Problem 1.3.3(a) with $b = c = 1$ and $n = 1$, suppose θ is discrete with frequency function $\pi(0) = \pi\left(-\frac{1}{2}\right) = \pi\left(\frac{1}{2}\right) = \frac{1}{3}$. Compute the Bayes risk of $\delta_{r,s}$ when

(a) $r = -s = -1$

(b) $r = -\frac{1}{2}s = -1$.

Which one of the rules is the better one from the Bayes point of view?

11. A decision rule δ is said to be *unbiased* if

$$E_\theta(l(\theta, \delta(\mathbf{X}))) \le E_\theta(l(\theta', \delta(\mathbf{X})))$$

for all $\theta, \theta' \in \Theta$.

(a) Show that if θ is real and $l(\theta, a) = (\theta - a)^2$, then this definition coincides with the definition of an unbiased estimate of θ.

(b) Show that if we use the $0-1$ loss function in testing, then a test function is unbiased in this sense if, and only if, the *power function*, defined by $\beta(\theta, \delta) = E_\theta(\delta(\mathbf{X}))$, satisfies

$$\beta(\theta', \delta) \ge \sup\{\beta(\theta, \delta) : \theta \in \Theta_0\},$$

for all $\theta' \in \Theta_1$.

12. In Problem 1.3.3, show that if $c \leq b$, $z > 0$, and

$$r = -s = -z\left(\frac{b}{b+c}\right)\Big/\sqrt{n},$$

then $\delta_{r,s}$ is unbiased

13. A (behavioral) *randomized test* of a hypothesis H is defined as any statistic $\varphi(\mathbf{X})$ such that $0 \leq \varphi(\mathbf{X}) \leq 1$. The interpretation of φ is the following. If $\mathbf{X} = \mathbf{x}$ and $\varphi(\mathbf{x}) = 0$ we decide Θ_0, if $\varphi(\mathbf{x}) = 1$, we decide Θ_1; but if $0 < \varphi(\mathbf{x}) < 1$, we perform a Bernoulli trial with probability $\varphi(\mathbf{x})$ of success and decide Θ_1 if we obtain a success and decide Θ_0 otherwise.

Define the nonrandomized test δ_u, $0 < u < 1$, by

$$\begin{aligned}\delta_u(\mathbf{X}) &= 1 \quad \text{if } \varphi(\mathbf{X}) \geq u \\ &= 0 \quad \text{if } \varphi(\mathbf{X}) < u.\end{aligned}$$

Suppose that $U \sim \mathcal{U}(0,1)$ and is independent of $\mathbf{X}$. Consider the following randomized test δ: Observe U. If $U = u$, use the test δ_u. Show that δ agrees with φ in the sense that,

$$P_\theta[\delta(\mathbf{X}) = 1] = 1 - P_\theta[\delta(\mathbf{X}) = 0] = E_\theta(\varphi(\mathbf{X})).$$

14. *Convexity of the risk set.* Suppose that the set of decision procedures is finite. Show that if δ_1 and δ_2 are two randomized procedures, then, given $0 < \alpha < 1$, there is a randomized procedure δ_3 such that $R(\theta, \delta_3) = \alpha R(\theta, \delta_1) + (1-\alpha)R(\theta, \delta_2)$ for all θ.

15. Suppose that $P_{\theta_0}(B) = 0$ for some event B implies that $P_\theta(B) = 0$ for all $\theta \in \Theta$. Further suppose that $l(\theta_0, a_0) = 0$. Show that the procedure $\delta(X) \equiv a_0$ is admissible.

16. In Example 1.3.4, find the set of μ where $MSE(\widehat{\mu}) \leq MSE(\bar{X})$. Your answer should depend on n, σ^2 and $\delta = |\mu - \mu_0|$.

17. In Example 1.3.4, consider the estimator

$$\widehat{\mu}_w = w\mu_0 + (1-w)\bar{X}.$$

If n, σ^2 and $\delta = |\mu - \mu_0|$ are known,

(a) find the value of w_0 that minimizes $MSE(\widehat{\mu}_w)$,

(b) find the minimum relative risk of $\widehat{\mu}_{w_0}$ to $\bar{X}$.

18. For Example 1.1.1, consider the loss function (1.3.1) and let δ_k be the decision rule "reject the shipment iff $X \geq k$."

(a) Show that the risk is given by (1.3.7).

(b) If $N = 10$, $s = r = 1$, $\theta_0 = .1$, and $k = 3$, plot $R(\theta, \delta_k)$ as a function of θ.

(c) Same as (b) except $k = 2$. Compare δ_2 and δ_3.

19. Consider a decision problem with the possible states of nature θ_1 and θ_2, and possible actions a_1 and a_2. Suppose the loss function $\ell(\theta, a)$ is