It is also seen that equality holds in (6.29) if and only if δ is a linear function of the ψ's in S (Problem 6.12). The problem of whether the Bhattacharyya bounds become sharp as $s \to \infty$ has been investigated for some one-parameter cases by Blight and Rao (1974).

(b) A different kind of extension avoids the need for regularity conditions by considering differences instead of derivatives. (See Hammersley 1950, Chapman and Robbins 1951, Kiefer 1952, Fraser and Guttman 1952, Fend 1959, Sen and Ghosh 1976, Chatterji 1982, and Klaassen 1984, 1985.)

(c) Applications of the inequality to the sequential case in which the number of observations is not a fixed integer but a random variable, say N, determined from the observations is provided by Wolfowitz (1947), Blackwell and Girshick (1947), and Seth (1949). Under suitable regularity conditions, (6.23) then continues to hold with n replaced by $E_\theta(N)$; see also Simons 1980, Govindarajulu and Vincze 1989, and Stefanov 1990.

(d) Other extensions include arbitrary convex loss functions (Kozek 1976); weighted loss functions (Mikulski and Monsour 1988); to the case that g and δ are vector-valued (Rao 1945, Cramér 1946b, Seth 1949, Shemyakin 1987, and Rao 1992); to nonparametric problems (Vincze 1992); location problems (Klaassen 1984); and density estimation (Brown and Farrell 1990).

7 Problems

Section 1

1.1 Verify (a) that (1.4) defines a probability distribution and (b) condition (1.5).

1.2 In Example 1.5, show that a_i^* minimizes (1.6) for $i = 0, 1$, and simplify the expression for a_0^*. [*Hint*: $\Sigma\kappa p^{\kappa-1}$ and $\Sigma\kappa(\kappa-1)p^{\kappa-2}$ are the first and second derivatives of $\Sigma p^\kappa = 1/q$.]

1.3 Let X take on the values $-1, 0, 1, 2, 3$ with probabilities $P(X = -1) = 2pq$ and $P(X = k) = p^k q^{3-k}$ for $k = 0, 1,2,3$.

(a) Check that this is a probability distribution.

(b) Determine the LMVU estimator at p_0 of (i) p, and (ii) pq, and decide for each whether it is UMVU.

1.4 For a sample of size n, suppose that the estimator $T(\mathbf{x})$ of $\tau(\theta)$ has expectation

$$E[T(\mathbf{X})] = \tau(\theta) + \sum_{k=1}^{\infty} \frac{a_k}{n^k},$$

where a_k may depend on θ but not on n.

(a) Show that the expectation of the jackknife estimator T_J of (1.3) is

$$E[T_J(\mathbf{X})] = \tau(\theta) - \frac{a_2}{n^2} + O(1/n^3).$$

(b) Show that if var $T \sim c/n$ for some constant c, then var $T_J \sim c'/n$ for some constant c'. Thus, the jackknife will reduce bias and not increase variance.

A *second-order jackknife* can be defined by jackknifing T_J, and this will result in further bias reduction, but may not maintain a variance of the same order (Robson and Whitlock 1964; see also Thorburn 1976 and Note 8.3).

1.5 (a) Any two random variables X and Y with finite second moments satisfy the covariance inequality $[\text{cov}(X, Y)]^2 \leq \text{var}(X) \cdot \text{var}(Y)$.

(b) The inequality in part (a) is an equality if and only if there exist constants a and b for which $P(X = aY + b) = 1$.

[*Hint*: Part (a) follows from the Schwarz inequality (Problem 1.7.20) with $f = X - E(X)$ and $g = Y - E(Y)$.]

1.6 An alternative proof of the Schwarz inequality is obtained by noting that

$$\int (f + \lambda g)^2 dP = \int f^2 dP + 2\lambda \int fg\, dP + \lambda^2 \int g^2 dP \geq 0 \quad \text{for all } \lambda,$$

so that this quadratic in λ has at most one root.

1.7 Suppose X is distributed on (0, 1) with probability density $p_\theta(x) = (1-\theta) + \theta/2\sqrt{x}$ for all $0 < x < 1, 0 \leq \theta \leq 1$. Show that there does not exist an LMVU estimator of θ. [*Hint*: Let $\delta(x) = a[x^{-1/2} + b]$ for $c < x < 1$ and $\delta(x) = 0$ for $0 < x < c$. There exist values a and b, and c such that $E_0(\delta) = 0$ and $E_1(\delta) = 1$ (and δ is unbiased) and that $E_0(\delta^2)$ is arbitrarily close to zero (Stein 1950).]

1.8 If δ and δ' have finite variance, so does $\delta' - \delta$. [*Hint*: Problem 1.5.]

1.9 In Example 1.9, (a) determine all unbiased estimators of zero; (b) show that no nonconstant estimator is UMVU.

1.10 If estimators are restricted to the class of linear estimators, characterization of best unbiased estimators is somewhat easier. Although the following is a consequence of Theorem 1.7, it should be established without using that theorem.

Let $\mathbf{X}_{p\times 1}$ satisfy $E(\mathbf{X}) = B\psi$ and $\text{var}(\mathbf{X}) = I$, where $B_{p\times r}$ is known, and $\psi_{r\times 1}$ is unknown. A *linear estimator* is an estimator of the form $a'\mathbf{X}$, where $a_{p\times 1}$ is a known vector. We are concerned with the class of estimators

$$\mathcal{D} = \{\delta(\mathbf{x}) : \delta(\mathbf{x}) = a'\mathbf{x}, \text{ for some known vector } a\}.$$

(a) For a known vector c, show that the estimators in $\mathcal{D}$ that are unbiased estimators of $c'\psi$ satisfy $a'B = c'$.

(b) Let $\mathcal{D}_c = \{\delta(\mathbf{x}) : \delta(\mathbf{x}) = a'\mathbf{x}, a'B = c'\}$ be the class of linear unbiased estimators of $c'\psi$. Show that the *best linear unbiased estimator* (BLUE) of $c'\psi$, the linear unbiased estimator with minimum variance, is $\delta^*(\mathbf{x}) = a^{*\prime}\mathbf{x}$, where $a^{*\prime} = a'B(B'B)^{-1}B'$ and $a^{*\prime}B = c$ with variance $\text{var}(\delta^*) = c'c$.

(c) Let $\mathcal{D}_0 = \{\delta(\mathbf{x}) : \delta(\mathbf{x}) = a'\mathbf{x}, a'B = 0.\}$ be the class of linear unbiased estimators of zero. Show that if $\delta \in \mathcal{D}_0$, then $\text{cov}(\delta, \delta^*) = 0$.

(d) Hence, establish the analog of Theorem 1.7 for linear estimators:

Theorem. *An estimator* $\delta^* \in \mathcal{D}_c$ *satisfies* $\text{var}(\delta^*) = \min_{\delta\in\mathcal{D}_c} \text{var}(\delta)$ *if and only if* $\text{cov}(\delta^*, U) = 0$, *where* U *is any estimator in* $\mathcal{D}_0$.

(e) Show that the results here can be directly extended to the case of $\text{var}(\mathbf{X}) = \Sigma$, where $\Sigma_{p\times p}$ is a known matrix, by considering the transformed problem with $\mathbf{X}^* = \Sigma^{1/2}\mathbf{X}$ and $B^* = \Sigma^{1/2}B$.

1.11 Use Theorem 1.7 to find UMVU estimators of some of the $\eta_\theta(d_i)$ in the dose-response model (1.6.16), with the restriction (1.6.17) (Messig and Strawderman 1993). Let the classes Δ and $\mathcal{U}$ be defined as in Theorem 1.7.

(a) Show that an estimator $U \in \mathcal{U}$ if and only if $U(x_1, x_2) = a[I(x_1 = 0) - I(x_2 = 0)]$ for an arbitrary constant $a < \infty$.

(b) Using part (a) and (1.7), show that an estimator δ is UMVU for its expectation only if it is of the form $\delta(x_1, x_2) = aI_{(0,0)}(x_1, x_2) + bI_{(0,1),(1,0),(2,0)}(x_1, x_2) + cI_{(1,1)}(x_1, x_2) + dI_{(2,1)}(x_1, x_2)$ where a, b, c, and d are arbitrary constants.

(c) Show that there does not exist a UMVU estimator of $\eta_\theta(d_1) = 1 - e^{-\theta}$, but the UMVU estimator of $\eta_\theta(d_2) = 1 - e^{-2\theta}$ is $\delta(x_1, x_2) = 1 - \frac{1}{2}[I(x_1 = 0) + I(x_2 = 0)]$.

(d) Show that the LMVU estimator of $1 - e^{-\theta}$ is $\delta(x_1, x_2) = \frac{x_1}{2} + \frac{1}{2(1+e^{-\theta})}[I(x_1 = 0) - I(x_2 = 0)]$.

1.12 Show that if $\delta(X)$ is a UMVU estimator of $g(\theta)$, it is the unique UMVU estimator of $g(\theta)$. (Do not assume completeness, but rather use the covariance inequality and the conditions under which it is an equality.)

1.13 If δ_1 and δ_2 are in Δ and are UMVU estimators of $g_1(\theta)$ and $g_2(\theta)$, respectively, then $a_1\delta_1 + a_2\delta_2$ is also in Δ and is UMVU for estimating $a_1g_1(\theta) + a_2g_2(\theta)$, for any real a_1 and a_2.

1.14 Completeness of T is not only sufficient but also necessary so that every $g(\theta)$ that can be estimated unbiasedly has only one unbiased estimator that is a function of T.

1.15 Suppose $X_1, \ldots, X_n$ are iid Poisson (λ).

(a) Show that $\bar{X}$ is the UMVU estimator for λ.

(b) For $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, we have that $ES^2 = E\bar{X} = \lambda$. To directly establish that var $S^2 >$ var $\bar{X}$, prove that $E(S^2|\bar{X}) = \bar{X}$.

Note: The identity $E(S^2|\bar{X}) = \bar{X}$ shows how completeness can be used in calculating conditional expectations.

1.16 (a) If $X_1, \ldots, X_n$ are iid (not necessarily normal) with $\text{var}(X_i) = \sigma^2 < \infty$, show that $\delta = \Sigma(X_i - \bar{X})^2/(n-1)$ is an unbiased estimator of σ^2.

(b) If the X_i take on the values 1 and 0 with probabilities p and $q = 1 - p$, the estimator δ of (a) depends only on $T = \Sigma X_i$ and hence is UMVU for estimating $\sigma^2 = pq$. Compare this result with that of Example 1.13.

1.17 If T has the binomial distribution $b(p, n)$ with $n > 3$, use Method 1 to find the UMVU estimator of p^3.

1.18 Let $X_1, \ldots, X_n$ be iid according to the Poisson distribution $P(\lambda)$. Use Method 1 to find the UMVU estimator of (a) λ^k for any positive integer k and (b) $e^{-\lambda}$.

1.19 Let $X_1, \ldots, X_n$ be distributed as in Example 1.14. Use Method 1 to find the UMVU estimator of θ^k for any integer $k > -n$.

1.20 Solve Problem 1.18(b) by Method 2, using the fact that an unbiased estimator of $e^{-\lambda}$ is $\delta = 1$ if $X_1 = 0$, and $\delta = 0$ otherwise.

1.21 In n Bernoulli trials, let $X_i = 1$ or 0 as the ith trial is a success or failure, and let $T = \Sigma X_i$. Solve Problem 1.17 by Method 2, using the fact that an unbiased estimator of p^3 is $\delta = 1$ if $X_1 = X_2 = X_3 = 1$, and $\delta = 0$ otherwise.

1.22 Let X take on the values 1 and 0 with probability p and q, respectively, and assume that $1/4 < p < 3/4$. Consider the problem of estimating p with loss function $L(p, d) = 1$ if $|d - p| \geq 1/4$, and 0 otherwise. Let δ^* be the randomized estimator which is Y_0 or Y_1 when $X = 0$ or 1 where Y_0 and Y_1 are distributed as $U(-1/2, 1/2)$ and $U(1/2, 3/2)$, respectively.

(a) Show that δ^* is unbiased.

(b) Compare the risk function of δ^* with that of X.

Section 2

2.1 If $X_1, \ldots, X_n$ are iid as $N(\xi, \sigma^2)$ with σ^2 known, find the UMVU estimator of (a) ξ^2, (b) ξ^3, and (c) ξ^4. [*Hint*: To evaluate the expectation of $\bar{X}^k$, write $\bar{X} = Y + \xi$, where Y is $N(0, \sigma^2/n)$ and expand $E(Y + \xi)^k$.]

2.2 Solve the preceding problem when σ is unknown.

2.3 In Example 2.1 with σ known, let $\delta = \Sigma c_i X_i$ be any linear estimator of ξ. If δ is biased, show that its risk $E(\delta - \xi)^2$ is unbounded. [*Hint*: If $\Sigma c_i = 1 + k$, the risk is $\geq k^2 \xi^2$.]

2.4 Suppose, as in Example 2.1, that $X_1, \ldots, X_n$ are iid as $N(\xi, \sigma^2)$, with one of the parameters known, and that the estimand is a polynomial in ξ or σ. Then, the UMVU estimator is a polynomial in $\bar{X}$ or $S^2 = \sum(X_i - \xi)^2$. The variance of any such polynomial can be estimated if one knows the moments $E(\bar{X}^k)$ and $E(S^k)$ for all $k = 1, 2, \ldots$. To determine $E(\bar{X}^k)$, write $\bar{X} = Y + \xi$, where Y is distributed as $N(0, \sigma^2/n)$. Show that

(a)

$$E(\bar{X}^k) = \sum_{r=0}^{k} \binom{k}{r} \xi^{k-r} E(Y^r)$$

with

$$E(Y^r) = \begin{cases} (r-1)(r-3)\cdots 3 \cdot 1(\sigma^2/n)^{r/2} & \text{when } r \geq 2 \text{ is even} \\ 0 & \text{when } r \text{ is odd} . \end{cases}$$

(b) As an example, consider the UMVU estimator S^2/n of σ^2. Show that $E(S^4) = n(n+2)\sigma^2$ and $\operatorname{var}\left(\frac{S^2}{n}\right) = \frac{2\sigma^4}{n}$ and that the UMVU estimator of this variance is $2S^4/n^2(n+2)$.

2.5 In Example 2.1, when both parameters are unknown, show that the UMVU estimator of ξ^2 is given by $\delta = \bar{X}^2 - \frac{S^2}{n(n-1)}$ where now $S^2 = \sum(X_i - \bar{X})^2$.

2.6 (a) Determine the variance of the estimator Problem 2.5.

(b) Find the UMVU estimator of the variance in part (a).

2.7 If X is a single observation from $N(\xi, \sigma^2)$, show that no unbiased estimator δ of σ^2exists when ξ is unknown. [*Hint*: For fixed $\sigma = a$, X is a complete sufficient statistic for ξ, and $E[\delta(X)] = a^2$ for all ξ implies $\delta(x) = a^2$ a.e.]

2.8 Let $X_i, i = 1, \ldots, n$, be independently distributed as $N(\alpha + \beta t_i, \sigma^2)$ where α, β, and σ^2 are unknown, and the t's are known constants that are not all equal. Find the UMVU estimators of α and β.

2.9 In Example 2.2 with $n = 1$, the UMVU estimator of p is the indicator of the event $X_1 \leq u$ whether σ is known or unknown.

2.10 Verify Equation (2.14), the density of $(X_1 - \bar{X})/S$ in normal sampling. [The UMVU estimator in (2.13) is used by Kiefer (1977) as an example of his estimated confidence approach.]

2.11 Assuming (2.15) with $\sigma = \tau$, determine the UMVU estimators of σ^2 and $(\eta - \xi)/\sigma$.

2.12 Assuming (2.15) with $\eta = \xi$ and $\sigma^2/\tau^2 = \gamma$, show that when γ is known:

(a) T' defined in Example 2.3(iii) is a complete sufficient statistic;

(b) δ_γ is UMVU for ξ.

2.13 Show that in the preceding problem with γ unknown,

(a) a UMVU estimator of ξ does not exist;

(b) the estimator $\hat{\xi}$ is unbiased under the conditions stated in Example 2.3. [*Hint*: (i) Problem 2.12(b) and the fact that δ_γ is unbiased for ξ even when $\sigma^2/\tau^2 \neq \gamma$. (ii) Condition on (S_X, S_Y).]

2.14 For the model (2.15) find the UMVU estimator of $P(X_1 < Y_1)$ when (a) $\sigma = \tau$ and (b) when σ and τ are arbitrary. [*Hint*: Use the conditional density (2.13) of X_1 given $\bar{X}$, S_X^2 and that of Y_1 given $\bar{Y}$, S_Y^2 to determine the conditional density of $Y_1 - X_1$ given $\bar{X}, \bar{Y}, S_X^2$, and S_Y^2.]

2.15 If $(X_1, Y_1), \ldots, (X_n, Y_n)$ are iid according to any bivariate distribution with finite second moments, show that $S_{XY}/(n-1)$ given by (2.17) is an unbiased estimator of $\text{cov}(X_i, Y_i)$.

2.16 In a sample size $N = n + k + 1$, some of the observations are missing. Assume that $(X_i, Y_i), i = 1, \ldots, n$, are iid according to the bivariate normal distribution (2.16), and that $U_1, \ldots, U_k$ and $V_1, \ldots, V_l$ are independent $N(\xi, \sigma^2)$ and $N(\eta, \tau^2)$, respectively.

(a) Show that the minimal sufficient statistics are complete when ξ and η are known but not when they are unknown.

(b) When ξ and η are known, find the UMVU estimators for σ^2, τ^2, and $\rho\sigma\tau$, and suggest reasonable unbiased estimators for these parameters when ξ and η are unknown.

2.17 For the family (2.22), show that the UMVU estimator of a when b is known and the UMVU estimator of b is known are as stated in Example 2.5. [*Hint*: Problem 6.18.]

2.18 Show that the estimators (2.23) are UMVU. [*Hint*: Problem 1.6.18.].

2.19 For the family (2.22) with $b = 1$, find the UMVU estimator of $P(X_1 \geq u)$ and of the density $e^{-(u-a)}$ of X_1 at u. [*Hint*: Obtain the estimator $\delta(X_{(1)})$ of the density by applying Method 2 of Section 2.1 and then the estimator of the probability by integration. Alternatively, one can first obtain the estimator of the probability as $P(X_1 \geq u | X_{(1)})$ using the fact that $X_1 - X_{(1)}$ is ancillary and that given $X_{(1)}$, X_1 is either equal to $X_{(1)}$ or distributed as $E(X_{(1)}, 1)$.]

2.20 Find the UMVU estimator of $P(X_1 \geq u)$ for the family (2.22) when both a and b are unknown.

2.21 Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independently distributed as $E(a, b)$ and $E(a', b')$, respectively.

(a) If a, b, a', and b' are completely unknown, $X_{(1)}, Y_{(1)}, \Sigma[X_i - X_{(1)}]$, and $\Sigma[Y_j - Y_{(1)}]$ jointly are sufficient and complete.

(b) Find the UMVU estimators of $a' - a$ and b'/b.

2.22 In the preceding problem, suppose that $b' = b$.

(a) Show that $X_{(1)}, Y_{(1)}$, and $\Sigma[X_i - X_{(1)}] + \Sigma[Y_j - Y_{(1)}]$ are sufficient and complete.

(b) Find the UMVU estimators of b and $(a' - a)/b$.

2.23 In Problem 2.21, suppose that $a' = a$.

(a) Show that the complete sufficient statistic of Problem 2.21(a) is still minimal sufficient but no longer complete.

(b) Show that a UMVU estimator for $a' = a$ does not exist.

(c) Suggest a reasonable unbiased estimator for $a' = a$.

2.24 Let $X_1, \ldots, X_n$ be iid according to the uniform distribution $U(\xi - b, \xi + b)$. If ξ, b are both unknown, find the UMVU estimators of ξ, b, and ξ/b. [*Hint*: Problem 1.6.30.]

2.25 Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be iid as $U(0, \theta)$ and $U(0, \theta')$, respectively. If $n > 1$, determine the UMVU estimator of θ/θ'.

2.26 Verify the ML estimators given in (2.24).

2.27 In Example 2.6(b), show that

(a) The bias of the ML estimator is 0 when $\xi = u$.

(b) At $\xi = u$, the ML estimator has smaller expected squared error than the UMVU estimator.

[Hint: In (b), note that $u - \bar{X}$ is always closer to 0 than $\sqrt{\frac{n}{n-1}}(u - \bar{X})$.]

2.28 Verify (2.26).

2.29 Under the assumptions of Lemma 2.7, show that:

(a) If b is replaced by any random variable B which is independent of X and not 0 with probability 1, then $R_\delta(\theta) < R_{\delta^*}(\theta)$.

(b) If squared error is replaced by any loss function of the form $L(\theta, \delta) = \rho(d - \theta)$ and δ is risk unbiased with respect to L, then $R_\delta(\theta) < R_{\delta^*}(\theta)$.

Section 3

3.1 (a) In Example 3.1, show that $\Sigma(X_i - \bar{X})^2 = T(n - T)/n$.

(b) The variance of $T(n-T)/n(n-1)$ in Example 3.1 is $(pq/n)[(q-p)^2 + 2pq/(n-1)]$.

3.2 If T is distributed as $b(p, n)$, find an unbiased estimator $\delta(T)$ of p^m $(m \le n)$ by Method 1, that is, using (1.10). [*Hint*: Example 1.13.]

3.3 (a) Use the method leading to (3.2) to find the UMVU estimator $\pi_k(T)$ of $P[X_1 + \cdots + X_m = k] = \binom{m}{k} p^k q^{m-k}$ $(m \le n)$.

(b) For fixed t and varying k, show that the $\pi_k(t)$ are the probabilities of a hypergeometric distribution.

3.4 If Y is distributed according to (3.3), use Method 1 of Section 2.1

(a) to show that the UMVU estimator of p^r $(r < m)$ is

$$\delta(y) = \frac{(m - r + y - 1)(m - r + y - 2)\ldots(m - r)}{(m + y - 1)(m + y - 2)\cdots m},$$

and hence in particular that the UMVU estimator of $1/p$, $1/p^2$ and p are, respectively, $(m + y)/m$, $(m + y)(m + y + 1)/m(m + 1)$, and $(m - 1)/(m + y - 1)$;

(b) to determine the UMVU estimator of $\text{var}(Y)$;

(c) to show how to calculate the UMVU estimator δ of $\log p$.

3.5 Consider the scheme in which binomial sampling is continued until at least a successes *and* b failures have been obtained. Show how to calculate a reasonable estimator of $\log(p/q)$. [*Hint*: To obtain an unbiased estimator of $\log p$, modify the UMVU estimator δ of Problem 3.4(c).]

3.6 If binomial sampling is continued until m successes have been obtained, let X_i $(i = 1, \ldots, m)$ be the number of failures between the $(i - 1)$st and ith success.

(a) The X_i are iid according to the *geometric distribution* $P(X_i = x) = pq^x$, $x = 0, 1, \ldots$.

(b) The statistic $Y = \Sigma X_i$ is sufficient for $(X_1, \ldots, X_m)$ and has the distribution (3.3).

3.7 Suppose that binomial sampling is continued until the number of successes equals the number of failures.

(a) This rule is closed if $p = 1/2$ but not otherwise.
(b) If $p = 1/2$ and N denotes the number of trials required, $E(N) = \infty$.

3.8 Verify Equation (3.7) with the appropriate definition of $N'(x, y)$ (a) for the estimation of p and (b) for the estimation of $p^a q^b$.

3.9 Consider sequential binomial sampling with the stopping points (0, 1) and $(2, y)$, $y = 0, 1, \ldots$. (a) Show that this plan is closed and simple. (b) Show that (X, Y) is not complete by finding a nontrivial unbiased estimator of zero.

3.10 In Example 3.4(ii), (a) show that the plan is closed but not simple, (b) show that (X, Y) is not complete, and (c) evaluate the unbiased estimator (3.7) of p.

3.11 *Curtailed single sampling*. Let $a, b < n$ be three non-negative integers. Continue observation until either a successes, b failures, or n observations have been obtained. Determine the UMVU estimator of p.

3.12 For any sequential binomial sampling plan, the coordinates (X, Y) of the end point of the sample path are minimal sufficient.

3.13 Consider any closed sequential binomial sampling plan with a set B of stopping points, and let B' be the set $B \cup \{(x_0, y_0)\}$ where (x_0, y_0) is a point not in B that has positive probability of being reached under plan B. Show that the sufficient statistic $T = (X, Y)$ is not complete for the sampling plan which has B' as its set of stopping points. [*Hint*: For any point $(x, y) \in B$, let $N(x, y)$ and $N'(x, y)$ denote the number of paths to (x, y) when the set of stopping points is B and B', respectively, and let $N(x_0, y_0) = 0$, $N'(x_0, y_0) = 1$. Then, the statistic $1 - [N(X, Y)/N'(X, Y)]$ has expectation 0 under B' for all values of p.]

3.14 For any sequential binomial sampling plan under which the point (1, 1) is reached with positive probability but is not a stopping point, find an unbiased estimator of pq depending only on (X, Y). Evaluate this estimator for

(a) taking a sample of fixed size $n > 2$;
(b) inverse binomial sampling.

3.15 Use (3.3) to determine $A(t, n)$ in (3.11) for the negative binomial distribution with $m = n$, and evaluate the estimators (3.13) of q^r, and (3.14).

3.16 Consider n binomial trials with success probability p, and let r and s be two positive integers with $r + s < n$. To the boundary $x + y = n$, add the boundary point (r, s), that is, if the number of successes in the first $r + s$ trials is exactly r, the process is stopped and the remaining $n - (r + s)$ trials are not performed.

(a) Show that U is an unbiased estimator of zero if and only if $U(k, n - k) = 0$ for $k = 0, 1, \ldots, r - 1$ and $k = n - s + 1, n - s + 2, \ldots, n$, and $U(k, n - k) = c_k U(r, s)$ for $k = r, \ldots, n - s$, where the c's are given constants $\neq 0$.
(b) Show that δ is the UMVU estimator of its expectation if and only if

$$\delta(k, n - k) = \delta(r, s) \quad \text{for } k = r, \ldots, n - s.$$

3.17 Generalize the preceding problem to the case that two points (r_1, s_1) and (r_2, s_2) with $r_i + s_i < n$ are added to the boundary. Assume that these two points are such that all $n + 1$ points $x + y = n$ remain boundary points. [*Hint*: Distinguish the three cases that the intervals (r_1, s_1) and (r_2, s_2) are (i) mutually exclusive, (ii) one contained in the other, and (iii) overlapping but neither contained in the other.]

3.18 If X has the Poisson distribution $P(\theta)$, show that $1/\theta$ does not have an unbiased estimator.

3.19 If $X_1, \ldots, X_n$ are iid according to (3.18), the Poisson distribution truncated on the left at 0, find the UMVU estimator of θ when (a) $n = 1$ and (b) $n = 2$.

3.20 Let $X_1, \ldots, X_n$ be a sample from the Poisson distribution truncated on the left at 0, i.e., with sample space $\mathcal{X} = \{1, 2, 3, \ldots\}$.

(a) For $t = \Sigma x_i$, the UMVU estimator of λ is (Tate and Goen 1958) $\hat{\lambda} = \frac{C^n_{t-1}}{C^n_t} t$ where $C^n_t = \frac{(-1)^n}{n!} \sum_{k=0}^{\infty} \binom{n}{k} (-1)^k k^t$ is a *Stirling number of the second kind.*

(b) An alternate form of the UMVU estimator is $\hat{\lambda} = \frac{t}{n}\left(1 - \frac{C^{n-1}_{t-1}}{C^n_t}\right)$. [Hint: Establish the identity $C^n_t = C^{n-1}_{t-1} + nC^n_{t-1}$.]

(c) The Cramér-Rao lower bound for the variance of unbiased estimators of λ is $\lambda(1 - e^{-\lambda})^2/[n(1 - e^{-\lambda} - \lambda e^{-\lambda})]$, and it is not attained by the UMVU estimator. (It is, however, the asymptotic variance of the ML estimator.)

3.21 Suppose that X has the Poisson distribution truncated on the right at a, so that it has the conditional distribution of Y given $Y \le a$, where Y is distributed as $P(\lambda)$. Show that λ does not have an unbiased estimator.

3.22 For the negative binomial distribution truncated at zero, evaluate the estimators (3.13) and (3.14) for $m = 1, 2$, and 3.

3.23 If $X_1, \ldots, X_n$ are iid $P(\lambda)$, consider estimation of $e^{-b\lambda}$, where b is known.

(a) Show that $\delta^* = (1 - b/n)^t$ is the UMVU estimator of $e^{-b\lambda}$.

(b) For $b > n$, describe the behavior of δ^*, and suggest why it might not be a reasonable estimator.

(The probability $e^{-b\lambda}$, for $b > n$, is that of an "unobservable" event, in that it can be interpreted as the probability of no occurrence in a time interval of length b. A number of such situations are described and analyzed in Lehmann (1983), where it is suggested that, in these problems, no reasonable estimator may exist.)

3.24 If $X_1, \ldots, X_n$ are iid according to the logarithmic series distribution of Problem 1.5.14, evaluate the estimators (3.13) and (3.14) for $n = 1, 2$, and 3.

3.25 For the multinomial distribution of Example 3.8,

(a) show that $p_0^{r_0} \cdots p_s^{r_s}$ has an unbiased estimator provided $r_0, \ldots, r_s$ are nonnegative integers with $\Sigma r_i \le n$;

(b) find the totality of functions that can be estimated unbiasedly;

(c) determine the UMVU estimator of the estimand of (a).

3.26 In Example 3.9 when $p_{ij} = p_{i+}p_{+j}$, determine the variances of the two unbiased estimators $\delta_0 = n_{ij}/n$ and $\delta_1 = n_{i+}n_{+j}/n^2$ of p_{ij}, and show directly that $\text{var}(\delta_0) > \text{var}(\delta_1)$ for all $n > 1$.

3.27 In Example 3.9, show that independence of A and B implies that $(n_{1+}, \ldots, n_{I+})$ and $(n_{+1}, \ldots, n_{+J})$ are independent with multinomial distributions as stated.

3.28 Verify (3.20).

3.29 Let X, Y, and g be such that $E[g(X, Y)|y]$ is independent of y. Then, $E[f(Y)g(X, Y)] = E[f(Y)]E[g(X, Y)]$, and hence $f(Y)$ and $g(X, Y)$ are uncorrelated, for all f.

3.30 In Example 3.10, show that the estimator δ_1 of p_{ijk} is unbiased for the model (3.20). [*Hint*: Problem 3.29.]

Section 4

4.1 Let $X_1, \ldots, X_n$ be iid with distribution F.

(a) Characterize the totality of functions $f(X_1, \ldots, X_n)$ which are unbiased estimators of zero for the class $\mathcal{F}_0$ of all distributions F having a density.

(b) Give one example of a nontrivial unbiased estimator of zero when (i) $n = 2$ and (ii) $n = 3$.

4.2 Let $\mathcal{F}$ be the class of all univariate distribution functions F that have a probability density function f and finite mth moment.

(a) Let $X_1, \ldots, X_n$ be independently distributed with common distribution $F \in \mathcal{F}$. For $n \geq m$, find the UMVU estimator of ξ^m where $\xi = \xi(F) = EX_i$.

(b) Show that for the case that $P(X_i = 1) = p$, $P(X_i = 0) = q$, $p + q = 1$, the estimator of (a) reduces to (3.2).

4.3 In the preceding problem, show that $1/\text{var}_F X_i$ does not have an unbiased estimator for any n.

4.4 Let $X_1, \ldots, X_n$ be iid with distribution $F \in \mathcal{F}$ where $\mathcal{F}$ is the class of all symmetric distributions with a probability density. There exists no UMVU estimator of the center of symmetry θ of F (if unbiasedness is required only for the distributions F for which the expectation of the estimator exists). [*Hint*: The UMVU estimator of θ when F is $U(\theta - 1/2, \theta + 1/2)$, which was obtained in Problem 2.24, is unbiased for all $F \in \mathcal{F}$; so is $\bar{X}$.]

4.5 If $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independently distributed according to F and $G \in \mathcal{F}_0$, defined in Problem 4.1, the order statistics $X_{(1)} < \cdots < X_{(m)}$ and $Y_{(1)} < \cdots < Y_{(n)}$ are sufficient and complete. [*Hint*: For completeness, generalize the second proof suggested in Problem 6.33.]

4.6 Under the assumptions of the preceding problem, find the UMVU estimator of $P(X_i < Y_j)$.

4.7 Under the assumptions of Problem 4.5, let $\xi = EX_i$ and $\eta = EY_j$. Show that $\xi^2\eta^2$ possesses an unbiased estimator if and only if $m \geq 2$ and $n \geq 2$.

4.8 Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid $F \in \mathcal{F}$, where $\mathcal{F}$ is the family of all distributions with probability density and finite second moments. Show that $\delta(X, Y) = \sum(X_i - \bar{X})(Y_i - \bar{Y})/(n-1)$ is UMVU for $\text{cov}(X, Y)$.

4.9 Under the assumptions of the preceding problem, find the UMVU estimator of

(a) $P(X_i \leq Y_i)$;

(b) $P(X_i \leq X_j$ and $Y_i \leq Y_j)$, $i \neq j$.

4.10 Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be iid with $F \in \mathcal{F}$, where $\mathcal{F}$ is the family of all bivariate densities. Show that the sufficient statistic T, which generalizes the order statistics to the bivariate case, is complete. [*Hint*: Generalize the second proof suggested in Problem 6.33. As an exponential family for (X, Y), take the densities proportional to $e^{Q(x,y)}$ where

$$Q(x, y) = (\theta_{01}x + \theta_{10}y) + (\theta_{02}x^2 + \theta_{11}xy + \theta_{20}y^2) + \cdots$$
$$+(\theta_{0n}x^n + \cdots + \theta_{n0}y^n) - x^{2n} - y^{2n}.]$$

Section 5

5.1 Under the assumptions of Problem 1.3, determine for each p_1, the value $L_V(p_1)$ of the LMVU estimator of p at p_1 and compare the function $L_V(p)$, $0 < p < 1$ with the variance $V_{p_0}(p)$ of the estimator which is LMVU at (a) $p_0 = 1/3$ and (b) $p_0 = 1/2$.

5.2 Determine the conditions under which equality holds in (5.1).

5.3 Verify $I(\theta)$ for the distributions of Table 5.1.

5.4 If X is normal with mean zero and standard deviation σ, determine $I(\sigma)$.

5.5 Find $I(p)$ for the negative binomial distribution.

5.6 If X is distributed as $P(\lambda)$, show that the information it contains about $\sqrt{\lambda}$ is independent of λ.

5.7 Verify the following statements, asserted by Basu (1988, Chapter 1), which illustrate the relationship between information, sufficiency, and ancillarity. Suppose that we let $I(\theta) = E_\theta\left[-\partial^2/\partial\theta^2 \log f(x|(\theta)\right]$ be the information in X about θ and let $J(\theta) = E_\theta\left[-\partial^2/\partial\theta^2 \log g(T|\theta)\right]$ be the information about θ contained in a statistic T, where $g(\cdot|\theta)$ is the density function of T. Define $\lambda(\theta) = I(\theta) - J(\theta)$, a measure of information lost by using T instead of X. Under suitable regularity conditions, show that

(a) $\lambda(\theta) \geq 0$ for all θ

(b) $\lambda(\theta) = 0$ if and only if T is sufficient for θ.

(c) If Y is ancillary but (T, Y) is sufficient, then $I(\theta) = E_\theta[J(\theta|Y)]$, where

$$J(\theta|y) = E_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log h(T|y,\theta)|Y = y\right]$$

and $h(t|y, \theta)$ is the conditional density of T given $Y = y$.

(Basu's "regularity conditions" are mainly concerned with interchange of integration and differentiation. Assume any such interchanges are valid.)

5.8 Find a function of θ for which the amount of information is independent of θ:

(a) for the gamma distribution $\Gamma(\alpha, \beta)$ with α known and with $\theta = \beta$;

(b) for the binomial distribution $b(p, n)$ with $\theta = p$.

5.9 For inverse binomial sampling (see Example 3.2):

(a) Show that the best unbiased estimator of p is given by $\delta^*(Y) = (m-1)/(Y+m-1)$.

(b) Show that the information contained in Y about p is $I(p) = m/p^2(1-p)$.

(c) Show that $\text{var}\delta^* > 1/I(p)$.

(The estimator δ^* can be interpreted as the success rate if we ignore the last trial, which we know must be a success.)

5.10 Show that (5.13) can be differentiated by differentiating under the integral sign when $p_\theta(x)$ is given by (5.24), for each of the distributions of Table 5.2. [*Hint*: Form the difference quotient and apply the dominated convergence theorem.]

5.11 Verify the entries of Table 5.2.

5.12 Evaluate (5.25) when f is the density of Student's t-distribution with ν degrees of freedom. [*Hint*: Use the fact that

$$\int_{-\infty}^{\infty}\frac{dx}{(1+x^2)^k} = \frac{\Gamma(1/2)\Gamma(k-1/2)}{\Gamma(k)}.]$$

5.13 For the distribution with density (5.24), show that $I(\theta)$ is independent of θ.

5.14 Verify (a) formula (5.25) and (b) formula (5.27).

5.15 For the location t density, calculate the information inequality bound for unbiased estimators of θ.

5.16 (a) For the scale family with density $(1/\theta)f(x/\theta)$, $\theta > 0$, the amount of information a single observation X has about θ is

$$\frac{1}{\theta^2}\int\left[\frac{yf'(y)}{f(y)}+1\right]^2 f(y)\,dy.$$

(b) Show that the information X contains about $\xi = \log\theta$ is independent of θ.

(c) For the Cauchy distribution C$(0, \theta)$, $I(\theta) = 1/(2\theta^2)$.

5.17 If $p_\theta(x)$ is given by 1.5.1 with $s = 1$ and $T(x) = \delta(x)$, show that var$[\delta(X)]$ attains the lower bound (5.31) and is the only estimator to do so. [*Hint*: Use (5.18) and (1.5.15).]

5.18 Show that if a given function $g(\theta)$ has an unbiased estimator, there exists an unbiased estimator δ which for all θ values attains the lower bound (5.1) for some $\psi(x, \theta)$ satisfying (5.2) if and only if $g(\theta)$ has a UMVU estimator δ_0. [*Hint*: By Theorem 5.1, $\psi(x, \theta) = \delta_0(x)$ satisfies (5.2). For any other unbiased δ, cov$(\delta - \delta_0, \delta_0) = 0$ and hence var$(\delta_0) = [\text{cov}(\delta, \delta_0)]^2/\text{var}(\delta_0)$, so that $\psi = \delta_0$ provides an attainable bound.] (Blyth 1974).

5.19 Show that if $E_\theta\delta = g(\theta)$, and var$(\delta)$ attains the information inequality bound (5.31), then

$$\delta(x) = g(\theta) + \frac{g'(\theta)}{I(\theta)}\frac{\partial}{\partial\theta}p_\theta(x).$$

5.20 If $E_\theta\delta = g(\theta)$, the information inequality lower bound is $IB(\theta) = [g'(\theta)]^2/I(\theta)$. If $\theta = h(\xi)$ where h is differentiable, show that $IB(\xi) = IB(\theta)$.

5.21 (Liu and Brown 1993) Let X be an observation from the normal mixture density

$$p_\theta(x) = \frac{1}{2\sqrt{2\pi}}\left\{e^{-(1/2)(x-\theta)^2} + e^{-(1/2)(x+\theta)^2}\right\}, \quad \theta \in \Omega,$$

where Ω is any neighborhood of zero. Thus, the random variable X is either $N(\theta, 1)$ or $N(-\theta, 1)$, each with probability 1/2. Show that $\theta = 0$ is a *singular point*, that is, if there exists an unbiased estimator of θ it will have infinite variance at $\theta = 0$.

5.22 Let $X_1, \ldots, X_n$ be a sample from the Poisson (λ) distribution truncated on the left at 0, i.e., with sample space $\mathcal{X} = \{1, 2, 3, \ldots\}$ (see Problem 3.20). Show that the Cramér-Rao lower bound for the variance of unbiased estimators of λ is

$$\frac{\lambda(1-e^{-\lambda})^2}{n(1-e^{-\lambda}-\lambda e^{-\lambda})}$$

and is not attained by the UMVU estimator. (It is, however, the asymptotic variance of the ML estimator.)

5.23 Let $X_1, \ldots, X_n$ be iid according to a density $p(x, \theta)$ which is positive for all x. Then, the variance of any unbiased estimator δ of θ satisfies

$$\text{var}_{\theta_0}(\delta) \geq \frac{(\theta-\theta_0)^2}{\left\{\int_{-\infty}^{\infty}\frac{[p(x,\theta)]^2}{p(x,\theta_0)}\right\}^n - 1}, \qquad \theta \neq \theta_0.$$

[*Hint*: Direct consequence of (5.6).]

5.24 If $X_1, \ldots, X_n$ are iid as $N(\theta, \sigma^2)$ where σ is known and θ is known to have one of the values $0, \pm 1, \pm 2, \ldots$, the inequality of the preceding problem shows that any unbiased estimator δ of the restricted parameter θ satisfies

$$\text{var}_{\theta_0}(\delta) \geq \frac{\Delta^2}{e^{n\Delta^2/\sigma^2} - 1}, \quad \Delta \neq 0,$$

where $\Delta = \theta - \theta_0$, and hence $\sup_{\Delta \neq 0} \text{var}_{\theta_0}(\delta) \geq 1/[e^{n/\sigma^2} - 1]$.

5.25 Under the assumptions of the preceding problem, let $\bar{X}^*$ be the integer closest to $\bar{X}$.

(a) The estimator $\bar{X}^*$ is unbiased for the restricted parameter θ.

(b) There exist positive constants a and b such that for all sufficiently large n, $\text{var}_\theta(\bar{X}^*) \leq ae^{-bn}$ for all integers θ.

[*Hint*: (b) One finds $P(\bar{X}^* = k) = \int_{I_k} \phi(t)\,dt$, where I_k is the interval $((k - \theta - 1/2)\sqrt{n}/\sigma, (k - \theta + 1/2)\sqrt{n}/\sigma)$, and hence

$$\text{var}(\bar{X}^*) \leq 4 \sum_{k=1}^{\infty} k \left\{ 1 - \Phi\left[\frac{\sqrt{n}}{\sigma} \left(k - \frac{1}{2} \right) \right] \right\}.$$

The result follows from the fact that for all $y > 0$, $1 - \Phi(y) \leq \phi(y)/y$. See, for example, Feller 1968, Chapter VII, Section 1. Note that $h(y) = \phi(y)/(1 - \Phi(y))$ is the *hazard function* for the standard normal distribution, so we have $h(y) \geq y$ for all $y > 0$. $(1 - \Phi(y))/\phi(y)$ is also known as *Mill's ratio* (see Stuart and Ord, 1987, Section 5.38.) Efron and Johnstone (1990) relate the hazard function to the information inequality].

Note. The surprising results of Problems 5.23–5.25 showing a lower bound and variance which decrease exponentially are due to Hammersley (1950), who shows that, in fact,

$$\text{var}(\bar{X}^*) \sim \sqrt{\frac{8\sigma^2}{\pi n}} e^{-n/8\sigma^2} \quad \text{as} \quad \frac{n}{\sigma^2} \to \infty.$$

Further results concerning the estimation of restricted parameters and properties of $\bar{X}^*$ are given in Khan (1973), Ghosh (1974), Ghosh and Meeden (1978), and Kojima, Morimoto, and Takeuchi (1982).

5.26 *Kiefer inequality.*

(a) Let X have density (with respect to μ) $p(x, \theta)$ which is > 0 for all x, and let Λ_1 and Λ_2 be two distributions on the real line with finite first moments. Then, any unbiased estimator δ of θ satisfies

$$\text{var}(\delta) \geq \frac{[\int \Delta\, d\Lambda_1(\Delta) - \int \Delta\, d\Lambda_2(\Delta)]^2}{\int \psi^2(x, \theta) p(x, \theta)\, d\mu(x)}$$

where

$$\psi(x, \theta) = \frac{\int_{\Omega_\theta} p(x, \theta + \Delta)[d\Lambda_1(\Delta) - d\Lambda_2(\Delta)]}{p(x, \theta)}$$

with $\Omega_\theta = \{\Delta : \theta + \Delta_\varepsilon \Omega\}$.

(b) If Λ_1 and Λ_2 assign probability 1 to $\Delta = 0$ and Δ, respectively, the inequality reduces to (5.6) with $g(\theta) = \theta$. [*Hint*: Apply (5.1).] (Kiefer 1952.)

5.27 Verify directly that the following families of densities satisfy (5.38).

(a) The exponential family of (1.5.1),

$$p_\eta(x) = h(x)e^{\eta T(x) - A(\eta)}.$$

(b) The location t family of Example 5.16.

(c) The logistic density of Table 1.4.1.

5.28 Extend condition (5.38) to vector-valued parameters, and show that it is satisfied by the exponential family (1.5.1) for $s > 1$.

5.29 Show that the assumption (5.36(b)) implies (5.38), so Theorem 5.15 is, in fact, a corollary of Theorem 5.10.

5.30 Show that (5.38) is satisfied if either of the following is true:

(a) $|\partial \log p_\theta / \partial\theta|$ is bounded.

(b) $[p_{\theta+\Delta}(x) - p_\theta(x)]/\Delta \to \partial \log p_\theta/\partial\theta$ uniformly.

5.31 (a) Show that if (5.38) holds, then the family of densities is *strongly differentiable* (see Note 8.6).

(b) Show that *weak differentiability* is implied by strong differentiability.

5.32 Brown and Gajek (1990) give two different sufficient conditions for (8.2) to hold, which are given below. Show that each implies (8.2). (Note that, in the progression from (a) to (b) the conditions become weaker, thus more widely applicable and harder to check.)

(a) For some $B < \infty$,

$$E_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2} p_\theta(X)/p_{\theta_0}(X)\right]^2 < B$$

for all θ in a neighborhood of θ.

(b) If $p_t^*(x) = \partial/\partial\theta p_\theta(x)|_{\theta=t}$, then

$$\lim_{\Delta\to 0} E_{\theta_0}\left[\frac{p^*_{\theta_0+\Delta}(X) - p^*_{\theta_0}(X)}{p_{\theta_0}(X)}\right]^2 = 0.$$

5.33 Let $\mathcal{F}$ be the class of all unimodal symmetric densities or, more generally, densities symmetric around zero and satisfying $f(x) \le f(0)$ for all x. Show that

$$\min_{f\in\mathcal{F}} \int x^2 f(x)dx = \frac{1}{12},$$

and that the minimum is attained by the uniform$(-\frac{1}{2}, \frac{1}{2})$ distribution. Thus, the uniform distribution has minimum variance among symmetric unimodal distributions. (See Example 4.8.6 for large-sample properties of the scale uniform.) [Hint: The side condition $\int f(x)dx = 1$, together with the method of undetermined multipliers, yields an equivalent problem, minimization of $\int(x^2 - a^2)f(x)dx$, where a is chosen to satisfy the constraint. A Neyman-Pearson type argument will now work.]

Section 6

6.1 For any random variables $(\psi_1, \ldots, \psi_s)$, show that the matrices $||E\psi_i\psi_j||$ and $C = ||\mathrm{cov}(\psi_i, \psi_j)||$ are positive semidefinite.

6.2 In this problem, we establish some facts about eigenvalues and eigenvectors of square matrices. (For a more general treatment, see, for example, Marshall and Olkin 1979, Chapter 20.)

We use the facts that a scalar $\lambda > 0$ is an *eigenvalue* of the $n \times n$ symmetric matrix A if there exists an $n \times 1$ vector p, the corresponding *eigenvector*, satisfying $Ap = \lambda p$. If A is nonsingular, there are n eigenvalues with corresponding linearly independent eigenvectors.

(a) Show that $A = P'D_\lambda P$, where D_λ is a diagonal matrix of eigenvalues of A and P is and $n \times n$ matrix whose rows are the corresponding eigenvalues that satisfies $P'P = PP' = I$, the identity matrix.

(b) Show that $\max_x \frac{x'Ax}{x'x}$ = largest eigenvalue of A.

(c) If B is a nonsingular symmetric matrix with eigenvector-eigenvalue representation $B = Q'D_\beta Q$, then $\max_x \frac{x'Ax}{x'Bx}$ = largest eigenvalue of A^*, where $A^* = D_\beta^{-1/2} Q A Q' D_\beta^{-1/2}$ and $D_\beta^{-1/2}$ is a diagonal matrix whose elements are the reciprocals of the square roots of the eigenvalues of B.

(d) For any square matrices C and D, show that the eigenvalues of the matrix CD are the same as the eigenvalues of the matrix DC, and hence that $\max_x \frac{x'Ax}{x'Bx}$ = largest eigenvalue of AB^{-1}.

(e) If $A = aa'$, where a is a $n \times 1$ vector (A is thus a rank-one matrix), then $\max_x \frac{x'aa'x}{x'Bx} = a'B^{-1}a$.

[Hint: For part (b) show that $\frac{x'Ax}{x'x} = \frac{y'D_\lambda y}{y'y} = \frac{\sum_i \lambda_i y_i^2}{\sum_i y_i^2}$, where $y = Px$, and hence the maximum is achieved at the vector y that is 1 at the coordinate of the largest eigenvalue and zero everywhere else.]

6.3 An alternate proof of Theorem 6.1 uses the method of Lagrange (or undetermined) multipliers. Show that, for fixed γ, the maximum value of $a'\gamma$, subject to the constraint that $a'Ca = 1$, is obtained by the solutions to

$$\frac{\partial}{\partial a_i}\left\{a'\gamma - \frac{1}{2}\lambda[a'Ca - 1]\right\} = 0,$$

where λ is the undetermined multiplier. (The solution is $a = \pm C^{-1}\gamma / \sqrt{\gamma' C^{-1}\gamma}$.)

6.4 Prove (6.11) under the assumptions of the text.

6.5 Verify (a) the information matrices of Table 6.1 and (b) Equations (6.15) and (6.16).

6.6 If $p(x) = (1-\varepsilon)\phi(x-\xi) + (\varepsilon/\tau)\phi[(x-\xi)/\tau]$ where ϕ is the standard normal density, find $I(\varepsilon, \xi, \tau)$.

6.7 Verify the expressions (6.20) and (6.21).

6.8 Let $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ be a partitioned matrix with A_{22} square and nonsingular, and let

$$B = \begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}.$$

Show that $|A| = |A_{11} - A_{12}A_{22}^{-1}A_{21}| \cdot |A_{22}|$.

6.9 (a) Let

$$A = \begin{pmatrix} a & b' \\ b & C \end{pmatrix}$$

where a is a scalar and b a column matrix, and suppose that A is positive definite. Show that $|A| \leq a|C|$ with equality holding if and only if $b = 0$.

(b) More generally, if the matrix A of Problem 6.8 is positive definite, show that $|A| \leq |A_{11}| \cdot |A_{22}|$ with equality holding if and only if $A_{12} = 0$.

[*Hint*: Transform A_{11} and the positive semidefinite $A_{12}A_{22}^{-1}A_{21}$ simultaneously to diagonal form.]

6.10 (a) Show that if the matrix A is nonsingular, then for any vector x, $(x'Ax)(x'A^{-1}x) > (x'x)^2$.

(b) Show that, in the notation of Theorem 6.6 and the following discussion,

$$\frac{\left[\frac{\partial}{\partial\theta_i} E_\theta \delta\right]^2}{I_{ii}(\theta)} = \frac{(\varepsilon_i'\alpha)^2}{\varepsilon_i' I(\theta)\varepsilon_i},$$

and if $\alpha = (0, \ldots, 0, \alpha_i, 0, \ldots 0)$, $\alpha' I(\theta)^{-1}\alpha = (\varepsilon_i'\alpha)^2 \varepsilon_i' I(\theta)^{-1}\varepsilon_i$, and hence establish (6.25).

6.11 Prove that (6.26) is necessary for equality in (6.25). [*Hint*: Problem 6.9(a).]

6.12 Prove the Bhattacharyya inequality (6.29) and show that the condition of equality is as stated.

8 Notes

8.1 Unbiasedness and Information

The concept of unbiasedness as "lack of systematic error" in the estimator was introduced by Gauss (1821) in his work on the theory of least squares. It has continued as a basic assumption in the developments of this theory since then.

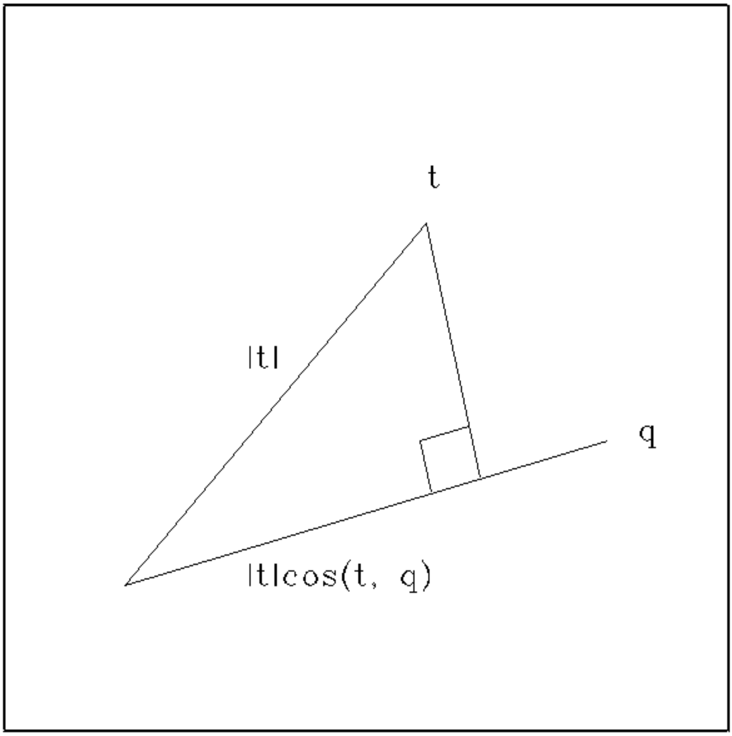
The amount of information that a data set contains about a parameter was introduced by Edgeworth (1908, 1909) and was developed more systematically by Fisher (1922 and later papers). The first version of the information inequality, and hence connections with unbiased estimation, appears to have been given by Fréchet (1943). Early extensions and rediscoveries are due to Darmois (1945), Rao (1945), and Cramér (1946b). The designation "information inequality," which replaced the earlier "Cramér-Rao inequality," was proposed by Savage (1954).

8.2 UMVU Estimators

The first UMVU estimators were obtained by Aitken and Silverstone (1942) in the situation in which the information inequality yields the same result (Problem 5.17). UMVU estimators as unique unbiased functions of a suitable sufficient statistic were derived in special cases by Halmos (1946) and Kolmogorov (1950) and were pointed out as a general fact by Rao (1947). An early use of Method 1 for determining such unbiased estimators is due to Tweedie (1947). The concept of completeness was defined, its implications for unbiased estimation developed, and Theorem 1.7 obtained, in Lehmann and Scheffé (1950, 1955, 1956).

Theorem 1.11 has been used to determine UMVU estimators in many special cases. Some applications include those of Abbey and David (1970, exponential distribution), Ahuja (1972, truncated Poisson), Bhattacharyya et al. (1977, censored), Bickel and Lehmann (1969, convex), Varde and Sathe (1969, truncated exponential), Brown and Cohen (1974, common mean), Downton (1973, $P(X \leq Y)$), Woodward and Kelley (1977, $P(X \leq Y)$), Iwase (1983, inverse Gaussian), and Kremers (1986, sum-quota sampling).

Figure 8.1. *Illustration of the information inequality*



8.3 Existence of Unbiased Esimators

Doss and Sethuraman (1989) show that the process of bias reduction may not always be the wisest course. If an estimand $g(\theta)$ does not have an unbiased estimator, and one tries to reduce the bias in a biased estimator δ, they show that as the bias goes to zero, $\text{var}(\delta) \to \infty$ (see Problem 1.4).

This result has implications for bias-reduction procedures such as the jackknife and the bootstrap. (For an introduction to the jackknife and the bootstrap, see Efron and Tibshirani 1993 or Shao and Tu 1995.) In particular, Efron and Tibshirani (1993, Section 10.6) discuss some practical implications of bias reduction, where they urge caution in its use, as large increases in standard errors can result.

Liu and Brown (1993) call a problem *singular* if there exists no unbiased estimator with finite variance. More precisely, if $\mathcal{F}$ is a family of densities, then if a problem is singular, there will be at least one member of $\mathcal{F}$, called a *singular point*, where any unbiased estimator of a parameter (or functional) will have infinite variance. There are many examples of singular problems, both in parametric and nonparametric estimation, with nonparametric density estimation being, perhaps, the best known. Two particularly simple examples of singular problems are provided by Example 1.2 (estimation of $1/p$ in a binomial problem) and Problem 5.21 (a mixture estimation problem).

8.4 Geometry of the Information Inequality

The information inequality can be interpreted as, and a proof can be based on, the fact that the length of the hypotenuse of a right triangle exceeds the length of each side.

For two vectors a and b, define $< t, q >= t'q$, with $< t, t >^2 = |t|^2$. For the triangle in Figure 8.1, using the fact that the cosine of the angle between t and q is $\cos(t, q) = t'q/|t||q|$ and the fact that the hypotenuse is the longest side, we have

$$|t| > |t|\cos(t, q) = |t| \left[\frac{< t, q >}{|t||q|} \right] = \frac{< t, q >}{|q|} .$$

If we define $< X, Y > = E[(X - EX)(Y - EY)]$ for random variables X and Y, applying the above inequality with this definition results in the covariance inequality (5.1), which, in turn, leads to the information inequality. See Fabian and Hannan (1977) for a rigorous development.

8.5 Fisher Information and the Hazard Function

Efron and Johnstone (1990) investigate an identity between the Fisher information number and the *hazard function*, h, defined by

$$h_\theta(x) = \lim_{\Delta \to 0} \Delta^{-1} P(x \leq X < x + \Delta | X \geq x) = \frac{f_\theta(x)}{1 - F_\theta(x)}$$

where f_θ and F_θ are the density and distribution function of the random variable X, respectively. The hazard function, $h(x)$, represents the conditional survival rate given survival up to time x and plays and important role in survival analysis. (See, for example, Kalbfleish and Prentice 1980, Cox and Oakes 1984, Fleming and Harrington 1991.)

Efron and Johnstone show that

$$I(\theta) = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log[f_\theta(x)]^2 f_\theta(x) dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log[h_\theta(x)]^2 f_\theta(x) dx.$$

They then interpret this identity and discuss its implications to, and connections with, survival analysis and statistical curvature of hazard models, among other things. They also note that this identity can be derived as a consequence of the more general result of James (1986), who showed that if $b(\cdot)$ is a continuous function of the random variable X, then

$$\text{var}[b(X)] = E[b(X) - \bar{b}(X)]^2, \quad \text{where} \quad \bar{b}(x) = E[b(X)|b(X) > x],$$

as long as the expectations exist.

8.6 Weak and Strong Differentiability

Research into determining necessary and sufficient conditions for the applicability of the Information Inequality bound has a long history (see, for example, Blyth and Roberts 1972, Fabian and Hannan 1977, Ibragimov and Has'minskii 1981, Section 1.7, Müller-Funk et al. 1989, Brown and Gajek 1990). What has resulted is a condition on the density sufficient to ensure (5.29).

The precise condition needed was presented by Fabian and Hannan (1977), who call it *weak differentiability*. The function $p_{\theta+\Delta}(x)/p_\theta(x)$ is *weakly differentiable* at θ if there is a measurable function q such that

$$\lim_{\Delta \to 0} \int h(x) \left\{ \left[\Delta^{-1} \left(\frac{p_{\theta+\Delta}(x)}{p_\theta(x)} - 1 \right) \right] - q(x) \right\} p_\theta(x)\, d\mu(x) = 0 \tag{8.1}$$

for all $h(\cdot)$ such that $\int h^2(x) p_\theta(x)\, d\mu(x) < \infty$. Weak differentiability is actually equivalent (necessary and sufficient) to the existence of a function $q_\theta(x)$ such that $(\partial/\partial\theta) E_\theta \delta = E\delta q$. Hence, it can replace condition (5.38) in Theorem 5.15.

Since weak differentiability is often difficult to verify, Brown and Gajek (1990) introduce the more easily verifiable condition of *strong differentiability*, which implies weak differentiability, and thus can also replace condition (5.38) in Theorem 5.15 (Problem 5.31). The function $p_{\theta+\Delta}(x)/p_\theta(x)$ is *strongly differentiable at* $\theta = \theta_0$ *with derivative* $q_{\theta_0}(x)$ if

$$\lim_{\Delta\to 0}\int\left\{\left[\Delta^{-1}\left(\frac{p_{\theta+\Delta}(x)}{p_\theta(x)}-1\right)\right]-q_{\theta_0}(x)\right\}^2 p_{\theta_0}(x)\,d\mu(x)=0. \tag{8.2}$$

These variations of the usual definition of differentiability are well suited for the information inequality problem. In fact, consider the expression in the square brackets in (8.1). If the limit of this expression exists, it is $q_\theta(x) = \partial \log p_\theta(x)/\partial\theta$. Of course, existence of this limit does not, by itself, imply condition (8.2); such an implication requires an integrability condition.

Brown and Gajek (1990) detail a number of easier-to-check conditions that imply (8.2). (See Problem 5.32.) Fabian and Hannan (1977) remark that if (8.1) holds and $\partial \log p_\theta(x)/\partial\theta$ exists, then it must be the case that $q_\theta(x) = \partial \log p_\theta(x)/\partial\theta$. However, the existence of one does not imply the existence of the other.

see how to present the transformations in the usual order, let us consider the sample space as the totality of possible samples s together with the labels and values of their elements. Suppose, for example, that the transformations are permutations of the labels. Since the same elements appear in many different samples, one must ensure that the transformations g of the samples are consistent, that is, that the transform of an element is independent of the particular sample in which it appears. If a transformation has this property, it will define a permutation of all the labels in the population and hence a transformation $\bar{g}$ of θ. Starting with g or $\bar{g}$ thus leads to the same result; the latter is more convenient because it provides the required consistency property automatically.

8 Problems

Section 1

1.1 Prove the parts of Theorem 1.4 relating to (a) risk and (b) variance.

1.2 In model (1.9), suppose that $n = 2$ and that f satisfies $f(-x_1, -x_2) = f(x_2, x_1)$. Show that the distribution of $(X_1 + X_2)/2$ given $X_2 - X_1 = y$ is symmetric about 0. Note that if X_1 and X_2 are iid according to a distribution which is symmetric about 0, the above equation holds.

1.3 If X_1 and X_2 are distributed according to (1.9) with $n = 2$ and f satisfying the assumptions of Problem 1.2, and if ρ is convex and even, then the MRE estimator of ξ is $(X_1 + X_2)/2$.

1.4 Under the assumptions of Example 1.18, show that (a) $E[X_{(1)}] = b/n$ and (b) $\text{med}[X_{(1)}] = b \log 2/n$.

1.5 For each of the three loss functions of Example 1.18, compare the risk of the MRE estimator to that of the UMVU estimator.

1.6 If T is a sufficient statistic for the family (1.9), show that the estimator (1.28) is a function of T only. [*Hint*: Use the factorization theorem.]

1.7 Let $X_i (i = 1, 2, 3)$ be independently distributed with density $f(x_i - \xi)$ and let $\delta = X_1$ if $X_3 > 0$ and $= X_2$ if $X_3 \leq 0$. Show that the estimator δ of ξ has constant risk for any invariant loss function, but δ is not location equivariant.

1.8 Prove Corollary 1.14. [*Hint*: Show that (a) $\phi(v) = E_0\rho(X - v) \to M$ as $v \to \pm\infty$ and (b) that ϕ is continuous; (b) follows from the fact (see TSH2, Appendix Section 2) that if $f_n, n = 1, 2, \ldots$ and f are probability densities such that $f_n(x) \to f(x)$ a.e., then $\int \psi f_n \to \int \psi f$ for any bounded ψ.]

1.9 Let $X_1, \ldots, X_n$ be distributed as in Example 1.19 and let the loss function be that of Example 1.15. Determine the totality of MRE estimators and show that the midrange is one of them.

1.10 Consider the loss function

$$\rho(t) = \begin{cases} -At & \text{if } t < 0 \\ Bt & \text{if } t \geq 0 \end{cases} \qquad (A, B \geq 0).$$

If X is a random variable with density f and distribution function F, show that $E\rho(X - v)$ is minimized for any v satisfying $F(v) = B/(A + B)$.

1.11 In Example 1.16, find the MRE estimator of ξ when the loss function is given by Problem 1.10.

1.12 Show that an estimator $\delta(X)$ of $g(\theta)$ is risk-unbiased with respect to the loss function of Problem 1.10 if $F_\theta[g(\theta)] = B/(A+B)$, where F_θ is the cdf of $\delta(X)$ under θ.

1.13 Suppose $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ have joint density $f(x_1 - \xi, \ldots, x_m - \xi; y_1 - \eta, \ldots, y_n - \eta)$ and consider the problem of estimating $\Delta = \eta - \xi$. Explain why it is desirable for the loss function $L(\xi, \eta; d)$ to be of the form $\rho(d - \Delta)$ and for an estimator δ of Δ to satisfy $\delta(\mathbf{x} + a, \mathbf{y} + b) = \delta(\mathbf{x}, \mathbf{y}) + (b - a)$.

1.14 Under the assumptions of the preceding problem, prove the equivalents of Theorems 1.4–1.17 and Corollaries 1.11–1.14 for estimators satisfying the restriction.

1.15 In Problem 1.13, determine the totality of estimators satisfying the restriction when $m = n = 1$.

1.16 In Problem 1.13, suppose the X's and Y's are independently normally distributed with known variances σ^2 and τ^2. Find conditions on ρ under which the MRE estimator is $\bar{Y} - \bar{X}$.

1.17 In Problem 1.13, suppose the X's and Y's are independently distributed as $E(\xi, 1)$ and $E(\eta, t)$, respectively, and that $m = n$. Find conditions on ρ under which the MRE estimator of Δ is $Y_{(1)} - X_{(1)}$.

1.18 In Problem 1.13, suppose that $\mathbf{X}$ and $\mathbf{Y}$ are independent and that the loss function is squared error. If $\hat{\xi}$ and $\hat{\eta}$ are the MRE estimators of ξ and η, respectively, the MRE estimator of Δ is $\hat{\eta} - \hat{\xi}$.

1.19 Suppose the X's and Y's are distributed as in Problem 1.17 but with $m \neq n$. Determine the MRE estimator of Δ when the loss is squared error.

1.20 For any density f of $\mathbf{X} = (X_1, \ldots, X_n)$, the probability of the set $A = \{\mathbf{x} : 0 < \int_{-\infty}^{\infty} f(\mathbf{x} - u)\, du < \infty\}$ is 1. [*Hint*: With probability 1, the integral in question is equal to the marginal density of $\mathbf{Y} = (Y_1, \ldots, Y_{n-1})$ where $Y_i = X_i - X_n$, and $P[0 < g(\mathbf{Y}) < \infty] = 1$ holds for any probability density g.]

1.21 Under the assumptions of Theorem 1.10, if there exists an equivariant estimator δ_0 of ξ with finite expected squared error, show that

(a) $E_0(|X_n| \mid \mathbf{Y}) < \infty$ with probability 1;

(b) the set $B = \{\mathbf{x} : \int |u| f(\mathbf{x} - u)\, du < \infty\}$ has probability 1.

[*Hint*: (a) $E|\delta_0| < \infty$ implies $E(|\delta_0| \mid \mathbf{Y}) < \infty$ with probability 1 and hence $E[\delta_0 - v(\mathbf{Y})| \mid \mathbf{Y}] < \infty$ with probability 1 for any $v(\mathbf{Y})$. (b) $P(B) = 1$ if and only if $E(|X_n| \mid \mathbf{Y}) < \infty$ with probability 1.]

1.22 Let δ_0 be location equivariant and let $\mathcal{U}$ be the class of all functions u satisfying (1.20) and such that $u(X)$ is an unbiased estimator of zero. Then, δ_0 is MRE if and only if $\text{cov}[\delta_0, u(X)] = 0$ for all $u \in \mathcal{U}$.[2] (Note the analogy with Theorem 2.1.7.)

Section 2

2.1 Show that the class $G(\mathcal{C})$ is a group.

2.2 In Example 2.2(ii), show that the transformations $\mathbf{x}' = -\mathbf{x}$ together with the identity transformation form a group.

2.3 Let $\{gX, g \in G\}$ be a group of transformations that leave the model (2.1) invariant. If the distributions $P_\theta, \theta \in \Omega$ are distinct, show that the induced transformations $\bar{g}$ are $1:1$ transformations of Ω. [*Hint*: To show that $\bar{g}\theta_1 = \bar{g}\theta_2$ implies $\theta_1 = \theta_2$, use the fact that $P_{\theta_1}(A) = P_{\theta_2}(A)$ for all A implies $\theta_1 = \theta_2$.]

[2] Communicated by P. Bickel.

2.4 Under the assumptions of Problem 2.3, show that

(a) the transformations $\bar{g}$ satisfy $\overline{g_2 g_1} = \bar{g}_2 \cdot \bar{g}_1$ and $(\bar{g})^{-1} = (\overline{g^{-1}})$;
(b) the transformations $\bar{g}$ corresponding to $g \in G$ form a group.
(c) establish (2.3) and (2.4).

2.5 Show that a loss function satisfies (2.9) if and only if it is of the form (2.10).

2.6 (a) The transformations g^* defined by (2.12) satisfy $(g_2 g_1)^* = g_2^* \cdot g_1^*$ and $(g^*)^{-1} = (g^{-1})^*$.
(b) If G is a group leaving (2.1) invariant and $G^* = \{g^*, g \in G\}$, then G^* is a group.

2.7 Let X be distributed as $N(\xi, \sigma^2)$, $-\infty < \xi < \infty, 0 < \sigma$, and let $h(\xi, \sigma) = \sigma^2$. The problem is invariant under the transformations $x' = ax + c; 0 < a, -\infty < c < \infty$. Show that the only equivariant estimator is $\delta(X) \equiv 0$.

2.8 Show that:

(a) If (2.11) holds, the transformations g^* defined by (2.12) are 1 : 1 from $\mathcal{H}$ onto itself.
(b) If $L(\theta, d) = L(\theta, d')$ for all θ implies $d = d'$, then g^* defined by (2.14) is unique, and is a 1 : 1 transformation from $\mathcal{D}$ onto itself.

2.9 If θ is the true temperature in degrees Celsius, then $\theta' = \bar{g}\theta = \theta + 273$ is the true temperature in degrees Kelvin. Given an observation X, in degrees Celsius:

(a) Show that an estimator $\delta(X)$ is functionally equivariant if it satisfies $\delta(x) + a = \delta(x + a)$ for all a.
(b) Suppose our estimator is $\delta(x) = (ax + b\theta_0)/(a + b)$, where x is the observed temperature in degrees Celsius, θ_0 is a prior guess at the temperature, and a and b are constants. Show that for a constant K, $\delta(x + K) \neq \delta(x) + K$, so δ does not satisfy the principle of functional equivariance.
(c) Show that the estimators of part (b) will not satisfy the principle of formal invariance.

2.10 To illustrate the difference between functional equivariance and formal invariance, consider the following.

To estimate the amount of electric power obtainable from a stream, one could use the estimate

$$\delta(x) = c \min\{100, x - 20\}$$

where x = stream flow in m^3/sec, 100 m^3/sec is the capacity of the pipe leading to the turbine, and 20 m^3/sec is the flow reduction necessary to avoid harming the trout. The constant c, in kilowatts $/\text{m}^3/\text{sec}$ converts the flow to a kilowatt estimate.

(a) If measurements were, instead, made in liters and watts, so $g(x) = 1000x$ and $\bar{g}(\theta) = 1000\,\theta$, show that functional equivariance leads to the estimate

$$\bar{g}(\delta(x)) = c \min\{10^5, g(x) - 20,000\}.$$

(b) The principle of formal invariance leads to the estimate $\delta(g(x))$. Show that this estimator is not a reasonable estimate of wattage.

(Communicated by L. LeCam.)

2.11 In an invariant probability model, write $X = (T, W)$, where T is sufficient for θ, and W is ancillary .

(a) If the group operation is transitive, show that any invariant statistic must be ancillary.
(b) What can you say about the invariance of an ancillary statistic?

2.12 In an invariant estimation problem, write $X = (T, W)$ where T is sufficient for θ, and W is ancillary. If the group of transformations is transitive, show:

(a) The best equivariant estimator δ^* is the solution to $\min_d E_\theta[L(\theta, d(x))|W = w]$.

(b) If e is the identity element of the group ($g^{-1}g = e$), then $\delta^* = \delta^*(t, w)$ can be found by solving, for each w, $\min_d E_e\{L[e, d(T, w)]|W = w\}$.

2.13 For the situation of Example 2.11:

(a) Show that the class of transformations is a group.

(b) Show that equivariant estimators must satisfy $\delta(n - x) = 1 - \delta(x)$.

(c) Show that, using an invariant loss, the risk of an equivariant estimator is symmetric about $p = 1/2$.

2.14 For the situation of Example 2.12:

(a) Show that the class of transformations is a group.

(b) Show that estimators of the form $\varphi(\bar{x}/s^2)s^2$, where $\bar{x} = 1/n\Sigma x_i$ and $s^2 = \Sigma(x_i - \bar{x})^2$ are equivariant, where φ is an arbitrary function.

(c) Show that, using an invariant loss function, the risk of an equivariant estimator is a function only of $\tau = \mu/\sigma$.

2.15 Prove Corollary 2.13.

2.16 (a) If g is the transformation (2.20), determine $\bar{g}$.

(b) In Example 2.12, show that (2.22) is not only sufficient for (2.14) but also necessary.

2.17 (a) In Example 2.12, determine the smallest group G containing both G_1 and G_2.

(b) Show that the only estimator that is invariant under G is $\delta(\mathbf{X}, \mathbf{Y}) \equiv 0$.

2.18 If $\delta(X)$ is an equivariant estimator of $h(\theta)$ under a group G, then so is $g^*\delta(X)$ with g^* defined by (2.12) and (2.13), provided G^* is commutative.

2.19 Show that:

(a) In Example 2.14(i), X is not risk-unbiased.

(b) The group of transformations $ax + c$ of the real line ($0 < a, -\infty < c < \infty$) is not commutative.

2.20 In Example 2.14, determine the totality of equivariant estimators of Δ under the smallest group G containing G_1 and G_2.

2.21 Let θ be real-valued and h strictly increasing, so that (2.11) is vacuously satisfied. If $L(\theta, d)$ is the loss resulting from estimating θ by d, suppose that the loss resulting from estimating $\theta' = h(\theta)$ by $d' = h(d)$ is $M(\theta', d') = L[\theta, h^{-1}(d')]$. Show that:

(a) If the problem of estimating θ with loss function L is invariant under G, then so is the problem of estimating $h(\theta)$ with loss function M.

(b) If δ is equivariant under G for estimating θ with loss function L, show that $h[\delta(X)]$ is equivariant for estimating $h(\theta)$ with loss function M.

(c) If δ is MRE for θ with L, then $h[\delta(X)]$ is MRE for $h(\theta)$ with M.

2.22 If $\delta(\mathbf{X})$ is MRE for estimating ξ in Example 2.2(i) with loss function $\rho(d - \xi)$, state an optimum property of $e^{\delta(\mathbf{X})}$ as an estimator of e^ξ.

2.23 Let X_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, s$, and W be distributed according to a density of the form

$$\left[\prod_{i=1}^{s} f_i(\mathbf{x}_i - \xi_i)\right] h(w)$$

where $\mathbf{x}_i - \xi_i = (x_{i1} - \xi_i, \ldots, x_{in_i} - \xi_i)$, and consider the problem of estimating $\theta = \Sigma c_i \xi_i$ with loss function $L(\xi_i, \ldots, \xi_s; d) = \rho(d - \theta)$. Show that:

(a) This problem remains invariant under the transformations

$$X'_{ij} = X_{ij} + a_i, \quad \xi'_i = \xi_i + a_i, \quad \theta' = \theta + \Sigma a_i c_i,$$
$$d' = d + \Sigma a_i c_i.$$

(b) An estimator δ of θ is equivariant under these transformations if

$$\delta(\mathbf{x}_1 + a_1, \ldots, \mathbf{x}_s + a_s, w) = \delta(\mathbf{x}_1, \ldots, \mathbf{x}_s, w) + \Sigma a_i c_i.$$

2.24 Generalize Theorem 1.4 to the situation of Problem 2.23.

2.25 If δ_0 is any equivariant estimator of θ in Problem 2.23, and if $\mathbf{y}_i = (x_{i1} - x_{in_i}, x_{i2} - x_{in_i}, \ldots, x_{in_i-1} - x_{in_i})$, show that the most general equivariant estimator of θ is of the form

$$\delta(\mathbf{x}_1, \ldots, \mathbf{x}_s, w) = \delta_0(\mathbf{x}_1, \ldots, \mathbf{x}_s, w) - v(\mathbf{y}_1, \ldots, \mathbf{y}_s, w).$$

2.26 (a) Generalize Theorem 1.10 and Corollary 1.12 to the situation of Problems 2.23 and 2.25. (b) Show that the MRE estimators of (a) can be chosen to be independent of W.

2.27 Suppose that the variables X_{ij} in Problem 2.23 are independently distributed as $N(\xi_i, \sigma^2)$, σ is known. Show that:

(a) The MRE estimator of θ is then $\Sigma c_i \bar{X}_i - v^*$, where $\bar{X}_i = (X_{i1} + \cdots + X_{in_i})/n_i$, and where v^* minimizes (1.24) with $X = \Sigma c_i \bar{X}_i$.

(b) If ρ is convex and even, the MRE estimator of θ is $\Sigma c_i \bar{X}_i$.

(c) The results of (a) and (b) remain valid when σ is unknown and the distribution of W depends on σ (but not the ξ's).

2.28 Show that the transformation of Example 2.11 and the identity transformation are the only transformations leaving the family of binomial distributions invariant.

Section 3

3.1 (a) A loss function L satisfies (3.4) if and only if it satisfies (3.5) for some γ.

(b) The sample standard deviation, the mean deviation, the range, and the MLE of τ all satisfy (3.7) with $r = 1$.

3.2 Show that if $\delta(\mathbf{X})$ is scale invariant, so is $\delta^*(\mathbf{X})$ defined to be $\delta(\mathbf{X})$ if $\delta(\mathbf{X}) \geq 0$ and $= 0$ otherwise, and the risk of δ^* is no larger than that of δ for any loss function (3.5) for which $\gamma(v)$ is nonincreasing for $v \leq 0$.

3.3 Show that the bias of any equivariant estimator of τ^r in (3.1) is proportional to τ^r.

3.4 A necessary and sufficient condition for δ to satisfy (3.7) is that it is of the form $\delta = \delta_0/u$ with δ_0 and u satisfying (3.7) and (3.9), respectively.

3.5 The function ρ of Corollary 3.4 with γ defined in Example 3.5 is strictly convex for $p \geq 1$.

3.6 Let X be a positive random variable. Show that:

(a) If $EX^2 < \infty$, then the value of c that minimizes $E(X/c - 1)^2$ is $c = EX^2/EX$.

(b) If Y has the gamma distribution with $\Gamma(\alpha, 1)$, then the value of w minimizing $E[(Y/w) - 1]^2$ is $w = \alpha + 1$.

3.7 Let X be a positive random variable.

(a) If $EX < \infty$, then the value of c that minimizes $E|X/c - 1|$ is a solution to $EXI(X \le c) = EXI(X \ge c)$, which is known as a *scale median*.

(b) Let Y have a χ^2-distribution with f degrees for freedom. Then, the minimizing value is $w = f + 2$. [*Hint*: (b) Example 1.5.9.]

3.8 Under the assumptions of Problem 3.7(a), the set of scale medians of X is an interval. If $f(x) > 0$ for all $x > 0$, the scale median of X is unique.

3.9 Determine the scale median of X when the distribution of X is (a) $U(0, \theta)$ and (b) $E(0, b)$.

3.10 Under the assumptions of Theorem 3.3:

(a) Show that the MRE estimator under the loss (3.13) is given by (3.14).

(b) Show that the MRE estimator under the loss (3.15) is given by (3.11), where $w^*(\mathbf{z})$ is any scale median of $\delta_0(\mathbf{x})$ under the distribution of $\mathbf{X}|\mathbf{Z}$.

[*Hint*: Problem 3.7.]

3.11 Let $X_1, \ldots, X_n$ be iid according to the uniform distribution $u(0, \theta)$.

(a) Show that the complete sufficient statistic $X_{(n)}$ is independent of Z [given by Equation (3.8)].

(b) For the loss function (3.13) with $r = 1$, the MRE estimator of θ is $X_{(n)}/w$, with $w = (n + 1)/(n + 2)$.

(c) For the loss function (3.15) with $r = 1$, the MRE estimator of θ is $[2^{1/(n+1)}]\, X_{(n)}$.

3.12 Show that the MRE estimators of Problem 3.11, parts (b) and (c), are risk-unbiased, but not mean-unbiased.

3.13 In Example 3.7, find the MRE estimator of $\text{var}(X_1)$ when the loss function is (a) (3.13) and (b) (3.15) with $r = 2$.

3.14 Let $X_1, \ldots, X_n$ be iid according to the exponential distribution $E(0, \tau)$. Determine the MRE estimator of τ for the loss functions (a) (3.13) and (b) (3.15) with $r = 1$.

3.15 In the preceding problem, find the MRE estimator of $\text{var}(X_1)$ when the loss function is (3.13) with $r = 2$.

3.16 Prove formula (3.19).

3.17 Let $X_1, \ldots, X_n$ be iid each with density $(2/\tau)[1 - (x/\tau)], 0 < x < \tau$. Determine the MRE estimator (3.19) of τ^r when (a) $n = 2$, (b) $n = 3$, and (c) $n = 4$.

3.18 In the preceding problem, find $\text{var}(X_1)$ and its MRE estimator for $n = 2, 3, 4$ when the loss function is (3.13) with $r = 2$.

3.19 (a) Show that the loss function L_s of (3.20) is convex and invariant under scale transformations.

(b) Prove Corollary 3.8.

(c) Show that for the situation of Example 3.7, if the loss function is L_s, then the UMVU estimator is also the MRE.

3.20 Let $X_1, \ldots, X_n$ be iid from the distribution $N(\theta, \theta^2)$.

(a) Show that this probability model is closed under scale transformations.

(b) Show that the MLE is equivariant.

[The MRE estimator is obtainable from Theorem 3.3, but does not have a simple form. See Eaton 1989, Robert 1991, 1994a for more details. Gleser and Healy (1976) consider a similar problem using squared error loss.]

3.21 (a) If δ_0 satisfies (3.7) and $c\delta_0$ satisfies (3.22), show that $c\delta_0$ cannot be unbiased in the sense of satisfying $E(c\delta_0) \equiv \tau^r$.

(b) Prove the statement made in Example 3.10.

3.22 Verify the estimator δ^* of Example 3.12.

3.23 If G is a group, a subset G_0 of G is a *subgroup* of G if G_0 is a group under the group operation of G.

(a) Show that the scale group (3.32) is a subgroup of the location-scale group (3.24)

(b) Show that any equivariant estimator of τ^r that is equivariant under (3.24) is also equivariant under (3.32); hence, in a problem that is equivariant under (3.32), the best scale equivariant estimator is at least as good as the best location-scale equivariant estimator.

(c) Explain why, in general, if $\mathcal{G}_0$ is a subgroup of $\mathcal{G}$, one can expect equivariance under $\mathcal{G}_0$ to produce better estimators than equivariance under $\mathcal{G}$.

3.24 For the situation of Example 3.13:

(a) Show that an estimator is equivariant if and only if it can be written in the form $\varphi(\bar{x}/s)s^2$.

(b) Show that the risk of an equivariant estimator is a function only of ξ/τ.

3.25 If $X_1, \ldots, X_n$ are iid according to $E(\xi, \tau)$, determine the MRE estimator of τ for the loss functions (a) (3.13) and (b) (3.15) with $r = 1$ and the MRE estimator of ξ for the loss function (3.43).

3.26 Show that δ satisfies (3.35) if and only if it satisfies (3.40) and (3.41).

3.27 Determine the bias of the estimator $\delta^*(\mathbf{X})$ of Example 3.18.

3.28 Lele (1993) uses invariance in the study of *mophometrics*, the quantitative analysis of biological forms. In the analysis of a biological object, one measures data $\mathbf{X}$ on k specific points called *landmarks*, where each landmark is typically two- or three-dimensional . Here we will assume that the landmark is two-dimensional (as is a picture), so $\mathbf{X}$ is a $k \times 2$ matrix. A model for $\mathbf{X}$ is

$$\mathbf{X} = (M + \mathbf{Y})\Gamma + \mathbf{t}$$

where $M_{k\times 2}$ is the mean form of the object, $\mathbf{t}$ is a fixed translation vector, and Γ is a 2×2 matrix that rotates the vector $\mathbf{X}$. The random variable $\mathbf{Y}_{k\times 2}$ is a *matrix normal* random variable, that is, each column of $\mathbf{Y}$ is distributed as $N(0, \Sigma_k)$, a k-variate normal random variable, and each row is distributed as $N(0, \Sigma_d)$, a bivariate normal random variable.

(a) Show that $\mathbf{X}$ is a matrix normal random variable with columns distributed as $N_k(M\Gamma_j, \Sigma_k)$ and rows distributed as $N_2(M_i\Gamma, \Gamma'\Sigma_d\Gamma)$, where Γ_j is the jth column of Γ and M_i is the ith row of M.

(b) For estimation of the shape of a biological form, the parameters of interest are M, Σ_k and Σ_d, with $\mathbf{t}$ and Γ being nuisance parameters. Show that, even if there were no nuisance parameters, Σ_k or Σ_d is not identifiable.

(c) It is usually assumed that the (1, 1) element of either Σ_k or Σ_d is equal to 1. Show that this makes the model identifiable.

(d) The form of a biological object is considered an inherent property of the form (a baby has the same form as an adult) and should not be affected by rotations, reflections, or translations. This is summarized by the transformation

$$\mathbf{X}' = \mathbf{X}P + b$$

where P is a 2×2 orthogonal matrix ($P'P = I$) and b is a $k \times 1$ vector. (See Note 9.3 for a similar group.) Suppose we observe n landmarks $\mathbf{X}_1, \cdots, \mathbf{X}_n$. Define the Euclidean distance between two matrices A and B to be $D(A, B) = \sum_{ij}(a_{ij} - b_{ij})^2$, and let the $n \times n$ matrix F have (i, j)th element $f_{ij} = D(\mathbf{X}_i, \mathbf{X}_j)$. Show that F is invariant under this group, that is $F(\mathbf{X}') = F(\mathbf{X})$. (Lele (1993) notes that F is, in fact, maximal invariant.)

3.29 In (9.1), show that the group $\mathbf{X}' = A\mathbf{X}+b$ induces the group $\mu' = A\mu+b, \Sigma' = A\Sigma A'$.

3.30 For the situation of Note 9.3, consider the equivariant estimation of μ.

(a) Show that an invariant loss is of the form $L(\mu, \Sigma, \delta) = L((\mu - \delta)'\Sigma^{-1}(\mu - \delta))$.

(b) The equivariant estimators are of the form $\bar{X} + c$, with $c = 0$ yielding the MRE estimator.

3.31 For $\mathbf{X}_1, \ldots, \mathbf{X}_n$ iid as $N_p(\mu, \Sigma)$, the cross-products matrix S is defined by

$$S = \{S_{ij}\} = \sum_{k=1}^{n}(x_{i_k} - \bar{x}_i)(x_{j_k} - \bar{x}_j)$$

where $\bar{x}_i = (1/n)\sum_{k=1}^{n} x_{i_k}$. Show that, for $\Sigma = I$,

(a) $E_I[\mathrm{tr}S] = E_I \sum_{i=1}^{p}\sum_{k=1}^{n}(X_{i_k} - \bar{X}_i)(X_{i_k} - \bar{X}_i) = p(n-1)$,

(b) $E_I[\mathrm{tr}S^2] = E_I \sum_{i=1}^{p}\sum_{j=1}^{p}\{\sum_{k=1}^{n}(X_{i_k} - \bar{X}_i)(X_{j_k} - \bar{X}_j)\}^2 = (n-1)(np - p - 1)$.

[These are straightforward, although somewhat tedious, calculations involving the chi-squared distribution. Alternatively, one can use the fact that S has a Wishart distribution (see, for example, Anderson 1984), and use the properties of that distribution.]

3.32 For the situation of Note 9.3:

(a) Show that equivariant estimators of Σ are of the form cS, where S is the cross-products matrix and c is a constant.

(b) Show that $E_I\{\mathrm{tr}[(cS - I)'(cS - I)]\}$ is minimized by $c = E_I\mathrm{tr}S/E_I\mathrm{tr}S^2$.

[*Hint*: For part (a), use a generalization of Theorem 3.3; see the argument leading to (3.29), and Example 3.11.]

3.33 For the estimation of Σ in Note 9.3:

(a) Show that the loss function in (9.2) is invariant.

(b) Show that Stein's loss $L(\delta, \Sigma) = \mathrm{tr}(\delta\Sigma^{-1}) - \log|\delta\Sigma^{-1}| - p$, where $|A|$ is the determinant of A, is an invariant loss with MRE estimator S/n.

(c) Show that a loss $L(\delta, \Sigma)$ is an invariant loss if and only if it can be written as a function of the eigenvalues of $\delta\Sigma^{-1}$.

[The univariate version of Stein's loss was seen in (3.20) and Example 3.9. Stein (1956b) and James and Stein (1961) used the multivariate version of the loss. See also Dey and Srinivasan 1985, and Dey et al. 1987.]

3.34 Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ have joint density

$$\frac{1}{\sigma^m \tau^n} f\left(\frac{x_1}{\sigma}, \ldots, \frac{x_m}{\sigma}; \frac{y_1}{\tau}, \ldots, \frac{y_n}{\tau}\right),$$

and consider the problem of estimating $\theta = (\tau/\sigma)^r$ with loss function $L(\sigma, \tau; d) = \gamma(d/\theta)$. This problem remains invariant under the transformations $X_i' = aX_i$, $Y_j' = bY_j$, $\sigma' = a\sigma$, $\tau' = b\tau$, and $d' = (b/a)^r d$ $(a, b > 0)$, and an estimator δ is equivariant under these transformations if $\delta(a\mathbf{x}, b\mathbf{y}) = (b/a)^r \delta(\mathbf{x}, \mathbf{y})$. Generalize Theorems 3.1 and 3.3, Corollary 3.4, and (3.19) to the present situation.

3.35 Under the assumptions of the preceding problem and with loss function $(d-\theta)^2/\theta^2$, determine the MRE estimator of θ in the following situations:

(a) $m = n = 1$ and X and Y are independently distributed as $\Gamma(\alpha, \sigma^2)$ and $\Gamma(\beta, \tau^2)$, respectively (α, β known).

(b) $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independently distributed as $N(0, \sigma^2)$ and $N(0, \tau^2)$, respectively.

(c) $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are independently distributed as $U(0, \sigma)$ and $U(0, \tau)$, respectively.

3.36 Generalize the results of Problem 3.34 to the case that the joint density of $\mathbf{X}$ and $\mathbf{Y}$ is

$$\frac{1}{\sigma^m \tau^n} f\left(\frac{x_1 - \xi}{\sigma}, \ldots, \frac{x_m - \xi}{\sigma}; \frac{y_1 - \eta}{\tau}, \ldots, \frac{y_n - \eta}{\tau}\right).$$

3.37 Obtain the MRE estimator of $\theta = (\tau/\sigma)^r$ with the loss function of Problem 3.35 when the density of Problem 3.36 specializes to

$$\frac{1}{\sigma^m \tau^n} \Pi_i f\left(\frac{x_i - \xi}{\sigma}\right) \Pi_j f\left(\frac{y_j - \eta}{\tau}\right)$$

and f is (a) normal, (b) exponential, or (c) uniform.

3.38 In the model of Problem 3.37 with $\tau = \sigma$, discuss the equivariant estimation of $\Delta = \eta - \xi$ with loss function $(d - \Delta)^2/\sigma^2$ and obtain explicit results for the three distributions of that problem.

3.39 Suppose in Problem 3.37 that an MRE estimator δ^* of $\Delta = \eta - \xi$ under the transformations $X_i' = a + bX_i$ and $Y_j' = a + bY_j$, $b > 0$, exists when the ratio $\tau/\sigma = c$ is known and that δ^* is independent of c. Show that δ^* is MRE also when σ and τ are completely unknown despite the fact that the induced group of transformations of the parameter space is not transitive.

3.40 Let $f(t) = \frac{1}{\pi}\frac{1}{1+t^2}$ be the Cauchy density, and consider the location-scale family

$$\mathcal{F} = \left\{\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), -\infty < \mu < \infty, 0 < \sigma < \infty\right\}.$$

(a) Show that this probability model is invariant under the transformation $x' = 1/x$.

(b) If $\mu' = \mu/(\mu^2+\sigma^2)$ and $\sigma' = \sigma/(\mu^2+\sigma^2)$, show that $P_{\mu,r}(X \in A) = P_{\mu',\sigma'}(X' \in A)$; that is, if X has the Cauchy density with location parameter μ and scale parameter σ, then X' has the Cauchy density with location parameter $\mu/(\mu^2 + \sigma^2)$ and scale parameter $\sigma/(\mu^2 + \sigma^2)$.

(c) Explain why this group of transformations of the sample and parameter spaces does not lead to an invariant estimation problem.

[See McCullaugh (1992) for a full development of this model, where it is suggested that the complex plane provides a more appropriate parameter space.]

3.41 Let (X_i, Y_i), $i = 1, \ldots, n$, be distributed as independent bivariate normal random variables with mean $(\mu, 0)$ and covariance matrix

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

(a) Show that the probability model is invariant under the transformations

$$(x', y') = (a + bx, by),$$
$$(\mu', \sigma'_{11}, \sigma'_{12}, \sigma'_{22}) = (a + b\mu, b^2\sigma_{11}, b^2\sigma_{12}, b^2\sigma_{22}).$$

(b) Using the loss function $L(\mu, d) = (\mu - d)^2/\sigma_{11}$, show that this is an invariant estimation problem, and equivariant estimators must be of the form $\delta = \bar{x} + \psi(u_1, u_2, u_3)\bar{y}$, where $u_1 = \Sigma(x_i - \bar{x})^2/\bar{y}^2$, $u_2 = \Sigma(y_i - \bar{y})^2/\bar{y}^2$, and $u_3 = \Sigma(x_i - \bar{x})(y_i - \bar{y})/\bar{y}^2$.

(c) Show that if δ has a finite second moment, then it is unbiased for estimating μ. Its risk function is a function of σ_{11}/σ_{22} and σ_{12}/σ_{22}.

(d) If the ratio σ_{12}/σ_{22} is known, show that $\bar{X} - (\sigma_{12}/\sigma_{22})\bar{Y}$ is the MRE estimator of μ.

[This problem illustrates the technique of *covariance adjustment*. See Berry, 1987.]

3.42 Suppose we let $X_1, \ldots, X_n$ be a sample from an exponential distribution $f(x|\mu, \sigma) = (1/\sigma)e^{-(x-\mu)/\sigma} I(x \geq \mu)$. The exponential distribution is useful in reliability theory, and a parameter of interest is often a quantile, that is, a parameter of the form $\mu + b\sigma$, where b is known. Show that, under quadratic loss, the MRE estimator of $\mu + b\sigma$ is $\delta_0 = x_{(1)} + (b - 1/n)(\bar{x} - x_{(1)})$, where $x_{(1)} = \min_i x_i$.

[Rukhin and Strawderman (1982) show that δ_0 is inadmissible, and exhibit a class of improved estimators.]

Section 4

4.1 (a) Suppose $X_i : N(\xi_i, \sigma^2)$ with $\xi_i = \alpha + \beta t_i$. If the first column of the matrix C leading to the canonical form (4.7) is $(1/\sqrt{n}, \ldots, 1/\sqrt{n})'$, find the second column of C.

(b) If $X_i : N(\xi_i, \sigma^2)$ with $\xi_i, = \alpha + \beta t_i + \gamma t_i^2$, and the first two columns of C are those of (a), find the third column under the simplifying assumptions $\Sigma t_i = 0$, $\Sigma t_i^2 = 1$. [*Note*: The orthogonal polynomials that are progressively built up in this way are frequently used to simplify regression analysis.]

4.2 Write out explicit expressions for the transformations (4.10) when Π_Ω is given by (a) $\xi_i = \alpha + \beta t_i$ and (b) $\xi_i = \alpha + \beta t_i + \gamma t_i^2$.

4.3 Use Problem 3.10 to prove (iii) of Theorem 4.3.

4.4 (a) In Example 4.7, determine $\hat{\alpha}$, $\hat{\beta}$, and hence $\hat{\xi}_i$ by minimizing $\Sigma(X_i - \alpha - \beta t_i)^2$.

(b) Verify the expressions (4.12) for α and β, and the corresponding expressions for $\hat{\alpha}$ and $\hat{\beta}$.

4.5 In Example 4.2, find the UMVU estimators of α, β, γ, and σ^2 when $\Sigma t_i = 0$ and $\Sigma t_i^2 = 1$.

4.6 Let X_{ij} be independent $N(\xi_{ij}, \sigma^2)$ with $\xi_{ij} = \alpha_i + \beta t_{ij}$. Find the UMVU estimators of the α_i and β.

4.7 (a) In Example 4.9, show that the vectors of the coefficients in the $\hat{\alpha}_i$ are not orthogonal to the vector of the coefficients of $\hat{\mu}$.

(b) Show that the conclusion of (a) is reversed if $\hat{\alpha}_i$ and $\hat{\mu}$ are replaced by $\hat{\hat{\alpha}}_i$ and $\hat{\hat{\mu}}$.

4.8 In Example 4.9, find the UMVU estimator of μ when the α_i are known to be zero and compare it with $\hat{\mu}$.

4.9 The coefficient vectors of the X_{ijk} given by (4.32) for $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$ are orthogonal to the coefficient vectors for the $\hat{\gamma}_{ij}$ given by (4.33).

4.10 In the model defined by (4.26) and (4.27), determine the UMVU estimators of α_i, β_j, and σ^2 under the assumption that the γ_{ij} are known to zero.

4.11 (a) In Example 4.11, show that

$$\Sigma\Sigma\Sigma(X_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2 = S^2 + S_\mu^2 + S_\alpha^2 + S_\beta^2 + S_\gamma^2$$

where $S^2 = \Sigma\Sigma\Sigma(X_{ijk} - X_{ij\cdot})^2$, $S_\mu^2 = IJm(X_{\dots} - \mu)^2$, $S_\alpha^2 = Jm\Sigma(X_{1\cdot\cdot} - X_{\dots} - \alpha_i)^2$, and S_β^2, $S^2\gamma$ are defined analogously.

(b) Use the decomposition of (a) to show that the least squares estimators of $\mu, \alpha_i, \dots$ are given by (4.32) and (4.33).

(c) Show that the *error sum of squares* S^2 is equal to $\Sigma\Sigma\Sigma(X_{ijk} - \hat{\hat{\xi}}_{ij})^2$ and hence in the canonical form to $\Sigma_{j=s+1}^n Y_j^2$.

4.12 (a) Show how the decomposition in Problem 4.11(a) must be modified when it is known that the γ_{ij} are zero.

(b) Use the decomposition of (a) to solve Problem 4.10.

4.13 Let X_{ijk} $(i = 1, \dots, I, j = 1, \dots, J, k = 1, \dots, K)$ be $N(\xi_{ijk}, \sigma^2)$ with

$$\xi_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$$

where $\Sigma\alpha_i = \Sigma\beta_j = \Sigma\gamma_k = 0$. Express μ, α_i, β_j, and γ_k in terms of the ξ's and find their UMVU estimators. Viewed as a special case of (4.4), what is the value of s?

4.14 Extend the results of the preceding problem to the model

$$\xi_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \varepsilon_{ik} + \lambda_{jk}$$

where

$$\sum_i \delta_{ij} = \sum_j \delta_{ij} = \sum_i \varepsilon_{ik} = \sum_k \varepsilon_{ik} = \sum_j \lambda_{jk} = \sum_k \lambda_{jk} = 0.$$

4.15 In the preceding problem, if it is known that the λ's are zero, determine whether the UMVU estimators of the remaining parameters remain unchanged.

4.16 (a) Show that under assumptions (4.35), if $\xi = \theta A$, then the least squares estimate of θ is $\mathbf{x}A(AA')^{-1}$.

(b) If (X, A) is multivariate normal with all parameters unknown, show that the least squares estimator of part (a) is a function of the complete sufficient statistic and, hence, prove part (a) of Theorem 4.14.

4.17 A generalization of the order statistics, to vectors, is given by the following definition.

Definition 8.1 The *c_j-order statistics* of a sample of vectors are the vectors arranged in increasing order according to their jth components.

Let $\mathbf{X}_i$, $i = 1, \dots, n$, be an iid sample of $p \times 1$ vectors, and let $X = (X_1, \dots, X_n)$ be a $p \times n$ matrix.

(a) If the distribution of $\mathbf{X}_i$ is completely unknown, show that, for any j, $j = 1, \dots, p$, the c_j-order statistics of $(X_1, \dots, X_n)$ are complete sufficient. (That is, the vectors $X_1, \dots, X_n$ are ordered according to their jth coordinate.)

(b) Let $Y_{1\times n}$ be a random variable with unknown distribution (possibly different from $\mathbf{X}_i$). Form the $(p-1)\times n$ matrix $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}$, and for any $j = 1, \ldots, p$, calculate the c_j-order statistics based on the columns of $\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}$. Show that these c_j-order statistics are sufficient.

[Hint: See Problem 1.6.33, and also TSH2, Chapter 4, Problem 12.]

(c) Use parts (a) and (b) to prove Theorem 4.14(b).

[Hint: Part (b) implies that only a symmetric function of (X, A) need be considered, and part (a) implies that an unconditionally unbiased estimator must also be conditionally unbiased. Theorem 4.12 then applies.]

4.18 The proof of Theorem 4.14(c) is based on two results. Establish that:

(a) For large values of θ, the unconditional variance of a linear unbiased estimator will be greater than that of the least squares estimator.

(b) For $\theta = 0$, the variance of $XA(AA')^{-1}$ is greater than that of $XA[E(AA')]^{-1}$. [You may use the fact that $E(AA')^{-1} - [E(AA')]^{-1}$ is a positive definite matrix (Marshall and Olkin 1979; Shaffer 1991). This is a multivariate extension of Jensen's inequality.]

(c) Parts (a) and (b) imply that no best linear unbiased estimator of $\Sigma\gamma_i\xi_i$ exists if EAA' is known.

4.19 (a) Under the assumptions of Example 4.15, find the variance of $\Sigma\lambda_i S_i^2$.

(b) Show that the variance of (a) is minimized by the values stated in the example.

4.20 In the linear model (4.4), a function $\Sigma c_i\xi_i$ with $\Sigma c_i = 0$ is called a *contrast*. Show that a linear function $\Sigma d_i\xi_i$ is a contrast if and only if it is translation invariant, that is, satisfies $\Sigma d_i(\xi_i + a) = \Sigma d_i\xi_i$ for all a, and hence if and only if it is a function of the differences $\xi_i - \xi_j$.

4.21 Determine which of the following are contrasts:

(a) The regression coefficients α, β, or γ of (4.2).

(b) The parameters μ, α_i, β_j, or γ_{ij} of (4.27).

(c) The parameters μ or α_i of (4.23) and (4.24).

Section 5

5.1 In Example 5.1:

(a) Show that the joint density of the Z_{ij} is given by (5.2).

(b) Obtain the joint multivariate normal density of the X_{ij} directly by evaluating their covariance matrix and then inverting it.

[*Hint*: The covariance matrix of $X_{11}, \ldots, X_{1n}; \ldots; X_{s1}, \ldots, X_{sn}$ has the form

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 & \ldots & 0 \\ 0 & \Sigma_2 & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & \Sigma_s \end{pmatrix}$$

where each Σ_i is an $n\times n$ matrix with a value a_i for all diagonal elements and a value b_i for all off-diagonal elements. For the inversion of Σ_i, see the next problem.]

5.2 Let $A = (a_{ij})$ be a nonsingular $n \times n$ matrix with $a_{ii} = a$ and $a_{ij} = b$ for all $i \neq j$. Determine the elements of A^{-1}. [*Hint*: Assume that $A^{-1} = (c_{ij})$ with $c_{ii} = c$ and $c_{ij} = d$ for all $i \neq j$, calculate c and d as the solutions of the two linear equations $\Sigma a_{1j} c_{j1} = 1$ and $\Sigma a_{1j} c_{j2} = 0$, and check the product AC.]

5.3 Verify the UMVU estimator of σ_A^2/σ^2 given in Example 5.1.

5.4 Obtain the joint density of the X_{ij} in Example 5.1 in the unbalanced case in which $j = 1, \ldots, n_i$, with the n_i not all equal, and determine a minimal set of sufficient statistics (which depends on the number of distinct values of n_i).

5.5 In the balanced one-way layout of Example 5.1, determine $\lim P(\hat{\sigma}_A^2 < 0)$ as $n \to \infty$ for $\sigma_A^2/\sigma^2 = 0, 0.2, 0.5$, 1, and $s = 3, 4, 5, 6$. [*Hint*: The limit of the probability can be expressed as a probability for a χ^2_{s-1} variable.]

5.6 In the preceding problem, calculate values of $P(\hat{\sigma}_A^2 < 0)$ for finite n. When would you expect negative estimates to be a problem? [The probability $P(\hat{\sigma}_A^2 < 0)$, which involves an F random variable, can also be expressed using the incomplete beta function, whose values are readily available through either extensive tables or computer packages. Searle et al. (1992, Section 3.5d) look at this problem in some detail.]

5.7 The following problem shows that in Examples 5.1–5.3 every unbiased estimator of the variance components (except σ^2) takes on negative values. (For some related results, see Pukelsheim 1981.)

Let X have distribution $P \in \mathcal{P}$ and suppose that T is a complete sufficient statistic for $\mathcal{P}$. If $g(P)$ is any U-estimable function defined over $\mathcal{P}$ and its UMVU estimator $\eta(T)$ takes on negative values with probability > 0, then show that this is true of every unbiased estimator of $g(P)$. [*Hint*: For any unbiased estimator δ, recall that $E(\delta|T) = \eta(T)$.]

5.8 Modify the car illustration of Example 5.1 so that it illustrates (5.5).

5.9 In Example 5.2, define a linear transformation of the X_{ijk} leading to the joint distribution of the Z_{ijk} stated in connection with (5.6), and verify the complete sufficient statistics (5.7).

5.10 In Example 5.2, obtain the UMVU estimators of the variance components σ_A^2, σ_B^2, and σ^2 when $\sigma_C^2 = 0$, and compare them to those obtained without this assumption.

5.11 For the X_{ijk} given in (5.8), determine a transformation taking them to variables Z_{ijk} with the distribution stated in Example 5.3.

5.12 In Example 5.3, obtain the UMVU estimators of the variance components σ_A^2, σ_B^2, and σ^2.

5.13 In Example 5.3, obtain the UMVU estimators of σ_A^2 and σ^2 when $\sigma_B^2 = 0$ so that the B terms in (5.8) drop out, and compare them with those of Problem 5.12.

5.14 In Example 5.4:

(a) Give a transformation taking the variables X_{ijk} into the W_{ijk} with density (5.11).

(b) Obtain the UMVU estimators of μ, α_i, σ_B^2, and σ^2.

5.15 A general class of models containing linear models of Types I and II, and mixed models as special cases assumes that the $1 \times n$ observation vector $\mathbf{X}$ is normally distributed with mean $\boldsymbol{\theta} A$ as in (4.13) and with covariance matrix $\Sigma_{i=1}^m \gamma_i V_i$ where the γ's are the components of variance and the V_i's are known symmetric positive semidefinite $n \times n$ matrices. Show that the following models are of this type and in each case specify the γ's and V's: (a) (5.1); (b) (5.5); (c) (5.5) without the terms C_{ij}; (d) (5.8); (e) (5.10).

5.16 Consider a nested three-way layout with

$$X_{ijkl} = \mu + a_i + b_{ij} + c_{ijk} + U_{ijkl}$$

$(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K; l = 1, \ldots, n)$ in the versions

(a) $a_i = \alpha_i, b_{ij} = \beta_{ij}, c_{ijk} = \gamma_{ijk}$;
(b) $a_i = \alpha_i, b_{ij} = \beta_{ij}, c_{ijk} = C_{ijk}$;
(c) $a_i = \alpha_i, b_{ij} = B_{ij}, c_{ijk} = C_{ijk}$;
(d) $a_i = A_i, b_{ij} = B_{ij}, c_{ijk} = C_{ijk}$;

where the α's, β's, and γ's are unknown constants defined uniquely by the usual conventions, and the A's, B's, C's, and U's are unobservable random variables, independently normally distributed with means zero and with variances $\sigma_A^2, \sigma_B^2, \sigma_C^2$ and σ^2.

In each case, transform the X_{ijkl} to independent variables Z_{ijkl} and obtain the UMVU estimators of the unknown parameters.

5.17 For the situation of Example 5.5, relax the assumption of normality to only assume that A_i and U_{ij} have zero means and finite second moments. Show that among all linear estimators (of the form $\sum c_{ij}x_{ij}$, c_{ij} known), the UMVU estimator of $\mu + \alpha_i$ (the best linear predictor) is given by (5.14).

[This is a Gauss-Markov theorem for prediction in mixed models. See Harville (1976) for generalizations.]

Section 6

6.1 In Example 6.1, show that $\gamma_{ij} = 0$ for all i, j is equivalent to $p_{ij} = p_{i+}p_{+j}$. [*Hint*: $\gamma_{ij} = \xi_{ij} - \xi_{i\cdot} - \xi_{\cdot j} + \xi_{\cdot\cdot} = 0$ implies $p_{ij} = a_i b_j$ and hence $p_{i+} = ca_i$ and $p_{+j} = b_j/c$ for suitable a_i, b_j, and $c > 0$.]

6.2 In Example 6.2, show that the conditional independence of A, B given C is equivalent to $\alpha_{ijk}^{ABC} = \alpha_{ij}^{AB} = 0$ for all i, j, and k.

6.3 In Example 6.1, show that the conditional distribution of the vectors $(n_{i1}, \ldots, n_{iJ})$ given the values of n_{i+} $(i = 1, \ldots, I)$ is that of I independent vectors with multinomial distribution $M(p_{1|i}, \ldots, p_{J|i}; n_{i+})$ where $p_{j|i} = p_{ij}/p_{i+}$.

6.4 Show that the distribution of the preceding problem also arises in Example 6.1 when the n subjects, rather than being drawn from the population at large, are randomly drawn: n_{1+} from Category $A_1, \ldots, n_{I+}$ from Category A_I.

6.5 An application of log linear models in genetics is through the *Hardy-Weinberg* model of mating. If a parent population contains alleles A, a with frequencies p and $1 - p$, then standard random mating assumptions will result in offspring with genotypes AA, Aa, and aa with frequencies $\theta_1 = p^2, \theta_2 = 2p(1 - p)$, and $\theta_3 = (1 - p)^2$.

(a) Give the full multinomial model for this situation, and show how the Hardy-Weinberg model is a non-full-rank submodel.
(b) For a sample $X_1, \ldots, X_n$ of n offspring, find the minimal sufficient statistic.

[See Brown (1986a) for a more detailed development of this model.]

6.6 A city has been divided into I major districts and the ith district into J_i subdistricts, all of which have populations of roughly equal size. From the police records for a given year, a random sample of n robberies is obtained. Write the joint multinomial distribution of the numbers n_{ij} of robberies in subdistrict (i, j) for this nested two-way layout as $e^{\Sigma\Sigma n_{ij}\xi_{ij}}$ with $\xi_{ij} = \mu + \alpha_i + \beta_{ij}$ where $\Sigma_i \alpha_i = \Sigma_j \beta_{ij} = 0$, and show that the assumption $\beta_{ij} = 0$ for all i, j is equivalent to the assumption that $p_{ij} = p_{i+}/J_i$ for all i, j.

6.7 Instead of a sample of fixed size n in the preceding problem, suppose the observations consist of all robberies taking place within a given time period, so that n is the value taken on by a random variable N. Suppose that N has a Poisson distribution with unknown expectation λ and that the conditional distribution of the n_{ij} given $N = n$ is the distribution assumed for the n_{ij} in the preceding problem. Find the UMVU estimator of λp_{ij} and show that no unbiased estimator p_{ij} exists. [*Hint*: See the following problem.]

6.8 Let N be an integer-valued random variable with distribution $P_\theta(N = n) = P_\theta(n)$, $n = 0, \ldots,$ for which N is complete. Given $N = n$, let X have the binomial distribution $b(p, n)$ for $n > 0$, with p unknown, and let $X = 0$ when $n = 0$. For the observations (N, X):

(a) Show that (N, X) is complete.

(b) Determine the UMVU estimator of $pE_\theta(N)$.

(c) Show that no unbiased estimator of any function $g(p)$ exists if $P_\theta(0) > 0$ for some θ.

(d) Determine the UMVU estimator of p if $P_\theta(0)$ for all θ.

Section 7

7.1 (a) Consider a population $\{a_1, \ldots, a_N\}$ with the parameter space defined by the restriction $a_1 + \cdots + a_N = A$ (known). A simple random sample of size n is drawn in order to estimate τ^2. Assuming the labels to have been discarded, show that $Y_{(1)}, \ldots, Y_{(n)}$ are not complete.

(b) Show that Theorem 7.1 need not remain valid when the parameter space is of the form $V_1 \times V_2 \times \cdots \times V_N$. [*Hint*: Let $N = 2, n = 1, V_1 = \{1, 2\}, V_2 = \{3, 4\}$.]

7.2 If $Y_1, \ldots, Y_n$ are the sample values obtained in a simple random sample of size n from the finite population (7.2), then (a) $E(Y_i) = \bar{a}$, (b) $\text{var}(Y_i) = \tau^2$, and (c) $\text{cov}(Y_i, Y_j) = -\tau^2/(N - 1)$.

7.3 Verify equations (a) (7.6), (b) (7.8), and (c) (7.13).

7.4 For the situation of Example 7.4:

(a) Show that $E\bar{Y}_{\nu-1} = E[\frac{1}{\nu-1}\sum_1^{\nu-1} Y_i] = \bar{a}$.

(b) Show that $[\frac{1}{\nu-1} - \frac{1}{N}]\frac{1}{\nu-2}\sum_1^{\nu-1}(Y_i - \bar{Y}_{\nu-1})^2$ is an unbiased estimator of $\text{var}(\bar{Y}_{\nu-1})$.

[Pathak (1976) proved (a) by first showing that $EY_1 = \bar{a}$, and then that $EY_1|T_0 = \bar{Y}_{\nu-1}$. To avoid trivialities, Pathak also assumes that $C_i + C_j < Q$ for all i, j, so that at least three observations are taken.]

7.5 Random variables $X_1, \ldots, X_n$ are *exchangeable* if any permutation of $X_1, \ldots, X_n$ has the same distribution.

(a) If $X_1, \ldots, X_n$ are iid, distributed as Bernoulli (p), show that given $\sum_1^n X_i = t$, $X_1, \ldots, X_n$ are exchangeable (but not independent).

(b) For the situation of Example 7.4, show that given $T = \{(C_1, X_1), \ldots, (C_\nu, X_\nu)\}$, the $\nu - 1$ preterminal observations are exchangeable.

The idea of exchangeability is due to deFinetti (1974), who proved a theorem that characterizes the distribution of exchangeable random variables as mixtures of iid random variables. Exchangeable random variables play a large role in Bayesian statistics; see Bernardo and Smith 1994 (Sections 4.2 and 4.3).

7.6 For the situation of Example 7.4, assuming that (a) and (b) hold:

(a) Show that $\hat{a}$ of (7.9) is UMVUE for $\bar{a}$.

(b) Defining $S^2 = \sum_{i=1}^{\nu}(Y_i - \bar{Y})/(\nu - 1)$, show that

$$\hat{\sigma}^2 = S^2 - \frac{MS_{[v]}\frac{\nu}{\nu-1} - S^2}{\nu - 2}$$

is UMVUE for τ^2 of (7.7), where $MS_{[v]}$ is the variance of the observations in the set (7.10).

[Kremers (1986) uses conditional expectation arguments (Rao-Blackwellization), and completeness, to establish these results. He also assumes that at least n_0 observations are taken. To avoid trivialities, we can assume $n_0 \geq 3$.]

7.7 In simple random sampling, with labels discarded, show that a necessary condition for $h(a_1, \ldots, a_N)$ to be U-estimable is that h is symmetric in its N arguments.

7.8 Prove Theorem 7.7.

7.9 Show that the approximate variance (7.16) for stratified sampling with $n_i = nN_i/N$ (proportional allocation) is never greater than the corresponding approximate variance τ^2/n for simple random sampling with the same total sample size.

7.10 Let V_p be the exact variance (7.15) and V_r the corresponding variance for simple random sampling given by (7.6) with $n = \Sigma n_i$, $N = \Sigma N_i$, $n_i/n = N_i/N$ and $\tau^2 = \Sigma\Sigma(a_{ij} - a_{..})^2/N$.

(a) Show that $V_r - V_p = \frac{N-n}{n(N-1)N}\left[\Sigma N_i(a_{i.} - a_{..})^2 - \frac{1}{N}\Sigma\frac{N-N_i}{N_i-1}N_i\tau_i^2\right]$.

(b) Give an example in which $V_r < V_p$.

7.11 The approximate variance (7.16) for stratified sampling with a total sample size $n = n_1 + \cdots + n_s$ is minimized when n_i is proportional to $N_i\tau_i$.

7.12 For sampling designs where the inclusion probabilities $\pi_i = \sum_{s:i\in s} P(s)$ of including the ith sample value Y_i is known, a frequently used estimator of the population total is the Horvitz-Thompson (1952) estimator $\delta_{HT} = \sum_i Y_i/\pi_i$.

(a) Show that δ_{HT} is an unbiased estimator of the population total.

(b) The variance of δ_{HT} is given by

$$\mathrm{var}(\delta_{HT}) = \sum_i Y_i^2\left[\frac{1}{\pi_i} - 1\right] + \sum_{i\neq j} Y_iY_j\left[\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right],$$

where π_{ij} are the *second-order inclusion probabilities* $\pi_{ij} = \sum_{s:i,j\in s} P(s)$.

Note that it is necessary to know the labels in order to calculate δ_{HT}, thus Theorem 7.5 precludes any overall optimality properties. See Hedayat and Sinha 1991 (Chapters 2 and 3) for a thorough treatment of δ_{HT}.

7.13 Suppose that an auxiliary variable is available for each element of the population (7.2) so that $\theta = \{(1, a_1, b_1), \ldots, (N, a_N, b_N)\}$. If $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ denote the values of a and b observed in a simple random sample of size n, and $\bar{Y}$ and $\bar{Z}$ denote their averages, then

$$\mathrm{cov}(\bar{Y}, \bar{Z}) = E(\bar{Y} - \bar{a})(\bar{Z} - \bar{b}) = \frac{N-n}{nN(N-1)}\Sigma(a_i - \bar{a})(b_i - \bar{b}).$$

7.14 Under the assumptions of Problem 7.13, if $B = b_1 + \cdots + b_N$ is known, an alternative unbiased estimator $\bar{a}$ is

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{Z_i}\right)\bar{b} + \frac{n(N-1)}{(n-1)N}\left[\bar{Y} - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{Z_i}\right)\bar{Z}\right].$$

[*Hint*: Use the facts that $E(Y_1/Z_1) = (1/N)\Sigma(a_i/b_i)$ and that by the preceding problem

$$E\left[\frac{1}{n-1}\Sigma\frac{Y_i}{Z_i}(Z_i - \bar{Z})\right] = \left[\frac{1}{N-1}\Sigma\frac{a_i}{b_i}(b_i - \bar{b})\right].]$$

7.15 In connection with cluster sampling, consider a set W of vectors $(a_1, \ldots, a_M)$ and the totality G of transformations taking $(a_1, \ldots, a_M)$ into $(a'_1, \ldots, a'_M)$ such that $(a'_1, \ldots, a'_M) \in W$ and $\Sigma a'_i = \Sigma a_i$. Give examples of W such that for any real number a_1 there exist $a_2, \ldots, a_M$ with $(a_1, \ldots, a_M) \in W$ and such that

(a) G consists of the identity transformation only;

(b) G consists of the identity and one other element;

(c) G is transitive over W.

7.16 For cluster sampling with unequal cluster sizes M_i, Problem 7.14 provides an alternative estimator of $\bar{a}$, with M_i in place of b_i. Show that this estimator reduces to $\bar{Y}$ if $b_1 = \cdots = b_N$ and hence when the M_i are equal.

7.17 Show that (7.17) holds if and only if δ depends only on X'', defined by (7.18).

9 Notes

9.1 History

The theory of equivariant estimation of location and scale parameters is due to Pitman (1939), and the first general discussions of equivariant estimation were provided by Peisakoff (1950) and Kiefer (1957). The concept of risk-unbiasedness (but not the term) and its relationship to equivariance were given in Lehmann (1951).

The linear models of Section 3.4 and Theorem 4.12 are due to Gauss. The history of both is discussed in Seal (1967); see also Stigler 1981. The generalization to exponential linear models was introduced by Dempster (1971) and Nelder and Wedderburn (1972).

The notions of *Functional Equivariance* and *Formal Invariance*, discussed in Section 3.2, have been discussed by other authors sometimes using different names. Functional Equivariance is called the *Principle of Rational Invariance* by Berger (1985, Section 6.1), *Measurement Invariance* by Casella and Berger (1990, Section 7.2.4) and *Parameter Invariance* by Dawid (1983). Schervish (1995, Section 6.2.2) argues that this principle is really only a reparametrization of the problem, and has nothing to do with invariance. This is almost in agreement with the principle of functional equivariance, however, it is still the case that when reparameterizing one must be careful to properly reparameterize the estimator, density, and loss function, which is part of the prescription of an invariant problem. This type of invariance is commonly illustrated by the example that if δ measures temperature in degrees Celsius, then $(9/5)\delta + 32$ should be used to measure temperature in degrees Fahrenheit (see Problems 2.9 and 2.10).

What we have called *Formal Invariance* was also called by that name in Casella and Berger (1990), but was called the *Invariance Principle* by Berger (1985) and *Context Invariance* by Dawid (1983).

(Problem 7.6), and hence toward a submodel (subspace) of dimension zero. In the analysis of variance, Example 7.7, the subspace of the submodel has dimension 1, $\{(\xi_1, \ldots, \xi_s) : \xi_i = \mu, i = 1, \ldots, s\}$, and in Example 7.8, it has dimension 2, $\{(\xi_1, \ldots, \xi_s) : \{i = \alpha + \beta t_i, i = 1, \ldots, s\}$. In general, the empirical Bayes strategies developed here will only work if the dimension of the submodel, r, is at least two fewer than that of the full model, s; that is, $s - r > 2$. This is a technical requirement, as the marginal distribution of interest is χ^2_{s-r}, and estimation is problematic if $s - r \leq 2$. The reason for this difficulty is the need to calculate the expectation $E(1/\chi^2_{s-r})$, which is infinite if $s - r \leq 2$. (See Problem 7.6; also see Problem 6.12 for an attempt at empirical Bayes if $s - r \leq 2$.)

In light of Theorem 7.5, we can improve the empirical Bayes estimators of Examples 7.7 and 7.8 by using their positive-part version. Moreover, Problem 7.8 shows that such an improvement will hold throughout the entire exponential family. Thus, the strategy of taking a positive part should always be employed in these cases of empirical Bayes estimation.

Finally, we note that Examples 7.7 and 7.8 can be greatly generalized. One can handle unequal n_i, unequal variances, full covariance matrices, general linear submodels, and more. In some cases, the algebra can become somewhat overwhelming, and details about performance of the estimators may become obscured. We examine a number of these cases in Problems 7.16–7.18.

8 Problems

Section 1

1.1 Verify the expressions for $\pi(\lambda|\bar{x})$ and $\delta^k(\bar{x})$ in Example 1.3.

1.2 Give examples of pairs of values (a, b) for which the beta density $B(a, b)$ is (a) decreasing, (b) increasing, (c) increasing for $p < p_0$ and decreasing for $p > p_0$, and (d) decreasing for $p < p_0$ and increasing for $p > p_0$.

1.3 In Example 1.5, if p has the improper prior density $\frac{1}{p(1-p)}$, show that the posterior density of p given x is proper, provided $0 < x < n$.

1.4 In Example 1.5, find the Jeffreys prior for p and the associated Bayes estimator δ_Λ.

1.5 For the estimator δ_Λ of Problem 1.4,

(a) calculate the bias and maximum bias;

(b) calculate the expected squared error and compare it with that of the UMVU estimator.

1.6 In Example 1.5, find the Bayes estimator δ of $p(1-p)$ when p has the prior $B(a, b)$.

1.7 For the situation of Example 1.5, the UMVU estimator of $p(1-p)$ is $\delta' = [x(x-1)]/[n(n-1)]$ (see Example 2.3.1 and Problem 2.3.1).

(a) Compare the estimator δ of Problem 1.6 with the UMVU estimator δ'.

(b) Compare the expected squared error of the estimator of $p(1-p)$ for the Jeffreys prior in Example 1.5 with that of δ'.

1.8 In analogy with Problem 1.2, determine the possible shapes of the gamma density $\Gamma(g, 1/\alpha), \alpha, g > 0$.

1.9 Let $X_1, \ldots, X_n$ be iid according to the Poisson distribution $P(\lambda)$ and let λ have a gamma distribution $\Gamma(g, \alpha)$.

(a) For squared error loss, show that the Bayes estimator $\delta_{\alpha,g}$ of λ has a representation analogous to (1.1.13).

(b) What happens to $\delta_{\alpha,g}$ as (i) $n \to \infty$, (ii) $\alpha \to \infty$, $g \to 0$, or both?

1.10 For the situation of the preceding problem, solve the two parts corresponding to Problem 1.5(a) and (b).

1.11 In Problem 1.9, if λ has the improper prior density $d\lambda/\lambda$ (corresponding to $\alpha = g = 0$), under what circumstances is the posterior distribution proper?

1.12 Solve the problems analogous to Problems 1.9 and 1.10 when the observations consist of a single random variable X having a negative binomial distribution $Nb(p, m)$, p has the beta prior $B(a, b)$, and the estimand is (a) p and (b) $1/p$.

Section 2

2.1 Referring to Example 1.5, suppose that X has the binomial distribution $b(p, n)$ and the family of prior distributions for p is the family of beta distributions $B(a, b)$.

(a) Show that the marginal distribution of X is the *beta-binomial* distribution with mass function

$$\binom{n}{x} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(x+a)\Gamma(n-x+b)}{\Gamma(n+a+b)}.$$

(b) Show that the mean and variance of the beta-binomial is given by

$$EX = \frac{na}{a+b} \quad \text{and} \quad \text{var } X = n\left(\frac{a}{a+b}\right)\left(\frac{b}{a+b}\right)\left(\frac{a+b+n}{a+b+1}\right).$$

[*Hint*: For part (b), the identities $EX = E[E(X|p)]$ and $\text{var } X = \text{var}[E(X|p)] + E[\text{var}(X|p)]$ are helpful.]

2.2 For the situation of Example 2.1, Lindley and Phillips (1976) give a detailed account of the effect of stopping rules, which we can illustrate as follows. Let X be the number of successes in n Bernoulli trials with success probability p.

(a) Suppose that the number of Bernoulli trials performed is a prespecifed number n, so that we have the binomial sampling model, $P(X = x) = \binom{n}{x} p^x(1-p)^{n-x}$, $x - 0, 1, \ldots, n$. Calculate the Bayes risk of the Bayes estimator (1.1.12) and the UMVU estimator of p.

(b) Suppose that the number of Bernoulli trials performed is a random variable N. The value $N = n$ was obtained when a prespecified number, x, of successes was observed so that we have the negative binomial sample model, $P(N = n) = \binom{n-1}{x-1} p^x(1-p)^{n-x}$, $n = x$. Calculate the Bayes risk of the Bayes estimator and the UMVU estimator of p.

(c) Calculate the mean squared errors of all three estimators under each model. If it is unknown which sampling mechanism generated the data, which estimator do you prefer overall?

2.3 Show that the estimator (2.2.4) tends in probability (a) to θ as $n \to \infty$, (b) to μ as $b \to 0$, and (c) to θ as $b \to \infty$.

2.4 Bickel and Mallows (1988) further investigate the relationship between unbiasedness and Bayes, specifying conditions under which these properties cannot hold simultaneously. In addition, they show that if a prior distribution is improper, then a posterior mean can be unbiased. Let $X \sim \frac{1}{\theta} f(x/\theta), x > 0$, where $\int_0^\infty t f(t) dt = 1$, and let $\pi(\theta) = \frac{1}{\theta^2} d\theta$, $\theta > 0$.

(a) Show that $E(X|\theta) = \theta$, so X is unbiased.

(b) Show that $\pi(\theta|x) = \frac{x^2}{\theta^3} f(x/\theta)$ is a proper density.

(c) Show that $E(\theta|x) = x$, and hence the posterior mean, is unbiased.

2.5 DasGupta (1994) presents an identity relating the Bayes risk to bias, which illustrates that a small bias can help achieve a small Bayes risk. Let $X \sim f(x|\theta)$ and $\theta \sim \pi(\theta)$. The Bayes estimator under squared error loss is $\delta^\pi = E(\theta|x)$. Show that the Bayes risk of δ^π can be written

$$r(\pi, \delta^\pi) = \int_\Theta \int_{\mathcal{X}} [\theta - \delta^\pi(x)]^2 f(x|\theta)\pi(\theta) dx d\theta = \int_\Theta \theta b(\theta)\pi(\theta) d\theta$$

where $b(\theta) = E[\delta^\pi(X)|\theta] - \theta$ is the bias of δ^π.

2.6 Verify the estimator (2.2.10).

2.7 In Example 2.6, verify that the posterior distribution of τ is $\Gamma(r + g - 1/2, 1/(\alpha + z))$.

2.8 In Example 2.6 with $\alpha = g = 0$, show that the posterior distribution given the X's of $\sqrt{n}(\theta - \bar{X})/\sqrt{Z/(n-1)}$ is Student's t-distribution with $n - 1$ degrees of freedom.

2.9 In Example 2.6, show that the posterior distribution of θ is symmetric about $\bar{x}$ when the joint prior of θ and σ is of the form $h(\sigma) d\sigma\, d\theta$, where h is an arbitrary probability density on $(0, \infty)$.

2.10 Rukhin (1978) investigates the situation when the Bayes estimator is the same for every loss function in a certain set of loss functions, calling such estimators *universal Bayes estimators*. For the case of Example 2.6, using the prior of the form of Problem 2.9, show that $\bar{X}$ is the Bayes estimator under every even loss function.

2.11 Let X and Y be independently distributed according to distributions P_ξ and Q_η, respectively. Suppose that ξ and η are real-valued and independent according to some prior distributions Λ and Λ'. If, with squared error loss, δ_Λ is the Bayes estimator of ξ on the basis of X, and $\delta'_{\Lambda'}$ is that of η on the basis of Y,

(a) show that $\delta'_{\Lambda'} - \delta_\Lambda$ is the Bayes estimator of $\eta - \xi$ on the basis of (X, Y);

(b) if $\eta > 0$ and $\delta^*_{\Lambda'}$ is the Bayes estimator of $1/\eta$ on the basis of Y, show that $\delta_\Lambda \cdot \delta^*_{\Lambda'}$ is the Bayes estimator of ξ/η on the basis of (X, Y).

2.12 For the density (2.2.13) and improper prior $(d\sigma/\sigma) \cdot (d\sigma_A/\sigma_A)$, show that the posterior distribution of (σ, σ_A) continues to be improper.

2.13 (a) In Example 2.7, obtain the Jeffreys prior distribution of (σ, τ).

(b) Show that for the prior of part (a), the posterior distribution of (σ, τ) is proper.

2.14 Verify the Bayes estimator (2.2.14).

2.15 Let $X \sim N(\theta, 1)$ and $L(\theta, \delta) = (\theta - \delta)^2$.

(a) Show that X is the limit of the Bayes estimators δ^{π_n}, where π_n is $N(0, 1)$. Hence, X is both generalized Bayes and a limit of Bayes estimators.

(b) For the prior measure $\pi(\theta) = e^{a\theta}, a > 0$, show that the generalized Bayes estimator is $X + a$.

(c) For $a > 0$, show that there is no sequence of proper priors for which $\delta^{\pi_n} \to X + a$.

This example is due to Farrell; see Kiefer 1966. Heath and Sudderth (1989), building on the work of Stone (1976), showed that inferences from this model are *incoherent*, and established when generalized Bayes estimators will lead to *coherent* (that is, noncontradictory) inferences. Their work is connected to the theory of "approximable by proper priors," developed by Stein (1965) and Stone (1965, 1970, 1976), which shows when generalized Bayes estimators can be looked upon as Bayes estimators.

2.16 (a) For the situation of Example 2.8, verify that $\delta(x) = x/n$ is a generalized Bayes estimator.

(b) If $X \sim N(0, 1)$ and $L(\theta, \delta) = (\theta - \delta)^2$, show that X is generalized Bayes under the improper prior $\pi(\theta) = 1$.

Section 3

3.1 For the situation of Example 3.1:

(a) Verify that the Bayes estimator will only depend on the data through $Y = \max_i X_i$.

(b) Show that $E(\Theta|y, a, b)$ can be expressed as

$$E(\Theta|y, a, b) = \frac{1}{b(n+a-1)} \frac{P(\chi^2_{2(n+a-1)} < 2/by)}{P(\chi^2_{2(n+a)} < 2/by)}$$

where χ^2_ν is a chi-squared random variable with ν degrees of freedom. (In this form, the estimator is particularly easy to calculate, as many computer packages will have the chi-squared distribution built in.)

3.2 Let $X_1, \ldots, X_n$ be iid from Gamma(a, b) where a is known.

(a) Verify that the conjugate prior for the natural parameter $\eta = -1/b$ is equivalent to an inverted gamma prior on b.

(b) Using the prior in part (a), find the Bayes estimator under the losses (i) $L(b, \delta) = (b - \delta)^2$ and (ii) $L(b, \delta) = (1 - \delta/b)^2$.

(c) Express the estimator in part (b)(i) in the form (3.3.9). Can the same be done for the estimator in part (b)(ii)?

3.3 (a) Prove Corollary 3.3.

(b) Verify the calculation of the Bayes estimator in Example 3.4.

3.4 Using Stein's identity (Lemma 1.5.15), show that if $X_i \sim p_{\eta_i}(x)$ of (3.3.7), then

$$E_\eta(-\nabla \log h(\mathbf{X})) = \boldsymbol{\eta},$$
$$R(\boldsymbol{\eta}, -\nabla \log h(\mathbf{X})) = \sum_{i=1}^p E_\eta \left[-\frac{\partial^2}{\partial X_i^2} \log h(\mathbf{X}) \right].$$

3.5 (a) If $X_i \sim$ Gamma(a, b), $i = 1, \ldots, p$, independent with a known, calculate $-\nabla \log h(\mathbf{x})$ and its expected value.

(b) Apply the results of part (a) to the situation where $X_i \sim N(0, \sigma_i^2)$, $i = 1, \ldots, p$, independent. Does it lead to an unbiased estimator of σ_i^2?

[*Note*: For part (b), squared error loss on the natural parameter $1/\sigma^2$ leads to the loss $L(\sigma^2, \delta) = (\sigma^2\delta - 1)^2/\sigma^4$ for estimation of σ^2.]

(c) If

$$X_i \sim \frac{\tan(a_i\pi)}{\pi} x^{a_i}(1-x)^{-1}, \; 0 < x < 1, \quad i = 1, \ldots, p, \text{ independent},$$

evaluate $-\nabla \log h(\mathbf{X})$ and show that it is an unbiased estimator of $\mathbf{a} = (a_1, \ldots, a_p)$.

3.6 For the situation of Example 3.6:

(a) Show that if δ is a Bayes estimator of θ, then $\delta' = \delta/\sigma^2$ is a Bayes estimator of η, and hence $R(\theta, \delta) = \sigma^4 R(\eta, \delta')$.

(b) Show that the risk of the Bayes estimator of η is given by

$$\frac{p\tau^4}{\sigma^2(\sigma^2+\tau^2)^2} + \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)^2 \sum a_i^2,$$

where $a_i = \eta_i - \mu/\sigma^2$.

(c) If $\Sigma a_i^2 = k$, a fixed constant, then the minimum risk is attained at $\eta_i = \mu/\sigma^2 + \sqrt{k/p}$.

3.7 If $\mathbf{X}$ has the distribution $p_\theta(\mathbf{x})$ of (1.5.1) show that, similar to Theorem 3.2, $E(\mathcal{T}\boldsymbol{\eta}(\theta)) = \nabla \log m_\pi(\mathbf{x}) - \nabla \log h(\mathbf{x})$.

3.8 (a) Use Stein's identity (Lemma 1.5.15) to show that if $X_i \sim p_{\eta_i}(x)$ of (3.3.18), then

$$E_\eta(-\nabla \log h(\mathbf{X})) = \sum_i \eta_i E_\eta \frac{\partial}{\partial X_j} T_i(\mathbf{X}).$$

(b) If X_i are iid from a gamma distribution Gamma(a, b), where the shape parameter a is known, use part (a) to find an unbiased estimator of $1/b$.

(c) If the X_i are iid from a beta(a, b) distribution, can the identity in part (a) be used to obtain an unbiased estimator of a when b is known, or an unbiased estimator of b when a is known?

3.9 For the natural exponential family $p_\eta(x)$ of (3.3.7) and the conjugate prior $\pi(\eta|k, \mu)$ of (3.3.19) establish that:

(a) $E(X) = A'(\eta)$ and var $X = A''(\eta)$, where the expectation is with respect to the sampling density $p_\eta(x)$.

(b) $EA'(\eta) = \mu$ and var$[A(\eta)] = (1/k)EA''(\eta)$, where the expectation is with respect to the prior distribution.

[The results in part (b) enable us to think of μ as a prior mean and k as a prior sample size.]

3.10 For each of the following situations, write the density in the form (3.7), and identify the natural parameter. Obtain the Bayes estimator of $A'(\eta)$ using squared loss and the conjugate prior. Express your answer in terms of the original parameters. (a) $X \sim$ binomial(p, n), (b) $X \sim$ Poisson(λ), and (c) $X \sim$ Gamma(a, b), a known.

3.11 For the situation of Problem 3.9, if $X_1, \ldots, X_n$ are iid as $p_\eta(x)$ and the prior is the conjugate $\pi(\eta|k, \mu)$, then the posterior distribution is $\pi(\eta|k+n, \frac{k\mu+n\bar{x}}{k+n})$.

3.12 If $X_1, \ldots, X_n$ are iid from a one-parameter exponential family, the Bayes estimator of the mean, under squared error loss using a conjugate prior, is of the form $a\bar{X} + b$ for constants a and b.

(a) If $EX_i = \mu$ and var $X_i = \sigma^2$, then no matter what the distribution of the X_i's, the mean squared error is

$$E[(a\bar{X}+b) - \mu]^2 = a^2 \text{ var } \bar{X} + [(a-1)\mu + b]^2.$$

(b) If μ is unbounded, then no estimator of the form $a\bar{X} + b$ can have finite mean squared error for $a \neq 1$.

(c) Can a conjugate-prior Bayes estimator in an exponential family have finite mean squared error?

[This problem shows why conjugate-prior Bayes estimators are considered "non-robust."]

Section 4

4.1 For the situation of Example 4.2:

(a) Show that the Bayes rule under a beta(α, α) prior is equivariant.

(b) Show that the Bayes rule under any prior that is symmetric about $1/2$ is equivariant.

4.2 The Bayes estimator of η in Example 4.7 is given by (4.22).

4.3 The Bayes estimator of τ in Example 4.5 is given by (4.22).

4.4 The Bayes estimators of η and τ in Example 4.9 are given by (4.31) and (4.32). (Recall Corollary 1.2.)

4.5 For each of the following situations, find a group G that leaves the model invariant and determine left- and right-invariant measures over G. The joint density of $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and the estimand are

(a) $f(\mathbf{x} - \eta, \mathbf{y} - \zeta)$, estimand $\eta - \zeta$;

(b) $f\left(\frac{\mathbf{x}-\eta}{\sigma}, \frac{\mathbf{y}-\zeta}{\tau}\right)$, estimand τ/σ;

(c) $f\left(\frac{\mathbf{x}-\eta}{\tau}, /, \frac{\mathbf{y}-\zeta}{\tau}\right)$, τ unknown; estimand $\eta - \zeta$.

4.6 For each of the situations of Problem 4.5, determine the MRE estimator if the loss is squared error with a scaling that makes it invariant.

4.7 For each of the situations of Problem 4.5:

(a) Determine the measure over Ω induced by the right-invariant Haar measure over $\bar{G}$;

(b) Determine the Bayes estimator with respect to the measure found in part (a), and show that it coincides with the MRE estimator.

4.8 In Example 4.9, show that the estimator

$$\hat{\tau}(\mathbf{x}) = \frac{\int\int \frac{1}{v^r} f\left(\frac{x_1-u}{v}, \ldots, \frac{x_n-u}{v}\right) dv du}{\int\int \frac{1}{v^{r+1}} f\left(\frac{x_1-u}{v}, \ldots, \frac{x_n-u}{v}\right) dv du}$$

is equivariant under scale changes; that is, it satisfies $\bar{\tau}(c\mathbf{x}) = c\hat{\tau}(\mathbf{x})$ for all values of r for which the integrals in $\hat{\tau}(\mathbf{x})$ exist.

4.9 If Λ is a left-invariant measure over G, show that Λ^* defined by $\Lambda^*(B) = \Lambda(B^{-1})$ is right invariant, where $B^{-1} = \{g^{-1} : g \in B\}$.

[*Hint*: Express $\Lambda^*(Bg)$ and $\Lambda^*(B)$ in terms of Λ.]

4.10 There is a correspondence between Haar measures and Jeffreys priors in the location and scale cases.

(a) Show that in the location parameter case, the Jeffreys prior is equal to the invariant Haar measure.

(b) Show that in the scale parameter case, the Jeffreys prior is equal to the invariant Haar measure.

(c) Show that in the location-scale case, the Jeffreys prior is equal to the left invariant Haar measure.

[Part c) is a source of some concern because, as mentioned in Section 4.4 (see the discussion following Example 4.9), the best-equivariant rule is Bayes against the right-invariant Haar measure (if it exists).]

4.11 For the model (3.3.23), find a measure ν in the (ξ, τ) plane which remains invariant under the transformations (3.3.24).

The next three problems contain a more formal development of left- and right-invariant Haar measures.

4.12 A measure Λ over a group G is said to be right invariant if it satisfies $\Lambda(Bg) = \Lambda(B)$ and left invariant if it satisfies $\Lambda(gB) = \Lambda(B)$. Note that if G is commutative, the two definitions agree.

(a) If the elements $g \in G$ are real numbers $(-\infty < g < \infty)$ and group composition is $g_2 \cdot g_1 = g_1 + g_2$, the measure ν defined by $\nu(B) = \int_B dx$ (i.e., Lebesgue measure) is both left and right invariant.

(b) If the elements $g \in G$ are the positive real numbers, and composition of g_2 and g_1 is multiplication of the two numbers, the measure ν defined by $\nu(B) = \int_B (1/y)\, dy$ is both left and right invariant.

4.13 If the elements $g \in G$ are pairs of real numbers $(a, b), b > 0$, corresponding to the transformations $gx = a + bx$, group composition by (1.4.8) is

$$(a_2, b_2) \cdot (a_1, b_1) = (a_2 + a_1 b_2, b_1 b_2).$$

Of the measures defined by

$$\nu(B) = \iint_B \frac{1}{y}\, dx\, dy \quad \text{and} \quad \nu(B) = \iint_B \frac{1}{y^2}\, dx\, dy,$$

the first is right but not left invariant, and the second is left but not right invariant.

4.14 The four densities defining the measures ν of Problem 4.12 and 4.13 $(dx, (1/y)dy, (1/y)dxdy, (1/y^2)dxdy)$ are the only densities (up to multiplicative constants) for which ν has the stated invariance properties in the situations of these problems.

[*Hint*: In each case, consider the equation

$$\int_B \pi(\theta)\, d\theta = \int_{gB} \pi(\theta)\, d\theta.$$

In the right integral, make the transformation to the new variable or variables $\theta' = g^{-1}\theta$. If J is the Jacobian of this transformation, it follows that

$$\int_B [\pi(\theta) - J\pi(g\theta)]\, d\theta = 0 \quad \text{for all } B$$

and, hence, that $\pi(\theta) = J\pi(g\theta)$ for all θ except in a null set N_g. The proof of Theorem 4 in Chapter 6 of TSH2 shows that N_g can be chosen independent of g. This proves in Problem 4.12(a) that for all $\theta \notin N$, $\pi(\theta) = \pi(\theta + c)$, and hence that $\pi(c) =$ constant a.e. The other three cases can be treated analogously.]

Section 5

5.1 For the model (3.3.1), let $\pi(\theta|x, \lambda)$ be a single-prior Bayes posterior and $\pi(\theta|x)$ be a hierarchical Bayes posterior. Show that $\pi(\theta|x) = \int \pi(\theta|x, \lambda) \cdot \pi(\lambda|x)\, d\lambda$, where $\pi(\lambda|x) = \int f(x|\theta)\pi(\theta|\lambda)\gamma(\lambda)\, d\theta / \iint f(x|\theta)\pi(\theta|\lambda)\gamma(\lambda)\, d\theta\, d\lambda$.

5.2 For the situation of Problem 5.1, show that:

(a) $E(\theta|x) = E[E(\theta|x, \lambda)]$;

(b) $\text{var}(\theta|x) = E[\text{var}(\theta|x, \lambda)] + \text{var}[E(\theta|x, \lambda)]$;

and hence that $\pi(\theta|x)$ will tend to have a larger variance than $\pi(\theta|x, \lambda_0)$.

5.3 For the model (3.3.3), show that:

(a) The marginal prior of θ, unconditional on τ^2, is given by

$$\pi(\theta) = \frac{\Gamma(a+\frac{1}{2})}{\sqrt{2\pi}\,\Gamma(a)b^a} \frac{1}{\left(\frac{1}{b}+\frac{\theta^2}{2}\right)^{a+1/2}},$$

which for $a = \nu/2$ and $b = 2/\nu$ is Student's t-distribution with ν degrees of freedom.

(b) The marginal posterior of τ^2 is given by

$$\pi(\tau^2|\bar{x}) = \frac{\left[\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right]^{1/2} e^{-\frac{1}{2}\frac{\bar{x}^2}{\sigma^2+\tau^2}} \frac{1}{(\tau^2)^{a+3/2}} e^{-1/b\tau^2}}{\int_0^\infty \left[\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right]^{1/2} e^{-frac12\frac{\bar{x}^2}{\sigma^2+\tau^2}} \frac{1}{(\tau^2)^{a+3/2}} e^{-1/b\tau^2} dt^2}.$$

5.4 Albert and Gupta (1985) investigate theory and applications of the hierarchical model

$$\begin{aligned} X_i|\theta_i &\sim b(\theta_i, n), \quad i = 1, \dots, p, \text{ independent}, \\ \theta_i|\eta &\sim \text{beta}[k\eta, k(1-\eta)], \ k \text{ known}, \\ \eta &\sim \text{Uniform}(0, 1). \end{aligned}$$

(a) Show that

$$\begin{aligned} E(\theta_i|\mathbf{x}) &= \left(\frac{n}{n+k}\right)\left(\frac{x_i}{n}\right) + \left(\frac{k}{n+k}\right) E(\eta|\mathbf{x}), \\ \text{var}(\theta_i|\mathbf{x}) &= \frac{k^2}{(n+k)(n+k+1)} \text{var}(\eta|\mathbf{x}). \end{aligned}$$

[Note that $E(\eta|\mathbf{x})$ and $\text{var}(\eta|\mathbf{x})$ are not expressible in a simple form.]

(b) Unconditionally on η, the θ_i's have conditional covariance

$$\text{cov}(\theta_i, \theta_j|\mathbf{x}) = \left(\frac{k}{n+k}\right)^2 \text{var}(\eta|\mathbf{x}), \quad i \neq j.$$

(c) Ignoring the prior distribution of η, show how to construct an empirical Bayes estimator of θ_i. (Again, this is not expressible in a simple form.)

[Albert and Gupta (1985) actually consider a more general model than given here, and show how to approximate the Bayes solution. They apply their model to a problem of nonresponse in mail surveys.]

5.5 (a) Analogous to Problem 1.7.9, establish that for any random variable X, Y, and Z,

$$\text{cov}(X, Y) = E[\text{cov}(X, Y)|Z] + \text{cov}[E(X|Z), E(Y|Z)].$$

(b) For the hierarchy

$$\begin{aligned} X_i|\theta_i &\sim f(x|\theta_i), \quad i = 1, \dots, p, \text{ independent}, \\ \Theta_i|\lambda &\sim \pi(\theta_i|\lambda), \quad i = 1, \dots, p, \text{ independent}, \\ \Lambda &\sim \gamma(\lambda), \end{aligned}$$

show that $\text{cov}(\Theta_i, \Theta_j|\mathbf{x}) = \text{cov}[E(\Theta_i|\mathbf{x}, \lambda), E(\Theta_j|\mathbf{x}, \lambda)]$.

(c) If $E(\Theta_i|\mathbf{x}, \lambda) = g(x_i) + h(\lambda)$, $i = 1, \ldots, p$, where $g(\cdot)$ and $h(\cdot)$ are known, then

$$\operatorname{cov}(\Theta_i, \Theta_j|\mathbf{x}) = \operatorname{var}[E(\Theta_i|\mathbf{x}, \lambda)].$$

[Part (c) points to what can be considered a limitation in the applicability of some hierarchical models, that they imply a positive correlation structure in the posterior distribution.]

5.6 The one-way random effects model of Example 2.7 (see also Examples 3.5.1 and 3.5.5) can be written as the hierarchical model

$$X_{ij}|\mu, \alpha_i \sim N(\mu + \alpha_i, \sigma^2), \quad j = 1, \ldots, n, \quad i = 1, \ldots, s,$$
$$\alpha_i \sim N(0, \sigma_A^2), \quad i = 1, \ldots, s.$$

If, in addition, we specify that $\mu \sim \text{Uniform}(-\infty, \infty)$, show that the Bayes estimator of $\mu + \alpha_i$ under squared error loss is given by (3.5.13), the UMVU predictor of $\mu + \alpha_i$.

5.7 Referring to Example 6.6:

(a) Using the prior distribution for $\gamma(b)$ given in (5.6.27), show that the mode of the posterior distribution $\pi(b|\mathbf{x})$ is $\hat{b} = (p\bar{x} + \alpha - 1)/(pa + \beta - 1)$, and hence the empirical Bayes estimator based on this $\hat{b}$ does not equal the hierarchical Bayes estimator (5.6.29).

(b) Show that if we estimate $b/(b+1)$ using its posterior expectation $E[b/(b+1)|\mathbf{x}]$, then the resulting empirical Bayes estimator is equal to the hierarchical Bayes estimator.

5.8 The method of Monte Carlo integration allows the calculation of (possibly complicated) integrals by using (possibly simple) generations of random variables.

(a) To calculate $\int h(x) f_X(x)\,dx$, generate a sample $X_1, \ldots, X_m$, iid, from $f_X(x)$. Then, $1/m \sum_{i=1}^m h(x_i) \to \int h(x) f_X(x)\,dx$ as $m \to \infty$.

(b) If it is difficult to generate random variable from $f_X(x)$, then generate pairs of random variables

$$Y_i \sim f_Y(y),$$
$$X_i \sim f_{X|Y}(x|y_i).$$

Then, $1/m \sum_{i=1}^m h(x_i) \to \int h(x) f_X(x)\,dx$ as $m \to \infty$.
[Show that if X is generated according to $Y \sim f_Y(y)$ and $X \sim f_{X|Y}(x|Y)$, then $P(X \le a) = \int_{-\infty}^a f_x(x)\,dx$.]

(c) If it is difficult to generate as in part (b), then generate

$$X_{m_i} \sim f_{X|Y}(x|Y_{m_{i-1}}),$$
$$Y_{m_i} \sim f_{Y|X}(y|X_{m_i}).$$

for $i = 1, \ldots, K$ and $m = 1, \ldots, M$.

Show that:

(i) for each m, $\{X_{m_i}\}$ is a Markov chain. If it is also an ergodic Markov chain $X_{m_i} \xrightarrow{\mathcal{L}} X$, as $i \to \infty$, where X has the stationary distribution of the chain.

(ii) If the stationary distribution of the chain is $f_X(x)$, then

$$\frac{1}{M} \sum_{m=1}^{M} h(x_{m_k}) \to \int h(x) f_X(x)\,dx$$

as $K, M \to \infty$.

[This is the basic theory behind the Gibbs sampler. For each k, we have generated independent random variables X_{m_k}, $m = 1, \ldots, M$, where X_{m_k} is distributed according to $f_{X|Y}(x|y_{m_{k-}})$. It is also the case that for each m and large k, X_{m_k} is approximately distributed according to $f_X(x)$, although the variables are not now independent. The advantages and disadvantages of these computational schemes (*one-long-chain* vs. *many-short-chains*) are debated in Gelman and Rubin 1992; see also Geyer and Thompson 1992 and Smith and Roberts 1992. The prevailing consensus leans toward one long chain.]

5.9 To understand the convergence of the Gibbs sampler, let $(X, Y) \sim f(x, y)$, and define

$$k(x, x') = \int f_{X|Y}(x|y) f_{Y|X}(y|x')\, dy.$$

(a) Show that the function $h^*(\cdot)$ that solves $h^*(x) = \int k(x, x') h^*(x')\, dx'$ is $h^*(x) = f_X(x)$, the marginal distribution of X.
(b) Write down the analogous integral equation that is solved by $f_Y(y)$.
(c) Define a sequence of functions recursively by $h_{i+1}(x) = \int k(x, x') h_i(x')\, dx'_I$ where $h_0(x)$ is arbitrary but satisfies $\sup_x \left|\frac{h_0(x)}{h^*(x)}\right| < \infty$. Show that

$$\int |h_{i+1}(x) - h^*(x)| dx < \int |h_i(x) - h^*(x)|\, dx$$

and, hence, $h_i(x)$ converges to $h^*(x)$.

[The method of part (c) is called *successive substitution*. When there are two variables in the Gibbs sampler, it is equivalent to *data augmentation* (Tanner and Wong 1987). Even if the variables are vector-valued, the above results establish convergence. If the original vector of variables contains more than two variables, then a more general version of this argument is needed (Gelfand and Smith 1990).]

5.10 A direct Monte Carlo implementation of substitution sampling is provided by the *data augmentation* algorithm (Tanner and Wong 1987). If we define

$$h_{i+1}(x) = \int \left[\int f_{X|Y}(x|y) f_{Y|X}(y|x')\, dy\right] h_i(x')\, dx',$$

then from Problem 5.9, $h_i(x) \to f_x(x)$ as $i \to \infty$.

(a) To calculate h_{i+1} using Monte Carlo integration:

(i) Generate $X'_j \sim h_i(x')$, $j = 1, \ldots, J$.
(ii) Generate, for each x'_j, $Y_{jk} \sim f_{Y|X}(y|x'_j)$, $k = 1, \ldots, K$.
(iii) Calculate $\hat{h}_{i+1}(x) = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{K} \sum_{k=1}^{K} f_{X|Y}(x|y_{jk})$.

Then, $\hat{h}_{i+1}(x) \to h_{i+1}(x)$ as $J, K \to \infty$, and hence the data augmentation algorithm converges.

(b) To implement (a)(i), we must be able to generate a random variable from a mixture distribution. Show that if $f_Y(y) = \sum_{i=1}^{n} a_i g_i(y)$, $\Sigma a_i = 1$, then the algorithm

(i) Select g_i with probability a_i
(ii) Generate $Y \sim g_i$

produces a random variable with distribution f_Y. Hence, show how to implement step (a)(i) by generating random variables from $f_{X|Y}$. Tanner and Wong (1987) note that this algorithm will work even if $J = 1$, which yields the approximation

$\hat{h}_{i+1}(x) = \frac{1}{K}\sum_{k=1}^{K} f_{X|Y}(x|y_k)$, identical to the Gibbs sampler. The data augmentation algorithm can also be seen as an application of the process of *multiple imputation* (Rubin 1976, 1987, Little and Rubin 1987).

5.11 Successive substitution sampling can be implemented via the Gibbs sampler in the following way. From Problem 5.8(c), we want to calculate

$$h_M = \frac{1}{M}\sum_{m=1}^{M} k(x|x_{m_k}) = \frac{1}{M}\sum_{m=1}^{M}\int f_{X|Y}(x|y)f_{X|Y}(y|x_{m_k})\,dy.$$

(a) Show that $h_M(x) \to f_X(x)$ as $M \to \infty$.

(b) Given x_{m_k}, a Monte Carlo approximation to $h_M(x)$ is

$$\hat{h}_M(x) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{J}\sum_{j=1}^{J} f_{X|Y}(x|y_{k_j})$$

where $Y_{k_j} \sim f_{Y|X}(y|x_{m_k})$ and $\hat{h}_M(x) \to h_M(x)$ as $J \to \infty$.

(c) Hence, as $M, J \to \infty$, $\hat{h}_M(x) \to f_X(x)$.

[This is the Gibbs sampler, which is usually implemented with $J = 1$.]

5.12 For the situation of Example 5.6, show that

(a)
$$E\left(\frac{1}{M}\sum_{i=1}^{M}\Theta_i\right) = E\left(\frac{1}{M}\sum_{i=1}^{M}E(\Theta|\mathbf{x},\tau_i)\right),$$

(b)
$$\text{var}\left(\frac{1}{M}\sum_{i=1}^{M}\Theta_i\right) \geq \text{var}\left(\frac{1}{M}\sum_{i=1}^{M}E(\Theta|\mathbf{x},\tau_i)\right).$$

(c) Discuss when equality might hold in (b). Can you give an example?

5.13 Show that for the hierarchy (5.5.1), the posterior distributions $\pi(\theta|\mathbf{x})$ and $\pi(\lambda|\mathbf{x})$ satisfy

$$\pi(\theta|\mathbf{x}) = \int\left[\int \pi(\theta|\mathbf{x},\lambda)\pi(\lambda|\mathbf{x},\theta')\,d\lambda\right]\pi(\theta'|\mathbf{x})\,d\theta',$$
$$\pi(\lambda|\mathbf{x}) = \int\left[\int \pi(\lambda|\mathbf{x},\theta)\pi(\theta|\mathbf{x},\lambda')\,d\theta\right]\pi(\lambda'|\mathbf{x})\,d\lambda',$$

and, hence, are stationary points of the Markov chains in (5.5.13).

5.14 Starting from a uniform random variable $U \sim \text{Uniform}(0, 1)$, it is possible to construct many random variables through transformations.

(a) Show that $-\log U \sim \exp(1)$.

(b) Show that $-\sum_{i=1}^{n}\log U_i \sim \text{Gamma}(n, 1)$, where $U_1, \ldots, U_n$ are iid as $U(0, 1)$.

(c) Let $X \sim \text{Exp}(a, b)$. Write X as a function of U.

(d) Let $X \sim \text{Gamma}(n, \beta)$, n an integer. Write X as a function of $U_1, \ldots, U_n$, iid as $U(0, 1)$.

5.15 Starting with a $U(0, 1)$ random variable, the transformations of Problem 5.14 will not get us normal random variables, or gamma random variables with noninteger shape parameters. One way of doing this is to use the *Accept-Reject Algorithm* (Ripley 1987, Section 3.2), an algorithm for simulating $X \sim f(x)$:

(i) Generate $Y \sim g(y)$, $U \sim U(0, 1)$, independent.
(ii) Calculate $\rho(Y) = \frac{1}{M}\frac{f(Y)}{g(Y)}$ where $M = \sup_t f(t)/g(t)$.
(iii) If $U < \rho(Y)$, set $X = Y$, otherwise return to i).

(a) Show that the algorithm will generate $X \sim f(x)$.
(b) Starting with $Y \sim \exp(1)$, show how to generate $X \sim N(0, 1)$.
(c) Show how to generate a gamma random variable with a noninteger shape parameter.

5.16 Consider the normal hierarchical model

$$\begin{aligned} X|\theta_1 &\sim n(\theta_1, \sigma_1^2), \\ \theta_1|\theta_2 &\sim n(\theta_2|\sigma_2^2), \\ &\vdots \\ \theta_{k-1}|\theta_k &\sim n(\theta_k, \sigma_k^2) \end{aligned}$$

where $\sigma_1^2, i = 1, \dots, k$, are known.

(a) Show that the posterior distribution of θ_i $(1 \le i \le k-1)$ is

$$\pi(\theta_i|x, \theta_k) = N(\alpha_i x + (1 - \alpha_i)\theta_k, \tau_i^2)$$

where $\tau_i^2 = (\Sigma_1^i \sigma_j^2)(\Sigma_{i+1}^k \sigma_j^2)/\Sigma_1^k \sigma_j^2$ and $\alpha_i = \tau_i^2/\Sigma_1^i \sigma_j^2$.
(b) Find an expression for the Kullback-Leibler information$K[\pi(\theta_i|x, \theta_k), \pi(\theta_i|\theta_k)]$ and show that it is a decreasing function of i.

5.17 The original proof of Theorem 5.7 (Goel and DeGroot 1981) used *Rényi's entropy function* (Rényi 1961)

$$R_\alpha(f, g) = \frac{1}{\alpha - 1}\log \int f^\alpha(x) g^{1-\alpha}(x)\, d\mu(x),$$

where f and g are densities, μ is a dominating measure, and α is a constant, $\alpha \neq 1$.

(a) Show that $R_\alpha(f, g)$ satisfies $R_\alpha(f, g) > 0$ and $R_\alpha(f, f) = 0$.
(b) Show that Theorem 5.7 holds if $R_\alpha(f, g)$ is used instead of $K[f, g]$.
(c) Show that $\lim_{\alpha\to 1} R_\alpha(f, g) = K[f, g]$, and provide another proof of Theorem 5.7.

5.18 The Kullback-Leibler information, $K[f, g]$ (5.5.25), is not symmetric in f and g, and a modification, called the *divergence*, remedies this. Define $J[f, g]$, the divergence between f and g, to be $J[f, g] = K[f, g] + K[g, f]$. Show that, analogous to Theorem 5.7, $J[\pi(\lambda|x), \gamma(\lambda)] < J[\pi(\theta|x), \pi(\theta)]$.

5.19 Goel and DeGroot (1981) define a Bayesian analog of Fisher information [see (2.5.10)] as

$$\mathcal{I}[\pi(\theta|x)] = \int_\Omega \left[\frac{\frac{\partial}{\partial x}\pi(\theta|x)}{\pi(\theta|x)}\right]^2 d\theta,$$

the information that x has about the posterior distribution. As in Theorem 5.7, show that $\mathcal{I}[\pi(\lambda|x)] < \mathcal{I}[\pi(\theta|x)]$, again showing that the influence of λ is less than that of θ.

5.20 Each of m spores has a probability τ of germinating. Of the r spores that germinate, each has probability ω of bending in a particular direction. If s bends in the particular direction, a probability model to describe this process is the *bivariate binomial*, with mass function

$$f(r, s|\tau, \omega, m) = \binom{m}{r}\tau^r(1-\tau)^{m-r}\binom{r}{s}\omega^s(1-\omega)^{r-s}.$$

(a) Show that the Jeffreys prior is $\pi_J(\tau, \omega) = (1-\tau)^{-1/2}\omega^{-1/2}(1-\omega)^{-1/2}$.

(b) If τ is considered a nuisance parameter, the reference prior is

$$\pi_R(\tau, \omega) = \tau^{-1/2}(1-\tau)^{-1/2}\omega^{-1/2}(1-\omega)^{-1/2}.$$

Compare the posterior means $E(\omega|r, s, m)$ under both the Jeffreys and reference priors. Is one more appropriate?

(c) What is the effect of the different priors on the posterior variance?

[Priors for the bivariate binomial have been considered by Crowder and Sweeting (1989), Polson and Wasserman (1990), and Clark and Wasserman (1993), who propose a reference/Jeffreys trade-off prior.]

5.21 Let $\mathcal{F} = \{f(x|\theta); \theta \in \Omega\}$ be a family of probability densities. The Kullback-Leibler information for discrimination between two densities in $\mathcal{F}$ can be written

$$\psi(\theta_1, \theta_2) = \int f(x|\theta_1) \log\left[\frac{f(x|\theta_1)}{f(x|\theta_2)}\right] dx.$$

Recall that the gradient of ψ is $\nabla\psi = \{(\partial/\partial\theta_i)\psi\}$ and the Hessian is $\nabla\nabla\psi = \{(\partial^2/\partial\theta_i\partial\theta_j)\psi\}$.

(a) If integration and differentiation can be interchanged, show that

$$\nabla\psi(\theta, \theta) = 0 \quad \text{and} \quad \det[\nabla\nabla\psi(\theta, \theta)] = I(\theta),$$

where $I(\theta)$ is the Fisher information of $f(x|\theta)$.

(b) George and McCulloch (1993) argue that choosing $\pi(\theta) = (\det[\nabla\nabla\psi(\theta, \theta)])^{1/2}$ is an appealing least informative choice of priors. What justification can you give for this?

Section 6

6.1 For the model (3.3.1), show that $\delta^\lambda(x)|_{\lambda=\hat{\lambda}} = \delta^{\hat{\lambda}(x)}$, where the Bayes estimator $\delta^\lambda(x)$ minimizes $\int L[\theta, d(x)]\pi(\theta|x, \lambda)\, d\theta$ and the empirical Bayes estimator $\delta^{\hat{\lambda}(x)}$ minimizes $\int L[\theta, d(x)]\pi(\theta|x, \hat{\lambda})\, d\theta$.

6.2 This problem will investigate conditions under which an empirical Bayes estimator is a Bayes estimator. Expression (6.6.3) is a true posterior expected loss if $\pi(\theta|\mathbf{x}, \hat{\lambda}(\mathbf{x}))$ is a true posterior.

From the hierarchy

$$X|\theta \sim f(x|\theta),$$
$$\Theta|\lambda \sim \pi(\theta|\lambda),$$

define the joint distribution of $\mathbf{X}$ and Θ to be $(\mathbf{X}, \theta) \sim g(\mathbf{x}, \theta) = f(\mathbf{x}|\theta)\pi(\theta|\hat{\lambda}(\mathbf{x}))$, where $\pi(\theta|\hat{\lambda}(\mathbf{x}))$ is obtained by substituting $\hat{\lambda}(\mathbf{x})$ for λ in $\pi(\theta|\lambda)$.

(a) Show that, for this joint density, the formal Bayes estimator is equivalent to the empirical Bayes estimator from the hierarchical model.

(b) If $f(\cdot|\theta)$ and $\pi(\cdot|\lambda)$ are proper densities, then $\int g(\mathbf{x}, \theta)\, d\theta < \infty$. However, $\int\int g(\mathbf{x}, \theta)\, d\mathbf{x} d\theta$ need not be finite.

6.3 For the model (6.3.1), the Bayes estimator $\delta^\lambda(x)$ minimizes $\int L(\theta, d(x)) \times \pi(\theta|x, \lambda)\, d\theta$ and the empirical Bayes estimator, $\delta^{\hat{\lambda}}(x)$, minimizes $\int L(\theta, d(x))\pi(\theta|x, \hat{\lambda}(x))\, d\theta$. Show that $\delta^\lambda(x)|_{\lambda=\hat{\lambda}(x)} = \delta^{\hat{\lambda}}(x)$.

6.4 For the situation of Example 6.1:

(a) Show that

$$\int_{-\infty}^{\infty} e^{-n/2\sigma^2(\bar{x}-\theta)^2} e^{(-1/2)\theta^2/\tau^2} d\theta = \sqrt{2\pi}\left(\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)^{1/2} e^{(-n/2)\bar{x}^2/\sigma^2+n\tau^2}$$

and, hence, establish (6.6.4).

(b) Verify that the marginal MLE of $\sigma^2 + n\tau^2$ is $n\bar{x}^2$ and that the empirical Bayes estimator is given by (6.6.5).

6.5 Referring to Example 6.2:

(a) Show that the Bayes risk, $r(\pi, \delta^\pi)$, of the Bayes estimator (6.6.7) is given by

$$r(\pi, \delta^\pi) = kE[\text{var}(p_k|x_k)] = \frac{kab}{(a+b)(a+b+1)(a+b+n)}.$$

(b) Show that the Bayes risk of the unbiased estimator $\mathbf{X}/n = (X_1/n, \ldots, X_k/n)$ is given by

$$r(\pi, \mathbf{X}/n) = \frac{kab}{n(a+b+1)(a+b)}.$$

6.6 Extend Theorem 6.3 to the case of Theorem 3.2; that is, if $\mathbf{X}$ has density (3.3.7) and $\boldsymbol{\eta}$ has prior density $\pi(\boldsymbol{\eta}|\gamma)$, then the empirical Bayes estimator is

$$E\left(\sum \eta_i \frac{\partial T_i(\mathbf{x})}{\partial x_j}|\mathbf{x}, \hat{\gamma}(\mathbf{x})\right) = \frac{\partial}{\partial x_j}\log m(\mathbf{x}|\hat{\gamma}(\mathbf{x})) - \frac{\partial}{\partial x_j}\log h(\mathbf{x}),$$

where $m(\mathbf{x}|\gamma)$ is the marginal distribution of $\mathbf{X}$ and $\hat{\gamma}(\mathbf{x})$ is the marginal MLE of γ.

6.7 (a) For $p_\eta(\mathbf{x})$ of (1.5.2), show that for any prior distribution $\pi(\eta|\lambda)$ that is dependent on a hyperparameter λ, the empirical Bayes estimator is given by

$$E\left[\sum_{i=1}^{s} \eta_i \frac{\partial}{\partial x_j} T_i(\mathbf{x})|\mathbf{x}, \hat{\lambda}\right] = \frac{\partial}{\partial x_j}\log m_\pi(\mathbf{x}|\hat{\lambda}(\mathbf{x})) - \frac{\partial}{\partial x_j}\log h(\mathbf{x}).$$

where $m_\pi(\mathbf{x}) = \int p_\theta(\mathbf{x})\pi(\theta)\,d\theta$.

(b) If $\mathbf{X}$ has the distribution $p_\theta(\mathbf{x})$ of (1.5.1), show that a similar formulas holds, that is,

$$E(\mathcal{T}\eta(\theta)|\hat{\lambda}) = \nabla \log m_\pi(\mathbf{x}|\hat{\lambda}) - \nabla \log h(\mathbf{x}),$$

where $\mathcal{T} = \{\partial T_i/\partial x_j\}$ is the Jacobian of $\mathcal{T}$and ∇a is the gradient vector of a, that is, $\nabla a = \{\partial a/\partial x_i\}$.

6.8 For each of the following situations, write the empirical Bayes estimator of the natural parameter (under squared error loss) in the form (6.6.12), using the marginal likelihood estimator of the hyperparameter λ. Evaluate the expressions as far as possible.

(a) $X_i \sim N(0, \sigma_i^2)$, $i = 1, \ldots, p$, independent; $1/\sigma_i^2 \sim$ Exponential(λ).

(b) $X_i \sim N(\theta_i, 1)$, $i = 1, \ldots, p$, independent, $\theta_i \sim DE(0, \lambda)$.

6.9 Strawderman (1992) shows that the James-Stein estimator can be viewed as an empirical Bayes estimator in an arbitrary location family. Let $\mathbf{X}_{p\times 1} \sim f(x-\theta)$, with $E\mathbf{X} = \theta$ and var $\mathbf{X} = \sigma^2 I$. Let the prior be $\theta \sim f^{*n}$, the n-fold convolution of f with itself. [The *convolution* of f with itself is $f^{*2}(x) = \int f(x-y)f(y)dy$. The *n-fold convolution* is $f^{*n}(x) = \int f*(n-1)(x)(x-y)f(y)dy$.] Equivalently, let $U_i \sim f$, $i = 0, \cdots, n$, iid, $\theta = \sum_1^n U_i$, and $\mathbf{X} = U_0 + \theta$.

(a) Show that the Bayes rule against squared error loss is $\frac{n}{n+1}\mathbf{x}$. Note that n is a prior parameter.

(b) Show that $|\mathbf{X}|^2/(p\sigma^2)$ is an unbiased estimator of $n+1$, and hence that an empirical Bayes estimator of θ is given by $\delta^{EB} = [1 - (p\sigma^2/|\mathbf{x}|^2)]\mathbf{x}$.

6.10 Show for the hierarchy of Example 3.4, where σ^2 and τ^2 are known but μ is unknown, that:

(a) The empirical Bayes estimator of θ_i, based on the marginal MLE of θ_i, is $\frac{\tau^2}{\sigma^2+\tau^2}X_i + \frac{\sigma^2}{\sigma^2+\tau^2}\bar{X}$.

(b) The Bayes risk, under sum-of-squared-errors loss, of the empirical Bayes estimator from part (a) is

$$p\sigma^2 - \frac{2(p-1)^2\sigma^4}{p(\sigma^2+\tau^2)} + (p-1)\left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)^2 \sum_{i=1}^{p} E(X_i - \bar{X})^2.$$

(c) The minimum risk of the empirical Bayes estimator is attained when all θ_is are equal.

[*Hint*: Show that $\sum_{i=1}^{p} E[(X_i - \bar{X})^2] = \sum_{i=1}^{p}(\theta_i - \bar{\theta})^2 + (p-1)\sigma^2$.]

6.11 For $E(\Theta|\mathbf{x})$ of (5.5.8), show that as $\nu \to \infty$, $E(\Theta|x) \to [p/(p+\sigma^2)]\bar{x}$, the Bayes estimator under a $N(0,1)$ prior.

6.12 (a) Show that the empirical Bayes $\delta^{EB}(\bar{x}) = (1 - \sigma^2/\max\{\sigma^2, p\bar{x}^2\})\bar{x}$ of (6.6.5) has bounded mean squared error.

(b) Show that a variation of $\delta^{EB}(\bar{x})$, of part (a), $\delta^v(\bar{x}) = [1 - \sigma^2/(v + p\bar{x}^2)]\bar{x}$, also has bounded mean squared error.

(c) For $\sigma^2 = \tau^2 = 1$, plot the risk functions of the estimators of parts (a) and (b).

[Thompson (1968a, 1968b) investigated the mean squared error properties of estimators like those in part (b). Although such estimators have smaller mean squared error than $\bar{x}$ for small values of θ, they always have larger mean squared error for larger values of θ.]

6.13 (a) For the hierarchy (5.5.7), with $\sigma^2 = 1$ and $p = 10$, evaluate the Bayes risk $r(\pi, \delta^\pi)$ of the Bayes estimator (5.5.8) for $\nu = 2, 5$, and 10.

(b) Calculate the Bayes risk of the estimator δ^v of Problem 6.12(b). Find a value of v that yields a good approximation to the risk of the hierarchical Bayes estimator. Compare it to the Bayes risk of the empirical Bayes estimator of Problem 6.12(a).

6.14 Referring to Example 6.6, show that the empirical Bayes estimator is also a hierarchical Bayes estimator using the prior $\gamma(b) = 1/b$.

6.15 The Taylor series approximation to the estimator (5.5.8) is carried out in a number of steps. Show that:

(a) Using a first-order Taylor expansion around the point $\bar{x}$, we have

$$\begin{aligned}\frac{1}{(1+\theta^2/\nu)^{(\nu+1)/2}} &= \frac{1}{(1+\bar{x}^2/\nu)^{(\nu+1)/2}} \\ &\quad - \frac{\nu+1}{\nu}\frac{\bar{x}}{(1+\bar{x}^2/\nu)^{(\nu+3)/2}}(\theta - \bar{x}) + R(\theta - \bar{x})\end{aligned}$$

where the remainder, $R(\theta - \bar{x})$, satisfies $R(\theta - \bar{x})/(\theta - \bar{x})^2 \to 0$ as $\theta \to \bar{x}$.

(b) The remainder in part (a) also satisfies

$$\int_{-\infty}^{\infty} R(\theta - \bar{x}) e^{-\frac{p}{2\sigma^2}(\theta - \bar{x})^2} d\theta = O(1/p^{3/2}).$$

(c) The numerator and denominator of (5.5.8) can be written

$$\int_{-\infty}^{\infty} \frac{1}{(1+\theta^2/\nu)^{(\nu+1)/2}} e^{-\frac{p}{2\sigma^2}(\theta-\bar{x})^2} d\theta = \frac{\sqrt{2\pi\sigma^2/p}}{(1+\bar{x}^2/\nu)^{(\nu+1)/2}} + O\left(\frac{1}{p^{3/2}}\right)$$

and

$$\int_{-\infty}^{\infty} \frac{\theta}{(1+\theta^2/\nu)^{(\nu+1)/2}} e^{-\frac{p}{2\sigma^2}(\theta-\bar{x})^2} d\theta$$
$$= \frac{\sqrt{2\pi\sigma^2/p}}{(1+\bar{x}^2/\nu)^{(\nu+1)/2}} \left[1 - \frac{(\nu+1)/\nu}{(1+\bar{x}^2/\nu)}\right] \bar{x} + O\left(\frac{1}{p^{3/2}}\right),$$

which yields (5.6.32).

6.16 For the situation of Example 6.7:

(a) Calculate the values of the approximation (5.6.32) for the values of Table 6.2. Are there situations where the estimator (5.6.32) is clearly preferred over the empirical Bayes estimator (5.6.5) as an approximation to the hierarchical Bayes estimator (5.5.8)?

(b) Extend the argument of Problem 6.15 to calculate the next term in the expansion and, hence, obtain a more accurate approximation to the hierarchical Bayes estimator (5.5.8). For the values of Table 6.2, is this new approximation to (5.5.8) preferable to (5.6.5) and (5.6.32)?

6.17 (a) Show that if $b(\cdot)$ has a bounded second derivative, then

$$\int b(\lambda) e^{-nh(\lambda)} d\lambda = b(\hat{\lambda}) \sqrt{\frac{2\pi}{nh''(\hat{\lambda})}} e^{-nh(\hat{\lambda})} + O\left(\frac{1}{n^{3/2}}\right)$$

where $h(\hat{\lambda})$ is the unique minimum of $h(\lambda)$, $h''(\lambda) \neq 0$, and $nh(\hat{\lambda}) \to$ constant as $n \to \infty$.

[*Hint*: Expand both $b(\cdot)$ and $h(\cdot)$ in Taylor series around $\hat{\lambda}$, up to second-order terms. Then, do the term-by-term integration.]

This is the *Laplace approximation* for an integral. For refinements and other developments of this approximation in Bayesian inference, see Tierney and Kadane 1986, Tierney, Kass, and Kadane 1989, and Robert 1994a (Section 9.2.3).

(b) For the hierarchical model (5.5.1), the posterior mean can be approximated by

$$E(\Theta|x) = e^{-nh(\hat{\lambda})} \left[\frac{2\pi}{nh''(\hat{\lambda})}\right]^{1/2} E(\Theta|x, \hat{\lambda}) + O\left(\frac{1}{n^{3/2}}\right)$$

where $h = \frac{1}{n} \log \pi(\lambda|x)$ and $\hat{\lambda}$ is the mode of $\pi(\lambda|x)$, the posterior distribution of λ.

(c) If $\pi(\lambda|x)$ is the normal distribution with mean $\hat{\lambda}$ and variance $\sigma^2 = [-(\partial^2/\partial\lambda^2) \times \log \pi(\lambda|x)|_{\lambda=\hat{\lambda}}]^{-1}$, then $E(\Theta|x) = E(\Theta|x, \hat{\lambda}) + O(1/n^{3/2})$.

(d) Show that the situation in part (c) arises from the hierarchy

$$X_i|\theta_i \sim N(\theta_i, \sigma^2),$$
$$\theta_i|\lambda \sim N(\lambda, \tau^2),$$
$$\lambda \sim \text{Uniform}(-\infty, \infty).$$

6.18 (a) Apply the Laplace approximation (5.6.33) to obtain an approximation to the hierarchical Bayes estimator of Example 6.6.

(b) Compare the approximation from part (a) with the empirical Bayes estimator (5.6.24). Which is a better approximation to the hierarchical Bayes estimator?

6.19 Apply the Laplace approximation (5.6.33) to the hierarchy of Example 6.7 and show that the resulting approximation to the hierarchical Bayes estimator is given by (5.6.32).

6.20 (a) Verify (6.6.37), that under squared error loss

$$r(\pi, \delta) = r(\pi, \delta^\pi) + E(\delta - \delta^\pi)^2.$$

(b) For $X \sim$ binomial(p, n), $L(p, \delta) = (p - \delta)^2$, and $\pi = \{\pi : \pi = \text{beta}(a, b), a > 0, b > 0\}$, determine whether $\hat{p} = x/n$ or $\delta^0 = (a_0 + x)/(a_0 + b_0 + n)$ is more robust, according to (6.6.37).

(c) Is there an estimator of the form $(c + x)/(c + d + n)$ that you would consider more robust, in the sense of (6.6.37), than either estimator in part (b)?

[In part (b), for fixed n and (a_0, b_0), calculate the Bayes risk of $\hat{p}$ and δ^0 for a number of (a, b) pairs.]

6.21 (a) Establish (6.6.39) and (6.6.40) for the class of priors given by (6.6.38).

(b) Show that the Bayes estimator based on $\pi(\theta) \in \pi$ in (6.6.38), under squared error loss, is given by (6.6.41).

Section 7

7.1 For the situation of Example 7.1:

(a) The empirical Bayes estimator of $\boldsymbol{\theta}$, using an unbiased estimate of $\tau^2/(\sigma^2 + \tau^2)$, is

$$\delta^{\text{EB}} = \left(1 - \frac{(p-2)\sigma^2}{|\mathbf{x}|^2}\right)\mathbf{x},$$

the James-Stein estimator.

(b) The empirical Bayes estimator of $\boldsymbol{\theta}$, using the marginal MLE of $\tau^2/(\sigma^2 + \tau^2)$, is

$$\delta^{\text{EB}} = \left(1 - \frac{p\sigma^2}{|\mathbf{x}|^2}\right)^+ \mathbf{x},$$

which resembles the positive-part Stein estimator.

7.2 Establish Corollary 7.2. Be sure to verify that the conditions on $g(\mathbf{x})$ are sufficient to allow the integration-by-parts argument. [Stein (1973, 1981) develops these representations in the normal case.]

7.3 The derivation of an unbiased estimator of the risk (Corollary 7.2) can be extended to a more general model in the exponential family, the model of Corollary 3.3, where $\mathbf{X} = X_1, \ldots, X_p$ has the density

$$p_{\boldsymbol{\eta}}(\mathbf{x}) = e^{\sum_{i=1}^p \eta_i x_i - A(\boldsymbol{\eta})} h(\mathbf{x}).$$

(a) The Bayes estimator of $\boldsymbol{\eta}$, under squared error loss, is

$$E(\eta_i|\mathbf{x}) = \frac{\partial}{\partial x_i} \log m(\mathbf{x}) - \frac{\partial}{\partial x_i} \log h(\mathbf{x}).$$

Show that the risk of $E(\boldsymbol{\eta}|\mathbf{X})]$ has unbiased estimator

$$\sum_{i=1}^p \left[\frac{\partial^2}{\partial x_i^2} (\log h(\mathbf{x}) - 2 \log m(\mathbf{x})) + \left(\frac{\partial}{\partial x_i} \log m(\mathbf{x}) \right)^2 \right].$$

[*Hint*: Theorem 3.5 and Problem 3.4.]

(b) Show that the risk of the empirical Bayes estimator

$$E(\eta_i|\mathbf{x}, \hat{\lambda}) = \frac{\partial}{\partial x_i} \log m(\mathbf{x}|\hat{\lambda}(\mathbf{x})) - \frac{\partial}{\partial x_i} \log h(\mathbf{x}).$$

of Theorem 6.3 has unbiased estimator

$$\sum_{i=1}^{p} \left[\frac{\partial^2}{\partial x_i^2} \left(\log h(\mathbf{x}) - 2 \log m(\mathbf{x}|\hat{\lambda}(\mathbf{x})) \right) + \left(\frac{\partial}{\partial x_i} \log m(\mathbf{x}|\hat{\lambda}(\mathbf{x})) \right)^2 \right].$$

(c) Use the results of part (b) to derive an unbiased estimator of the risk of the positive-part Stein estimator of (7.7.10).

7.4 Verify (7.7.9), the expression for the Bayes risk of δ^{τ_0}. (Problem 3.12 may be helpful.)

7.5 A general version of the empirical Bayes estimator (7.7.3) is given by

$$\delta^c(\mathbf{x}) = \left(1 - \frac{c\sigma^2}{|\mathbf{x}|^2}\right) \mathbf{x},$$

where c is a positive constant.

(a) Use Corollary 7.2 to verify that

$$E_\theta |\boldsymbol{\theta} - \delta^c(\mathbf{X})|^2 = p\sigma^2 + c\sigma^4[c - 2(p-2)] E_\theta \frac{1}{|\mathbf{X}|^2}.$$

(b) Show that the Bayes risk, under $\Theta \sim N_p(0, \tau^2 I)$, is given by

$$r(\pi, \delta^c) = \sigma^2 \left[p + \frac{c\sigma^2}{\sigma^2 + \tau^2} \left(\frac{c}{p-2} - 2 \right) \right]$$

and is minimized by choosing $c = p - 2$.

7.6 For the model

$$\mathbf{X}|\boldsymbol{\theta} \sim N_p(\boldsymbol{\theta}, \sigma^2 I),$$
$$\boldsymbol{\theta}|\tau^2 \sim N_p(\mu, \tau^2 I):$$

Show that:

(a) The empirical Bayes estimator, using an unbiased estimator of $\tau^2/(\sigma^2 + \tau^2)$, is the Stein estimator

$$\delta_i^{\text{JS}}(\mathbf{x}) = \mu_i + \left(1 - \frac{(p-2)\sigma^2}{\Sigma(x_i - \mu_i)^2}\right)(x_i - \mu_i).$$

(b) If $p \geq 3$, the Bayes risk, under squared error loss, of δ^{JS} is $r(\tau, \delta^{\text{JS}}) = r(\tau, \delta^\tau) + 2\sigma^4/(\sigma^2 + \tau^2)$, where $r(\tau, \delta^\tau)$ is the Bayes risk of the Bayes estimator.

(c) If $p < 3$, the Bayes risk of δ^{JS} is infinite. [*Hint*: Show that if $Y \sim \chi_m^2$, $E(1/Y) < \infty \Longleftrightarrow m < 3$].

7.7 For the model

$$\mathbf{X}|\boldsymbol{\theta} \sim N_p(\boldsymbol{\theta}, \sigma^2 I),$$
$$\boldsymbol{\theta}|\tau^2 \sim N(\mu, \tau^2 I)$$

the Bayes risk of the ordinary Stein estimator

$$\delta_i(\mathbf{x}) = \mu_i + \left(1 - \frac{(p-2)\sigma^2}{\Sigma(x_i - \mu_i)^2}\right)(x_i - \mu_i)$$

is uniformly larger than its positive-part version

$$\delta_i^+(\mathbf{x}) = \mu_i + \left(1 - \frac{(p-2)\sigma^2}{\Sigma(x_i - \mu_i)^2}\right)^+ (x_i - \mu_i).$$

7.8 Theorem 7.5 holds in greater generality than just the normal distribution. Suppose $\mathbf{X}$ is distributed according to the multivariate version of the exponential family $p_\eta(x)$ of (33.7),

$$p_\eta(\mathbf{x}) = e^{\eta'\mathbf{x} - A(\eta)} h(\mathbf{x}), \quad -\infty < x_i < \infty,$$

and a multivariate conjugate prior distribution [generalizing (3.19)] is used.

(a) Show that $E(\mathbf{X}|\eta) = \nabla A(\eta)$.

(b) If $\mu = 0$ in the prior distribution (see 3.19), show that $r(\tau, \delta) \geq r(\tau, \delta^+)$, where $\delta(\mathbf{x}) = [1 - B(\mathbf{x})]\mathbf{x}$ and $\delta^+(\mathbf{x}) = [1 - B(\mathbf{x})]^+\mathbf{x}$.

(c) If $\mu \neq 0$, the estimator $\delta(\mathbf{x})$ would be modified to $\mu + \delta(\mathbf{x} - \mu)$. Establish a result similar to part (b) for this estimator.

[*Hint*: For part (b), the proof of Theorem 7.5, modified to use the Bayes estimator $E(\nabla A(\eta)|\mathbf{x}, k, \mu)$ as in (3.21), will work.]

7.9 (a) For the model (7.7.15), show that the marginal distribution of X_i is negative binomial$(a, 1/b + 1)$; that is,

$$P(X_i = x) = \binom{a + x - 1}{x} \left(\frac{b}{b+1}\right)^x \left(\frac{1}{b+1}\right)^a$$

with $EX_i = ab$ and var $X_i = ab(b + 1)$.

(b) If $X_1, \ldots, X_m$ are iid according to the negative binomial distribution in part (a), show that the conditional distribution of $X_j | \sum_1^m X_i$ is the *negative hypergeometric* distribution, given by

$$P\left(X_j = x | \sum_1^m X_i = t\right) = \frac{\binom{a+x-1}{x}\binom{(m-1)a + t - x - 1}{t - x}}{\binom{ma + t - 1}{t}}$$

with $EX_j = t/m$ and var $X_j = (m-1)t(ma + t)/m^2(ma + 1)$.

7.10 For the situation of Example 7.6:

(a) Show that the Bayes estimator under the loss $L_k(\lambda, \delta)$ of (7.7.16) is given by (7.7.17).

(b) Verify (7.7.19) and (7.7.20).

(c) Evaluate the Bayes risks $r(0, \delta^1)$ and $r(1, \delta^0)$. Which estimator, δ^0 or δ^1, is more robust?

7.11 For the situation of Example 7.6, evaluate the Bayes risk of the empirical Bayes estimator (7.7.20) for $k = 0$ and 1. What values of the unknown hyperparameter b are least and which are most favorable to the empirical Bayes estimator?

[*Hint*: Using the posterior expected loss (7.7.22) and Problem 7.9(b), the Bayes risk can be expressed as an expectation of a function of ΣX_i only. Further simplification seems unlikely.]

7.12 Consider a hierarchical Bayes estimator for the Poisson model (7.7.15) with loss (7.7.16). Using the distribution (5.6.27) for the hyperparameter b, show that the Bayes estimator is

$$\left(\frac{p\bar{x}+\alpha-k}{p\bar{x}+pa+\alpha+\beta-k}\right)(a+x_i-k).$$

[*Hint*: Show that the Bayes estimator is $E(\lambda^{1-k}|\mathbf{x})/E(\lambda^{-k}|\mathbf{x})$ and that

$$E(\lambda^r|\mathbf{x})=\frac{\Gamma(p\bar{x}+pa+\alpha+\beta)\Gamma(p\bar{x}+\alpha+r)}{\Gamma(p\bar{x}+\alpha)\Gamma(p\bar{x}+pa+\alpha+\beta+r)}\frac{\Gamma(a+x_i+r)}{\Gamma(a+x_i)}.\Bigg]$$

7.13 Prove the following: Two matrix results that are useful in calculating estimators from multivariate hierarchical models are

(a) For any vector a of the form $a=(I-\frac{1}{s}J)b$, $\mathbf{1}'a=\Sigma a_i=0$.

(b) If B is an idempotent matrix (that is, $B^2=I$) and a is a scalar, then

$$(I+aB)^{-1}=I-\frac{a}{1+a}B.$$

7.14 For the situation of Example 7.7:

(a) Show how to derive the empirical Bayes estimator δ^L of (7.7.28).

(b) Verify the Bayes risk of δ^L of (7.7.29).

For the situation of Example 7.8:

(c) Show how to derive the empirical Bayes estimator δ^{EB_2} of (7.7.33).

(d) Verify the Bayes risk of δ^{EB_2}, (7.7.34).

7.15 The empirical Bayes estimator (7.7.27) can also be derived as a hierarchical Bayes estimator. Consider the hierarchical model

$$\begin{aligned} X_{ij}|\xi_i &\sim N(\xi_i,\sigma^2), \quad j=1,\ldots,n, \quad i=1,\ldots,s,\\ \xi_i|\mu &\sim N(\mu,\tau^2), \quad i=1,\ldots,s,\\ \mu &\sim \text{Uniform}(-\infty,\infty) \end{aligned}$$

where σ^2 and τ^2 are known.

(a) Show that the Bayes estimator, with respect to squared error loss, is

$$E(\xi_i|\mathbf{x})=\frac{\sigma^2}{\sigma^2+n\tau^2}E(\mu|\mathbf{x})+\frac{n\tau^2}{\sigma^2+n\tau^2}\bar{x}_i$$

where $E(\mu|\mathbf{x})$ is the posterior mean of μ.

(b) Establish that $E(\mu|\mathbf{x})=\bar{x}=\Sigma x_{ij}/ns$. [This can be done by evaluating the expectation directly, or by showing that the posterior distribution of $\xi_i|\mathbf{x}$ is

$$\xi_i|\mathbf{x}\sim N\left[\frac{\sigma^2}{\sigma^2+n\tau^2}\bar{x}+\frac{n\tau^2}{\sigma^2+n\tau^2}\bar{x}_i,\ \frac{\sigma^2}{\sigma^2+n\tau^2}\left(n\tau^2+\frac{\sigma^2}{s}\right)\right].$$

Note that the ξ_i's are not independent a posteriori. In fact,

$$\boldsymbol{\xi}|\mathbf{x}\sim N_s\left(\frac{n\tau^2}{\sigma^2+n\tau^2}M,\ \frac{n\sigma^2\tau^2}{\sigma^2+n\tau^2}M\right),$$

where $M=I+(\sigma^2/n\tau^2)J$].

(c) Show that the empirical Bayes estimator (7.7.32) can also be derived as a hierarchical Bayes estimator, by appending the specification $(\alpha, \beta) \sim \text{Uniform}(\Re^2)$ [that is, $\pi(\alpha, \beta) = d\alpha\, d\beta$, $-\infty < \alpha, \beta < \infty$] to the hierarchy (7.7.30).

7.16 Generalization of model (7.7.23) to the case of unequal n_i is, perhaps, not as straightforward as one might expect. Consider the generalization

$$X_{ij}|\xi_i \sim N(\xi_i, \sigma^2), \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, s,$$
$$\xi_i|\mu \sim N(\mu, \tau_i^2), \quad i = 1, \ldots, s.$$

We also make the assumption that $\tau_i^2 = \tau^2/n_i$. Show that:

(a) The above model is equivalent to

$$\mathbf{Y} \sim N_s(\boldsymbol{\lambda}, \sigma^2 I),$$
$$\boldsymbol{\lambda} \sim N_s(\mathbf{Z}\mu, \tau^2 I)$$

where $Y_i = \sqrt{n_i}\bar{X}_i$, $\lambda_i = \sqrt{n_i}\xi_i$ and $\mathbf{z} = (\sqrt{n_1}, \ldots, \sqrt{n_s})'$.

(b) The Bayes estimator of ξ_i, using squared error loss, is

$$\frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}\bar{x}_i.$$

(c) The marginal distribution of Y_i is $Y_i \sim N_s(\mathbf{Z}\mu, (\sigma^2 + \tau^2)I)$, and an empirical Bayes estimator of ξ is

$$\delta_i^{EB} = \bar{x} + \left(1 - \frac{(s-3)\sigma^2}{\Sigma n_i(\bar{x}_i - \bar{x})^2}\right)(\bar{x}_i - \bar{x})$$

where $\bar{x}_i = \Sigma_j x_{ij}/n_i$ and $\bar{x} = \Sigma_i n_i \bar{x}_i / \Sigma_i n_i$.

[Without the assumption that $\tau_i^2 = \tau^2/n_i$, one cannot get a simple empirical Bayes estimator. If $\tau_i^2 = \tau^2$, the likelihood estimation can be used to get an estimate of τ^2 to be used in the empirical Bayes estimator. This is discussed by Morris (1983a).]

7.17 (Empirical Bayes estimation in a general case). A general version of the hierarchical models of Examples 7.7 and 7.8 is

$$\mathbf{X}|\boldsymbol{\xi} \sim N_s(\boldsymbol{\xi}, \sigma^2 I),$$
$$\boldsymbol{\xi}|\boldsymbol{\beta} \sim N_s(\mathbf{Z}\boldsymbol{\beta}, \tau^2 I)$$

where σ^2 and $\mathbf{Z}_{s\times r}$, of rank r, are known and τ^2 and $\boldsymbol{\beta}_{r\times 1}$ are unknown. Under this model show that:

(a) The Bayes estimator of ξ, under squared error loss, is

$$E(\boldsymbol{\xi}|\mathbf{x}, \boldsymbol{\beta}) = \frac{\sigma^2}{\sigma^2 + \tau^2} z\boldsymbol{\beta} + \frac{\tau^2}{\sigma^2 + \tau^2}\mathbf{x}.$$

(b) Marginally, the distribution of $\mathbf{X}|\boldsymbol{\beta}$ is $\mathbf{X}|\boldsymbol{\beta} \sim N_s(\mathbf{Z}\boldsymbol{\beta}, (\sigma^2 + \tau^2)I)$.

(c) Under the marginal distribution in part (b),

$$E[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}] = E\hat{\boldsymbol{\beta}} = \boldsymbol{\beta},$$
$$E\left[\frac{s-r-2}{|\mathbf{X} - \mathbf{Z}\hat{\boldsymbol{\beta}}|^2}\right] = \frac{1}{\sigma^2 + \tau^2},$$

and, hence, an empirical Bayes estimator of ξ is

$$\delta^{EB} = \mathbf{Z}\hat{\boldsymbol{\beta}} + \left(1 - \frac{(s-r-2)\sigma^2}{|\mathbf{x} - \mathbf{Z}\hat{\boldsymbol{\beta}}|^2}\right)(\mathbf{x} - \mathbf{Z}\hat{\boldsymbol{\beta}}).$$

(d) The Bayes risk of δ^{EB} is $r(\tau, \delta^\tau) + (r+2)\sigma^4/(\sigma^2+\tau^2)$, where $r(\tau, \delta^\tau)$ is the risk of the Bayes estimator.

7.18 (Hierarchical Bayes estimation in a general case.) In a manner similar to the previous problem, we can derive hierarchical Bayes estimators for the model

$$\begin{aligned}\mathbf{X}|\boldsymbol{\xi} &\sim N_s(\boldsymbol{\xi}, \sigma^2 I),\\ \boldsymbol{\xi}|\boldsymbol{\beta} &\sim N_s(\mathbf{Z}\boldsymbol{\beta}, \tau^2 I),\\ \boldsymbol{\beta} &\sim \text{Uniform}(\Re^r)\end{aligned}$$

where σ^2 and $\mathbf{Z}_{s\times r}$, of rank r, are known and τ^2 is unknown.

(a) The prior distribution of $\boldsymbol{\xi}$, unconditional on β, is proportional to

$$\pi(\boldsymbol{\xi}) = \int_{\Re^r} \pi(\boldsymbol{\xi}|\boldsymbol{\beta})\, d\boldsymbol{\beta} \propto e^{-\frac{1}{2}\frac{\xi'(I-H)\xi}{\tau^2}},$$

where $H = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ projects from $\Re^s$ to $\Re^r$.
[*Hint*: Establish that

$$\begin{aligned}(\boldsymbol{\xi} - \mathbf{Z}\boldsymbol{\beta})'(\boldsymbol{\xi} - \mathbf{Z}\boldsymbol{\beta}) &= \boldsymbol{\xi}'(I-H)\boldsymbol{\xi}\\ &\quad + [\boldsymbol{\beta} - (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\xi}]'\mathbf{Z}'\mathbf{Z}[\boldsymbol{\beta} - (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\xi}]\end{aligned}$$

to perform the integration on $\boldsymbol{\beta}$.]

(b) Show that

$$\boldsymbol{\xi}|\mathbf{x} \sim N_s\left(\frac{\tau^2}{\sigma^2+\tau^2}M, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}M\right)$$

where $M = I + (\sigma^2/\tau^2)H$, and hence that the Bayes estimator is given by

$$\frac{\sigma^2}{\sigma^2+\tau^2}H\mathbf{x} + \frac{\tau^2}{\sigma^2+\tau^2}\mathbf{x},$$

where $\mathbf{Z}\hat{\boldsymbol{\beta}} = H\mathbf{x}$.
[*Hint*: Establish that

$$\begin{aligned}&\frac{1}{\tau^2}\boldsymbol{\xi}'(I-H)\boldsymbol{\xi} + \frac{1}{\sigma^2}(\mathbf{x}-\xi)'(\mathbf{x}-\xi)\\ &= \frac{\sigma^2+\tau^2}{\sigma^2\tau^2}\left[\left(\boldsymbol{\xi} - \frac{\tau^2}{\sigma^2+\tau^2}M\mathbf{x}\right)' M^{-1}\left(\boldsymbol{\xi} - \frac{\tau^2}{\sigma^2+\tau^2}M\mathbf{x}\right)\right]\\ &\quad + \frac{1}{\sigma^2+\tau^2}\mathbf{x}'(I-H)\mathbf{x}\end{aligned}$$

where $M^{-1} = I - \frac{\sigma^2}{\sigma^2+\tau^2}H$.]

(c) Marginally, $\mathbf{X}'(I-H)\mathbf{X} \sim (\sigma^2+\tau^2)\chi^2_{s-r}$. This leads us to the empirical Bayes estimator

$$H\mathbf{x} + \left(1 - \frac{(s-r-2)\sigma^2}{\mathbf{x}'(I-H)\mathbf{x}}\right)(\mathbf{x} - H\mathbf{x})$$

which is equal to the empirical Bayes estimator of Problem 7.17(c).

[The model in this and the previous problem can be substantially generalized. For example, both $\sigma^2 I$ and $\tau^2 I$ can be replaced by full, positive definite matrices. At the cost of an increase in the complexity of the matrix calculations and the loss of simple answers, hierarchical and empirical Bayes estimators can be computed. The covariances, either scalar or matrix, can also be unknown, and inverted gamma (or inverted Wishart) prior distributions can be accommodated. Calculations can be implemented via the Gibbs sampler.

Note that these generalizations encompass the "unequal n_i" case (see Problem 7.16), but there are no simple solutions for this case. Many of these estimators also possess a minimax property, which will be discussed in Chapter 5.]

7.19 As noted by Morris (1983a), an analysis of variance-type hierarchical model, with unequal n_i, will yield closed-form empirical Bayes estimators if the prior variances are proportional to the sampling variances. Show that, for the model

$$\begin{aligned} X_{ij}|\xi_i &\sim N(\xi_i, \sigma^2), \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, s, \\ \boldsymbol{\xi}|\boldsymbol{\beta} &\sim N_s(\mathbf{Z}\boldsymbol{\beta}, \tau^2 D^{-1}) \end{aligned}$$

where σ^2 and $\mathbf{Z}_{s\times r}$, of full rank r, are known, τ^2 is unknown, and $D = \text{diag}(n_1, \ldots, n_s)$, an empirical Bayes estimator is given by

$$\delta^{\text{EB}} = \mathbf{Z}\hat{\boldsymbol{\beta}} + \left(1 - \frac{(s-r-2)\sigma^2}{(\bar{\mathbf{x}} - \mathbf{Z}\hat{\boldsymbol{\beta}})' D(\bar{\mathbf{x}} - \mathbf{Z}\hat{\boldsymbol{\beta}})}\right)(\bar{\mathbf{x}} - \mathbf{Z}\hat{\boldsymbol{\beta}})$$

with $\bar{\mathbf{x}}_i = \Sigma_j x_{ij}/n_i$, $\bar{\mathbf{x}} = \{\bar{x}_i\}$, and $\hat{\boldsymbol{\beta}} = (\mathbf{Z}'D\mathbf{Z})^{-1}\mathbf{Z}'D\bar{\mathbf{x}}$.

7.20 An entertaining (and unjustifiable) result which abuses a hierarchical Bayes calculation yields the following derivation of the James-Stein estimator. Let $X \sim N_p(\boldsymbol{\theta}, I)$ and $\boldsymbol{\theta}|\tau^2 \sim N_p(0, \tau^2 I)$.

(a) Verify that conditional on τ^2, the posterior and marginal distributions are given by

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{x}, \tau^2) &= N_p\left(\frac{\tau^2}{\tau^2+1}\mathbf{x}, \frac{\tau^2}{\tau^2+1}I\right), \\ m(\mathbf{x}|\tau^2) &= N_p[0, (\tau^2+1)I]. \end{aligned}$$

(b) Show that, taking $\pi(\tau^2) = 1, -1 < \tau^2 < \infty$, we have

$$\begin{aligned} &\iint_{\Re^p} \boldsymbol{\theta}\pi(\boldsymbol{\theta}|\mathbf{x}, \tau^2) m(\mathbf{x}|\tau^2)\, d\boldsymbol{\theta}\, d\tau^2 \\ &= \frac{\mathbf{x}}{(2\pi)^{p/2}(|\mathbf{x}|^2)^{p/2-1}}\left[\Gamma\left(\frac{p-2}{2}\right)2^{(p-2)/2} - \frac{\Gamma(p/2)2^{p/2}}{|\mathbf{x}|^2}\right] \end{aligned}$$

and

$$\begin{aligned} &\int\int_{\Re^p} \pi(\boldsymbol{\theta}|\mathbf{x}, \tau^2) m(\mathbf{x}|\tau^2) d\boldsymbol{\theta}\, d\tau^2 \\ &= \frac{1}{(2\pi)^{p/2}(|\mathbf{x}|^2)^{p/2-1}}\Gamma\left(\frac{p-2}{2}\right)2^{(p-2)/2} \end{aligned}$$

and hence

$$E(\boldsymbol{\theta}|\mathbf{x}) = \left(1 - \frac{p-2}{|\mathbf{x}|^2}\right)\mathbf{x}.$$

(c) Explain some implications of the result in part (b) and, why it cannot be true. [Try to reconcile it with (3.3.12).]

(d) Why are the calculations in part (b) unjustified?

a finite sample space (Gutmann 1982a; see also Brown 1981). Gutmann (1982b) also demonstrates a sequential context in which the Stein effect does not hold (see Problem 7.29). ‖

Finally, we look at the admissibility of linear estimators. There has always been interest in characterizing admissibility of linear estimators, partly due to the ease of computing and using linear estimators, and also due to a search for a converse to Karlin's theorem (Theorem 2.14) (which gives sufficient conditions for admissibility of linear estimators). Note that we are concerned with the admissibility of linear estimators in the class of all estimators, not just in the class of linear estimators. (This latter question was addressed by La Motte (1982).)

Example 7.25 Admissible linear estimators. Let $X \sim N_r(\boldsymbol{\theta}, I)$, and consider estimation of $\varphi'\boldsymbol{\theta}$, where $\varphi_{r\times 1}$ is a known vector, and $L(\varphi'\boldsymbol{\theta}, \delta) = (\varphi'\boldsymbol{\theta} - \delta)^2$. For $r = 1$, the results of Karlin (1958); see also Meeden and Ghosh (1977), show that ax is admissible if and only if $0 \le a \le \varphi$. This result was generalized by Cohen (1965a) to show that $a'x$ is admissible if and only if a is in the sphere:

$$\{a : (a - \varphi/2)'(a - \varphi/2) \le \varphi'\varphi/4\} \tag{7.31}$$

(see Problem 7.30). Note that the extension to known covariance matrix is straightforward, and (7.31) becomes an ellipse.

For the problem of estimating $\boldsymbol{\theta}$, the linear estimator $C\mathbf{x}$, where C is an $r \times r$ symmetric matrix, is admissible if and only if all of the eigenvalues of C are between 0 and 1, with at most two equal to 1 (Cohen 1966).

Necessary and sufficient conditions for admissibility of linear estimators have also been described for multivariate Poisson estimation (Brown and Farrell, 1985a, 1985b) and for estimation of the scale parameters in the multivariate gamma distribution (Farrell et al., 1989). This latter result also has application to the estimation of variance components in mixed models. ‖

8 Problems

Section 1

1.1 For the situation of Example 1.2:

(a) Plot the risk functions of $\delta^{1/4}$, $\delta^{1/2}$, and $\delta^{3/4}$ for $n = 5, 10, 25$.

(b) For each value of n in part (a), find the range of prior values of p for which each estimator is preferred.

(c) If an experimenter has no prior knowledge of p, which of $\delta^{1/4}$, $\delta^{1/2}$, and $\delta^{3/4}$ would you recommend? Justify your choice.

1.2 The principle of *gamma-minimaxity* [first used by Hodges and Lehmann (1952); see also Robbins 1964 and Solomon 1972a, 1972b)] is a Bayes/frequentist synthesis. An estimator δ^* is *gamma-minimax* if

$$\inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi, \delta) = \sup_{\pi\in\Gamma} r(\pi, \delta^*)$$

where Γ is a specified class of priors. Thus, the estimator δ^* minimizes the maximum Bayes risk over those priors in the class Γ. (If Γ = all priors, then δ^* would be minimax.)

(a) Show that if $\Gamma = \{\pi_0\}$, that is, Γ consists of one prior, then the Bayes estimator is Γ minimax.

(b) Show that if $\Gamma = \{\text{all priors}\}$, then the minimax estimator is Γ minimax.

(c) Find the Γ-minimax estimator among the three estimators of Example 1.2.

1.3 Classes of priors for Γ-minimax estimation have often been specified using moment restrictions.

(a) For $X \sim b(p, n)$, find the Γ-minimax estimator of p under squared error loss, with

$$\Gamma_\mu = \left\{\pi(p) : \pi(p) = \text{beta}(a, b),\ \mu = \frac{a}{a+b}\right\}$$

where μ is considered fixed and known.

(b) For $X \sim N(\theta, 1)$, find the Γ-minimax estimator of θ under squared error loss, with

$$\Gamma_{\mu,\tau} = \left\{\pi(\theta) : E(\theta) = \mu,\ \text{var}\ \theta = \tau^2\right\}$$

where μ and τ are fixed and known.

[*Hint*: In part (b), show that the Γ-minimax estimator is the Bayes estimator against a normal prior with the specified moments (Jackson et al. 1970; see Chen, Eichenhauer-Herrmann, and Lehn 1990 for a multivariate version). This somewhat nonrobust Γ-minimax estimator is characteristic of estimators derived from moment restrictions and shows why robust Bayesians tend to not use such classes. See Berger 1985, Section 4.7.6 for further discussion.]

1.4 (a) For the random effects model of Example 4.2.7 (see also Example 3.5.1), show that the restricted maximum likelihood (REML) likelihood of σ_A^2 and σ^2 is given by (4.2.13), which can be obtained by integrating the original likelihood against a uniform $(-\infty, \infty)$ prior for μ.

(b) For $n_i = n$ in

$$X_{ij} = \mu + A_i + u_{ij} \quad (j = 1, \ldots, n_i, \quad i = 1, \ldots, s)$$

calculate the expected value of the REML estimate of σ_A^2 and show that it is biased. Compare REML to the unbiased estimator of σ_A^2. Which do you prefer?

(Construction of REML-type marginal likelihoods, where some effects are integrated out against priors, becomes particularly useful in nonlinear and generalized linear models. See, for example, Searle et al. 1992, Section 9.4 and Chapter 10.)

1.5 Establishing the fact that (9.1) holds, so S^2 is conditionally biased, is based on a number of steps, some of which can be involved. Define $\phi(a, \mu, \sigma^2) = (1/\sigma^2)E_{\mu,\sigma^2}[S^2 \mid |\bar{x}|/s < a]$.

(a) Show that $\phi(a, \mu, \sigma^2)$ only depends on μ and σ^2 through μ/σ. Hence, without loss of generality, we can assume $\sigma = 1$.

(b) Use the fact that the density $f(s \mid |\bar{x}|/s < a, \mu)$ has monotone likelihood ratio to establish $\phi(a, \mu, 1) \geq \phi(a, 0, 1)$.

(c) Show that

$$\lim_{a\to\infty} \phi(a, 0, 1) = 1 \quad \text{and} \quad \lim_{a\to 0} \phi(a, 0, 1) = \frac{E_{0,1}S^3}{E_{0,1}S} = \frac{n}{n-1}.$$

(d) Combine parts (a), (b), and (c) to establish (19.1).

The next three problems explore conditional properties of estimators. A detailed development of this theory is found in Robinson (1979a, 1979b), who also explored the relationship between admissibility and conditional properties.

1.6 Suppose that $X \sim f(x|\theta)$, and $T(x)$ is used to estimate $\tau(\theta)$. One might question the worth of $T(x)$ if there were some set $A \in \mathcal{X}$ for which $T(x) > \tau(\theta)$ for $x \in A$ (or if the reverse inequality holds). This leads to the conditional principle of never using an estimator if there exists a set $A \in \mathcal{X}$ for which $E_\theta\{[T(X) - \tau(\theta)]I(X \in A)\} \geq 0 \quad \forall\theta$, with strict inequality for some θ (or if the equivalent statement holds with the inequality reversed). Show that if $T(x)$ is the posterior mean of $\tau(\theta)$ against a proper prior, where both the prior and $f(x|\theta)$ are continuous in θ, then no such A can exist. (If such an A exists, it is called a *semirelevant* set. Elimination of semirelevant sets is an extremely strong requirement. A weaker requirement, elimination of *relevant* sets, seems more appropriate.)

1.7 Show that if there exists a set $A \in \mathcal{X}$ and an $\varepsilon > 0$ for which $E_\theta\{[T(X) - \tau(\theta)]I(X \in A)\} > \varepsilon$, then $T(x)$ is inadmissible for estimating $\tau(\theta)$ under squared error loss. (A set A satisfying the this inequality is an example of a *relevant* set.)

[*Hint*: Consider the estimator $T(x) + \varepsilon I(x \in A)$]

1.8 To see why elimination of semirelevant sets is too strong a requirement, consider the estimation of θ based on observing $X \sim f(x - \theta)$. Show that for any constant a, the Pitman estimator X satisfies

$$E_\theta[(X - \theta)I(X < a)] \leq 0 \quad \forall\theta \quad \text{or} \quad E_\theta[(X - \theta)I(X > a)] \geq 0 \quad \forall\theta,$$

with strict inequality for some θ. Thus, there are semirelevant sets for the Pitman estimator, which is, by most accounts, a fine estimator.

1.9 In Example 1.7, let $\delta^*(X) = X/n$ with probability $1 - \varepsilon$ and $= 1/2$ with probability ε. Determine the risk function of δ^* and show that for $\varepsilon = 1/(n+1)$, its risk is constant and less than $\sup R(p, X/n)$.

1.10 Find the bias of the minimax estimator (1.11) and discuss its direction.

1.11 In Example 1.7,

(a) determine c_n and show that $c_n \to 0$ as $n \to \infty$,

(b) show that $R_n(1/2)/r_n \to 1$ as $n \to \infty$.

1.12 In Example 1.7, graph the risk functions of X/n and the minimax estimator (1.11) for $n = 1, 4, 9, 16$, and indicate the relative positions of the two graphs for large values of n.

1.13 (a) Find two points $0 < p_0 < p_1 < 1$ such that the estimator (1.11) for $n = 1$ is Bayes with respect to a distribution Λ for which $P_\Lambda(p = p_0) + P_\Lambda(p = p_1) = 1$.

(b) For $n = 1$, show that (1.11) is a minimax estimator of p even if it is known that $p_o \leq p \leq p_1$.

(c) In (b), find the values p_0 and p_1 for which $p_1 - p_0$ is as small as possible.

1.14 Evaluate (1.16) and show that its maximum is $1 - \alpha$.

1.15 Let $X = 1$ or 0 with probabilities p and q, respectively, and consider the estimation of p with loss $= 1$ when $|d - p| \geq 1/4$, and 0 otherwise. The most general randomized estimator is $\delta = U$ when $X = 0$, and $\delta = V$ when $X = 1$ where U and V are two random variables with known distributions.

(a) Evaluate the risk function and the maximum risk of δ when U and V are uniform on $(0, 1/2)$ and $(1/2, 1)$, respectively.

(b) Show that the estimator δ of (a) is minimax by considering the three values $p = 0$, $1/2, 1$.

[*Hint*: (b) The risk at $p = 0, 1/2, 1$ is, respectively, $P(U > 1/4)$, $1/2[P(U < 1/4) + P(V > 3/4)]$, and $P(V < 3/4)$].

1.16 Show that the problem of Example 1.8 remains invariant under the transformations

$$X' = n - X, \quad p' = 1 - p, \quad d' = 1 - d.$$

This illustrates that randomized equivariant estimators may have to be considered when $\bar{G}$ is not transitive.

1.17 Let r_Λ be given by (1.3). If $r_\Lambda = \infty$ for some Λ, show that any estimator δ has unbounded risk.

1.18 In Example 1.9, show that no linear estimator has constant risk.

1.19 Show that the risk function of (1.22) depends on p_1 and p_2 only through $p_1 + p_2$ and takes on its maximum when $p_1 + p_2 = 1$.

1.20 (a) In Example 1.9, determine the region in the (p_1, p_2) unit square in which (1.22) is better than the UMVU estimator of $p_2 - p_1$ for $m = n = 2, 8, 18$, and 32.

(b) Extend Problems 1.11 and 1.12 to Example 1.9.

1.21 In Example 1.14, show that $\bar{X}$ is minimax for the loss function $(d - \theta)^2/\sigma^2$ without any restrictions on σ.

1.22 (a) Verify (1.37).

(b) Show that equality holds in (1.39) if and only if $P(X_i = 0) + P(X_i = 1) = 1$.

1.23 In Example 1.16(b), show that for any $k > 0$, the estimator

$$\delta = \frac{\sqrt{n}}{1 + \sqrt{n}} \frac{1}{n} \sum_{i=1}^{n} X_i^k + \frac{1}{2(1 + \sqrt{n})}$$

is a Bayes estimator for the prior distribution Λ over $\mathcal{F}_0$ for which (1.36) was shown to be Bayes.

1.24 Let X_i $(i = 1, \ldots, n)$ and Y_j $(j = 1, \ldots, n)$ be independent with distributions F and G, respectively. If $F(1) - F(0) = G(1) - G(0) = 1$ but F and G are otherwise unknown, find a minimax estimator for $E(Y_j) - E(X_i)$ under squared error loss.

1.25 Let X_i $(i = 1, \ldots, n)$ be iid with unknown distribution F. Show that

$$\delta = \frac{\text{No. of } X_i \le 0}{\sqrt{n}} \cdot \frac{1}{1 + \sqrt{n}} + \frac{1}{2(1 + \sqrt{n})}$$

is minimax for estimating $F(0) = P(X_i \le 0)$ with squared error loss. [*Hint*: Consider the risk function of δ.]

1.26 Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independently distributed as $N(\xi, \sigma^2)$ and $N(\eta, \tau^2)$, respectively, and consider the problem of estimating $\Delta = \eta - \xi$ with squared error loss.

(a) If σ and τ are known, $\bar{Y} - \bar{X}$ is minimax.

(b) If σ and τ are restricted by $\sigma^2 \le A$ and $\tau^2 \le B$, respectively (A, B known and finite), $\bar{Y} - \bar{X}$ continues to be minimax.

1.27 In the linear model (3.4.4), show that $\Sigma a_i \hat{\xi}_i$ (in the notation of Theorem 3.4.4) is minimax for estimating $\theta = \Sigma a_i \xi_i$ with squared error loss, under the restriction $\sigma^2 \le M$. [*Hint*: Treat the problem in its canonical form.]

1.28 For the random variable X whose distribution is (1.42), show that x must satisfy the inequalities stated below (1.42).

1.29 Show that the estimator defined by (1.43)

(a) has constant risk,

(b) is Bayes with respect to the prior distribution specified by (1.44) and (1.45).

1.30 Show that for fixed X and n, (1.43) $\to$ (1.11) as $N \to \infty$.

1.31 Show that var($\bar{Y}$) given by (3.7.6) takes on its maximum value subject to (1.41) when all the a's are 0 or 1.

1.32 (a) If $R(p, \delta)$ is given by (1.49), show that $\sup R(p, \delta) \cdot 4(1+\sqrt{n})^2 \to 1$ as $n \to \infty$.

(b) Determine the smallest value of n for which the Bayes estimator of Example 1.18 satisfies (1.48) for $r = 1$ and $b = 5, 10$, and 20.

1.33 (Efron and Morris 1971)

(a) Show that the estimator δ of (1.50) is the estimator that minimizes $|\delta - c\bar{x}|$ subject to the constraint $|\delta - \bar{x}| \leq M$. In this sense, it is the estimator that is closest to a Bayes estimator, $c\bar{x}$, while not straying too far from a minimax estimator, $\bar{x}$.

(b) Show that for the situation of Example 1.19, $R(\theta, \delta)$ is bounded for δ of (1.50).

(c) For the situation of Example 1.19, δ of (1.50) satisfies $\sup_\theta R(\theta, \delta) = (1/n) + M^2$.

Section 2

2.1 Lemma 2.1 has been extended by Berger (1990a) to include the case where the estimand need not be restricted to a finite interval, but, instead, attains a maximum or minimum at a finite parameter value.

Lemma 8.1 *Let the estimand $g(\theta)$ be nonconstant with global maximum or minimum at a point $\theta^* \in \Omega$ for which $f(x|\theta^*) > 0$ a.e. (with respect to a dominating measure μ), and let the loss $L(\theta, d)$ satisfy the assumptions of Lemma 2.1. Then, any estimator δ taking values above the maximum of $g(\theta)$, or below the minimum, is inadmissible.*

(a) Show that if θ^* minimizes $g(\theta)$, and if $\hat{g}(x)$ is an unbiased estimator of $g(\theta)$, then there exists $\epsilon > 0$ such that the set $A_\epsilon = \{x \in \mathcal{X} : \hat{g}(x) < g(\theta) - \epsilon\}$ satisfies $P(A_\epsilon) > 0$. A similar conclusion holds if $g(\theta^*)$ is a maximum.

(b) Suppose $g(\theta^*)$ is a minimum. (The case of a maximum is handled similarly.) Show that the estimator

$$\delta(x) = \begin{cases} \hat{g}(x) & \text{if } \hat{g}(x) \geq g(\theta^*) \\ g(\theta^*) & \text{if } \hat{g}(x) < g(\theta^*) \end{cases}$$

satisfies $R(\delta, \theta) - R(\hat{g}(x), \theta) < 0$.

(c) For the situation of Example 2.3, apply Lemma 8.1 to establish the inadmissibility of the UMVU estimator of σ_A^2. Also, explain why the hypotheses of Lemma 8.1 are not satisfied for the estimation of σ^2

2.2 Determine the Bayes risk of the estimator (2.4) when θ has the prior distribution $N(\mu, \tau^2)$.

2.3 Prove part (d) in the second proof of Example 2.8, that there exists a sequence of values $\theta_i \to -\infty$ with $b'(\theta_i) \to 0$.

2.4 Show that an estimator $aX + b$ $(0 \leq a \leq 1)$ of $E_\theta(X)$ is inadmissible (with squared error loss) under each of the following conditions:

(a) if $E_\theta(X) \geq 0$ for all θ, and $b < 0$;

(b) if $E_\theta(X) \leq k$ for all θ, and $ak + b > k$.

[*Hint*: In (b), replace X by $X' = k - X$ and $aX + b$ by $k - (aX + b) = aX' + k - b - ak$, respectively and use (a).]

2.5 Show that an estimator $[1/(1+\lambda)+\varepsilon]X$ of $E_\theta(X)$ is inadmissible (with squared error loss) under each of the following conditions:

(a) if $\text{var}_\theta(X)/E_\theta^2(X) > \lambda > 0$ and $\varepsilon > 0$,
(b) if $\text{var}_\theta(X)/E_\theta^2(X) < \lambda$ and $\varepsilon < 0$.

[*Hint*: (a) Differentiate the risk function of the estimator with respect to ε to show that it decreases as ε decreases (Karlin 1958).]

2.6 Show that if $\text{var}_\theta(X)/E_\theta^2(X) > \lambda > 0$, an estimator $[1/(1+\lambda)+\varepsilon]X+b$ is inadmissible (with squared error loss) under each of the following conditions:

(a) if $E_\theta(X) > 0$ for all θ, $b > 0$ and $\varepsilon > 0$;
(b) if $E_\theta(X) < 0$ for all θ, $b < 0$ and $\varepsilon > 0$ (Gupta 1966).

2.7 Brown (1986a) points out a connection between the information inequality and the unbiased estimator of the risk of Stein-type estimators.

(a) Show that (2.7) implies

$$R(\theta, \delta) \geq \frac{[1 + b'(\theta)]^2}{n} + b^2(\theta) \geq \frac{1}{n} + \frac{2b'(\theta)}{n} + b^2(\theta)$$

and, hence, if $R(\theta, \delta) \leq R(\theta, \bar{X})$, then $\frac{2b'(\theta)}{n} + b^2 \leq 0$.

(b) Show that a nontrivial solution $b(\theta)$ would lead to an improved estimator $x - g(x)$, for $p = 1$, in Corollary 4.7.2.

2.8 A density function $f(x|\theta)$ is *variation reducing of order* $n + 1$ (VR_{n+1}) if, for any function $g(x)$ with k ($k \leq n$) sign changes (ignoring zeros), the expectation $E_\theta g(X) = \int g(x)f(x|\theta)\,dx$ has at most k sign changes. If $E_\theta g(X)$ has exactly k sign changes, they are in the same order.

Show that $f(x|\theta)$ is VR_2 if and only if it has monotone likelihood ratio. (See TSH2, Lemma 2, Section 3.3 for the "if" implication).

Brown et al. (1981) provide a thorough introduction to this topic, including VR characterizations of many families of distributions (the exponential family is VR_∞, as is the χ^2_ν with ν the parameter, and the noncentral $\chi^2_\nu(\lambda)$ in λ). There is an equivalence between VR_n and TP_n, Karlin's (1968) *total positivity of order* n, in that $VR_n = TP_n$.

2.9 For the situation of Example 2.9, show that:

(a) without loss of generality, the restriction $\theta \in [a, b]$ can be reduced to $\theta \in [-m, m]$, $m > 0$.

(b) If Λ is the prior distribution that puts mass $1/2$ on each of the points $\pm m$, then the Bayes estimator against squared error loss is

$$\delta^\Lambda(\bar{x}) = m\frac{e^{mn\bar{x}} - e^{-mn\bar{x}}}{e^{mn\bar{x}} + e^{-mn\bar{x}}} = m\tanh(mn\bar{x}).$$

(c) For $m < 1/\sqrt{n}$,

$$\max_{\theta \in [-m,m]} R(\theta, \delta(\bar{X})) = \max\left\{R(-m, \delta^\Lambda(\bar{X})),\, R(m, \delta^\Lambda(\bar{X}))\right\}$$

and hence, by Corollary 1.6, δ^Λ is minimax.

[*Hint*: Problem 2.8 can be used to show that the derivative of the risk function can have at most one sign change, from negative to positive, and hence any interior extrema can only be a minimum.]

(d) For $m > 1.05/\sqrt{n}$, δ^Λ of part (b) is no longer minimax. Explain why this is so and suggest an alternate estimator in this case.
[*Hint*: Consider $R(0, \delta^\Lambda)$].

2.10 For the situation of Example 2.10, show that:

(a) $\max_{\theta \in [-m,m]} R(\theta, a\bar{X} + b) = \max\{R(-m, a\bar{X} + b), R(m, a\bar{X} + b)\}$.

(b) The estimator $a^*\bar{X}$, with $a^* = m^2/(\frac{1}{n} + m^2)$, is the linear minimax estimator for all m with minimax risk a^*/n.

(c) $\bar{X}$ is the linear minimax estimator for $m = \infty$.

2.11 Suppose X has distribution F_ξ and Y has distribution G_η, where ξ and η vary independently. If it is known that $\eta = \eta_0$, then any estimator $\delta(X, Y)$ can be improved upon by

$$\delta^*(x) = E_Y\delta(x, Y) = \int \delta(x, y)\, dG_{\eta_0}(y).$$

[*Hint*: Recall the proof of Theorem 1.6.1.]

2.12 In Example 2.13, prove that the estimator $aY + b$ is inadmissible when $a > 1/(r+1)$.
[*Hint*: Problems 2.4–2.6]

2.13 Let $X_1, \ldots, X_n$ be iid according to a $N(0, \sigma^2)$ density, and let $S^2 = \sum X_i^2$. We are interested in estimating σ^2 under squared error loss using linear estimators $cS^2 + d$, where c and d are constants. Show that:

(a) admissibility of the estimator $aY + b$ in Example 2.13 is equivalent to the admissibility of $cS^2 + d$, for appropriately chosen c and d.

(b) the risk of $cS^2 + d$ is given by $R(cS^2 + d, \sigma^2) = 2nc^2\sigma^2 + [(nc - 1)\sigma^2 + d]^2$

(c) for $d = 0$, $R(cS^2, \sigma^2) < R(0, \sigma^2)$ when $c < 2/(n + 2)$, and hence the estimator $aY + b$ in Example 2.13 is inadmissible when $a = b = 0$.

[This exercise illustrates the fact that constants are not necessarily admissible estimators.]

2.14 For the situation of Example 2.15, let $Z = \bar{X}/S$.

(a) Show that the risk, under squared error loss, of $\delta = \varphi(z)s^2$ is minimized by taking

$$\varphi(z) = \varphi^*_{\mu,\sigma}(z) = E(S^2/\sigma^2|z)/E((S^2/\sigma^2)^2|z).$$

(b) Stein (1964) showed that $\varphi^*_{\mu,\sigma}(z) \le \varphi^*_{0,1}(z)$ for every μ, σ. Assuming this is so, deduce that $\varphi_s(Z)S^2$ dominates $[1/(n + 1)]S^2$ in squared error loss, where

$$\varphi_s(z) = \min\left\{\varphi^*_{0,1}(z), \frac{1}{n+1}\right\}.$$

(c) Show that $\varphi^*_{0,1}(z) = (1 + z^2)/(n + 2)$, and, hence, $\varphi_s(Z)S^2$ is given by (2.31).

(d) The best equivariant estimator of the form $\varphi(Z)S^2$ was derived by Brewster and Zidek (1974) and is given by

$$\varphi_{BZ}(z) = \frac{E(S^2|Z \le z)}{E(S^4|Z \le z)},$$

where the expectation is calculated assuming $\mu = 0$ and $\sigma = 1$. Show that $\varphi_{BZ}(Z)S^2$ is generalized Bayes against the prior

$$\pi(\mu, \sigma) = \frac{1}{\sigma} \int_0^\infty u^{-1/2}(1+u)^{-1} e^{-un\mu^2/\sigma^2}\, du\, d\mu\, d\sigma.$$

[Brewster and Zidek did not originally derive their estimator as a Bayes estimator, but rather first found the estimator and then found the prior. Brown (1968) considered a family of estimators similar to those of Stein (1964), which took different values depending on a cutoff point for z^2. Brewster and Zidek (1974) showed that the number of cutoff points can be arbitrarily large. They constructed a sequence of estimators, with decreasing risks and increasingly dense cutoffs, whose limit was the best equivalent estimator.]

2.15 Show the equivalence of the following relationships: (a) (2.26) and (2.27), (b) (2.34) and (2.35) when $c = \sqrt{(n-1)/(n+1)}$, and (c) (2.38) and (2.39).

2.16 In Example 2.17, show that the estimator $aX/n + b$ is inadmissible for all (a, b) outside the triangle (2.39).

[*Hint*: Problems 2.4–2.6.]

2.17 Prove admissibility of the estimators corresponding to the interior of the triangle (2.39), by applying Theorem 2.4 and using the results of Example 4.1.5.

2.18 Use Theorem 2.14 to provide an alternative proof for the admissibility of the estimator $a\bar{X} + b$ satisfying (2.6), in Example 2.5.

2.19 Determine which estimators $aX + b$ are admissible for estimating $E(X)$ in the following situations, for squared error loss:

(a) X has a Poisson distribution.

(b) X has a negative binomial distribution (Gupta 1966).

2.20 Let X have the Poisson(λ) distribution, and consider the estimation of λ under the loss $(d-\lambda)^2/\lambda$ with the restriction $0 \le \lambda \le m$, where m is known.

(a) Using an argument similar to that of Example 2.9, show that X is not minimax, and a least favorable prior distribution must have a set $w_\wedge$ [of (1.5)] consisting of a finite number of points.

(b) Let Λ_a be a prior distribution that puts mass $a_i, i = 1, \ldots, k$, at parameter points $b_i, i = 1, \ldots, k$. Show that the Bayes estimator associated with this prior is

$$\delta^{\Lambda_a}(x) = \frac{1}{E(\lambda^{-1}|x)} = \frac{\sum_{i=1}^k a_i b_i^x e^{-b_i}}{\sum_{i=1}^k a_i b_i^{x-1} e^{-b_i}}.$$

(c) Let m_0 be the solution to $m = e^{-m} (m_0 \approx .57)$. Show that for $0 \le \lambda \le m, m \le m_0$ a one-point prior $(a_i = 1, b_1 = m)$ yields the minimax estimator. Calculate the minimax risk and compare it to that of X.

(d) Let m_1 be the first positive zero of $(1 + \delta^\Lambda(m))^2 = 2 + m^2/2$, where Λ is a two-point prior $(a_1 = a, b_1 = 0; a_2 = 1-a, b_2 = m)$. Show that for $0 \le \lambda \le m, m_0 < m \le m_1$, a two-point prior yields the minimax estimator (use Corollary 1.6). Calculate the minimax risk and compare it to that of X.

[As m increases, the situation becomes more complex and exact minimax solutions become intractable. For these cases, linear approximations can be quite satisfactory. See Johnstone and MacGibbon 1992, 1993.]

2.21 Show that the conditions (2.41) and (2.42) of Example 2.22 are not only sufficient but also necessary for admissibility of (2.40).

2.22 Let X and Y be independently distributed according to Poisson distributions with $E(X) = \xi$ and $E(Y) = \eta$, respectively. Show that $aX + bY + c$ is admissible for estimating ξ with squared error loss if and only if either $0 \le a < 1, b \ge 0, c \ge 0$ or $a = 1, b = c = 0$ (Makani 1972).

2.23 Let X be distributed with density $\frac{1}{2}\beta(\theta)e^{\theta x}e^{-|x|}$, $|\theta| < 1$.

(a) Show that $\beta(\theta) = 1 - \theta^2$.

(b) Show that $aX + b$ is admissible for estimating $E_\theta(X)$ with squared error loss if and only if $0 \le a \le 1/2$.

[*Hint*: (b) To see necessity, let $\delta = (1/2 + \varepsilon)X + b$ $(0 < \varepsilon \le 1/2)$ and show that δ is dominated by $\delta' = (1 - \frac{1}{2}\alpha + \alpha\varepsilon)X + (b/\alpha)$ for some α with $0 < \alpha < 1/(1/2 - \varepsilon)$.]

2.24 Let X be distributed as $N(\theta, 1)$ and let θ have the improper prior density $\pi(\theta) = e^\theta$ $(-\infty < \theta < \infty)$. For squared error loss, the formal Bayes estimator of θ is $X + 1$, which is neither minimax nor admissible. (See also Problem 2.15.)

Conditions under which the formal Bayes estimator corresponding to an improper prior distribution for θ in Example 3.4 is admissible are given by Zidek (1970).

2.25 Show that the natural parameter space of the family (2.16) is $(-\infty, \infty)$ for the normal (variance known), binomial, and Poisson distribution but not in the gamma or negative binomial case.

Section 3

3.1 Show that Theorem 3.2.7 remains valid for almost equivariant estimators.

3.2 Verify the density (3.1).

3.3 In Example 3.3, show that a loss function remains invariant under G if and only if it is a function of $(d - \theta)^*$.

3.4 In Example 3.3, show that neither of the loss functions $[(d - \theta)^{**}]^2$ or $|(d - \theta)^{**}|$ is convex.

3.5 Let Y be distributed as $G(y - \eta)$. If $T = [Y]$ and $X = Y - T$, find the distribution of X and show that it depends on η only through $\eta - [\eta]$.

3.6 (a) If $X_1, \ldots, X_n$ are iid with density $f(x - \theta)$, show that the MRE estimator against squared error loss [the Pitman estimator of (3.1.28)] is the Bayes estimator against right-invariant Haar measure.

(b) If $X_1, \ldots, X_n$ are iid with density $1/\tau f[(x - \mu)/\tau]$, show that:

(i) Under squared error loss, the Pitman estimator of (3.1.28) is the Bayes estimator against right-invariant Haar measure.

(ii) Under the loss (3.3.17), the Pitman estimator of (3.3.19) is the Bayes estimator against right-invariant Haar measure.

3.7 Prove formula (3.9).

3.8 Prove (3.11).

[*Hint*: In the term on the left side, lim inf can be replaced by lim. Let the left side of (3.11) be A and the right side B, and let $A_N = \inf h(a, b)$, where the inf is taken over $a \le -N, b \ge N, N = 1, 2, \ldots$, so that $A_N \to A$. There exist (a_N, b_N) such that $|h(a_N, b_N) - A_N| \le 1/N$. Then, $h(a_N, b_N) \to A$ and $A \ge B$.]

3.9 In Example 3.8, let $h(\theta)$ be the length of the path θ after cancellation. Show that h does not satisfy conditions (3.2.11).

3.10 Discuss Example 3.8 for the case that the random walk instead of being in the plane is (a) on the line and (b) in three-space.

3.11 (a) Show that the probabilities (3.17) add up to 1.

(b) With p_k given by (3.17), show that the risk (3.16) is infinite.

[*Hint*: (a) $1/k(k+1) = (1/k) - 1/(k+1)$.]

3.12 Show that the risk $R(\theta, \delta)$ of (3.18) is finite.

[*Hint*: $R(\theta, \delta) < \Sigma_{k>M|k+\theta|} 1/(k+1) \le \Sigma_{c<k<d} 1/(k+1) < \int_c^{d+1} dx/x$, where $c = M|\theta|/(M+1)$ and $d = M|\theta|/(M-1)$. The reason for the second inequality is that values of k outside (c, d) make no contribution to the sum.]

3.13 Show that the two estimators δ^* and δ^{**}, defined by (3.20) and (3.21), respectively, are equivariant.

3.14 Prove the relations (3.22) and (3.23).

3.15 Let the distribution of X depend on parameters θ and ϑ, let the risk function of an estimator $\delta = \delta(x)$ of θ be $R(\theta, \vartheta; \delta)$, and let $r(\theta, \delta) = \int R(\theta, \vartheta; \delta)\, dP(\vartheta)$ for some distribution P. If δ_0 minimizes $\sup_\theta r(\theta, \delta)$ and satisfies $\sup_\theta r(\theta, \delta_0) = \sup_{\theta,\vartheta} R(\theta, \vartheta; \delta_0)$, show that δ_0 minimizes $\sup_{\theta,\vartheta} R(\theta, \vartheta; \delta)$.

Section 4

4.1 In Example 4.2, show that an estimator δ is equivariant if and only if it satisfies (4.11) and (4.12).

4.2 Show that a function μ satisfies (4.12) if and only if it depends only on ΣX_i^2.

4.3 Verify the Bayes estimator (4.15).

4.4 Let X_i be independent with binomial distribution $b(p_i, n_i)$, $i = 1, \ldots, r$. For estimating $p = (p_1, \ldots, p_r)$ with average squared error loss (4.17), find the minimax estimator of p, and determine whether it is admissible.

4.5 Establishing the admissibility of the normal mean in two dimensions is quite difficult, made so by the fact that the conjugate priors fail in the limiting Bayes method. Let

$$X \sim N_2(\boldsymbol{\theta}, I) \quad \text{and} \quad L(\boldsymbol{\theta}, \delta) = |\boldsymbol{\theta} - \delta|^2.$$

The conjugate priors are $\boldsymbol{\theta} \sim N_2(0, \tau^2 I)$, $\tau^2 > 0$.

(a) For this sequence of priors, verify that the limiting Bayes argument, as in Example 2.8, results in inequality (4.18), which does not establish admissibility.

(b) Stein (in James and Stein 1961), proposed the sequence of priors that works to prove X is admissible by the limiting Bayes method. A version of these priors, given by Brown and Hwang (1982), is

$$g_n(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } |\boldsymbol{\theta}| \le 1 \\ 1 - \dfrac{\log|\boldsymbol{\theta}|}{\log n} & \text{if } 1 \le |\boldsymbol{\theta}| \le n \\ 0 & \text{if } |\boldsymbol{\theta}| \ge n \end{cases}$$

for $n = 2, 3, \ldots$. Show that $\delta^{g_n}(x) \to x$ a.e. as $n \to \infty$.

(c) A special case of the very general results of Brown and Hwang (1982) state that for the prior $\pi_n(\boldsymbol{\theta})^2 g(\boldsymbol{\theta})$, the limiting Bayes method (Blyth's method) will establish the admissibility of the estimator $\delta^g(x)$ [the generalized Bayes estimator against $g(\boldsymbol{\theta})$] if

$$\int_{\{\boldsymbol{\theta}:|\boldsymbol{\theta}|>1\}} \frac{g(\boldsymbol{\theta})\,d\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2[\max\{\log|\boldsymbol{\theta}|, \log 2\}]^2} < \infty.$$

Show that this holds for $g(\boldsymbol{\theta}) = 1$ and that $\delta^g(x) = x$, so x is admissible.

[Stein (1956b) originally established the admissibility of X in two dimensions using an argument based on the information inequality. His proof was complicated by the fact that he needed some additional invariance arguments to establish the result. See Theorem 7.19 and Problem 7.19 for more general statements of the Brown/Hwang result.]

4.6 Let $X_1,\ X_2, \ldots, X_r$ be independent with $X_i \sim N(\theta_i, 1)$. The following heuristic argument, due to Stein (1956b), suggests that it should be possible, at least for large r and hence large $|\theta|$, to improve on the estimator $\mathbf{X} = (X_1, X_2, \ldots, X_r)$.

(a) Use a Taylor series argument to show

$$|\mathbf{x}|^2 = r + |\boldsymbol{\theta}|^2 + O_p[(r + |\boldsymbol{\theta}|^2)^{1/2}],$$

so, with high probability, the true $\boldsymbol{\theta}$ is in the sphere $\{\boldsymbol{\theta} : |\boldsymbol{\theta}|^2 \le |\mathbf{x}|^2 - r\}$. The usual estimator $\mathbf{X}$ is approximately the same size as $\boldsymbol{\theta}$ and will almost certainly be outside of this sphere.

(b) Part (a) suggested to Stein an estimator of the form $\delta(\mathbf{x}) = [1 - h(|\mathbf{x}|^2)]\mathbf{x}$. Show that

$$|\boldsymbol{\theta} - \delta(\mathbf{x})|^2 = (1-h)^2|\mathbf{x} - \boldsymbol{\theta}|^2 - 2h(1-h)\boldsymbol{\theta}'(\mathbf{x} - \boldsymbol{\theta}) + h^2|\boldsymbol{\theta}|^2.$$

(c) Establish that $\boldsymbol{\theta}'(\mathbf{x} - \boldsymbol{\theta})/|\boldsymbol{\theta}| = Z \sim N(0, 1)$, and $|\mathbf{x} - \boldsymbol{\theta}|^2 \approx r$, and, hence,

$$|\boldsymbol{\theta} - \delta(\mathbf{x})|^2 \approx (1-h)^2 r + h^2|\boldsymbol{\theta}|^2 + O_p[(r + |\boldsymbol{\theta}|^2)^{1/2}].$$

(d) Show that the leading term in part (c) is minimized at $h = r/(r + |\boldsymbol{\theta}|^2)$, and since $|\mathbf{x}|^2 \approx r + |\boldsymbol{\theta}|^2$, this leads to the estimator $\delta(\mathbf{x}) = \left(1 - \frac{r}{|\mathbf{x}|^2}\right)\mathbf{x}$ of (4.20).

4.7 If S^2 is distributed as χ^2_r, use (2.2.5) to show that $E(S^{-2}) = 1/(r-2)$.

4.8 In Example 4.7, show that $\mathcal{R}$ is nonsingular for ρ_1 and ρ_2 and singular for ρ_3 and ρ_4.

4.9 Show that the function ρ_2 of Example 4.7 is convex.

4.10 In Example 4.7, show that X is admissible for (a) ρ_3 and (b) ρ_4.

[*Hint*: (a) It is enough to show that X_1 is admissible for estimating θ_1 with loss $(d_1 - \theta_1)^2$. This can be shown by letting $\theta_2, \ldots, \theta_r$ be known. (b) Note that X is admissible minimax for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$ when $\theta_1 = \cdots = \theta_r$.]

4.11 In Example 4.8, show that $\mathbf{X}$ is admissible under the assumptions (ii)(a).

[*Hint*:

i. If $v(t) > 0$ is such that

$$\int \frac{1}{v(t)} e^{-t^2/2\tau^2}\,dt < \infty,$$

show that there exists a constant $k(\tau)$ for which

$$\lambda_\tau(\boldsymbol{\theta}) = k(\tau)\left[\Sigma v(\theta_j)\right] \exp\left(-\frac{1}{2\tau^2}\Sigma\theta_j^2\right) / \Pi v(\theta_j)$$

is a probability density for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)$.

ii. If the X_i are independent $N(\theta_i, 1)$ and $\boldsymbol{\theta}$ has the prior $\lambda_\tau(\boldsymbol{\theta})$, the Bayes estimator of $\boldsymbol{\theta}$ with loss function (4.27) is $\tau^2 X/(1+\tau^2)$.

iii. To prove X admissible, use (4.18) with $\lambda_\tau(\boldsymbol{\theta})$ instead of a normal prior.]

4.12 Let $\mathcal{L}$ be a family of loss functions and suppose there exists $L_0 \in \mathcal{L}$ and a minimax estimator δ_0 with respect to L_0 such that in the notation of (4.29),

$$\sup_{L,\theta} R_L(\theta, \delta_0) = \sup_{\theta} R_{L_0}(\theta, \delta_0).$$

Then, δ_0 is minimax with respect to $\mathcal{L}$; that is, it minimizes $\sup_{L,\theta} R_L(\theta, \delta)$.

4.13 Assuming (4.25), show that $E = 1 - [(r-2)^2/r|\mathbf{X} - \mu|^2]$ is the unique unbiased estimator of the risk (5..4.25), and that E is inadmissible. [The estimator E is also unbiased for estimation of the loss $L(\boldsymbol{\theta}, \delta)$. See Note 9.5.]

4.14 A natural extension of risk domination under a particular loss is to risk domination under a class of losses. Hwang (1985) defines *universal domination* of δ by δ' if the inequality

$$E_\theta L(|\boldsymbol{\theta} - \delta'(\mathbf{X})|) \leq E_\theta L(|\boldsymbol{\theta} - \delta(\mathbf{X})|) \text{ for all } \boldsymbol{\theta}$$

holds for all loss functions $L(\cdot)$ that are nondecreasing, with at least one loss function producing nonidentical risks.

(a) Show that δ' universally dominates δ if and only if it *stochastically dominates* δ, that is, if and only if

$$P_\theta(|\boldsymbol{\theta} - \delta'(\mathbf{X})| > k) \leq P_\theta(|\boldsymbol{\theta} - \delta(\mathbf{X})| > k)$$

for all k and $\boldsymbol{\theta}$ with strict inequality for some $\boldsymbol{\theta}$.

[*Hint*: For a positive random variable Y, recall that $EY = \int_0^\infty P(Y > t)\, dt$. Alternatively, use the fact that stochastic ordering on random variables induces an ordering on expectations. See Lemma 1, Section 3.3 of TSH2.]

(b) For $X \sim N_r(\boldsymbol{\theta}, I)$, show that the James-Stein estimator $\delta^c(\mathbf{x}) = (1 - c/|\mathbf{x}|^2)\mathbf{x}$ does not universally dominate $\mathbf{x}$. [From (a), it only need be shown that $P_\theta(|\boldsymbol{\theta} - \delta^c(\mathbf{X})| > k) > P_\theta(|\boldsymbol{\theta} - \mathbf{X}| > k)$ for some $\boldsymbol{\theta}$ and k. Take $\boldsymbol{\theta} = 0$ and find such a k.]

Hwang (1985) and Brown and Hwang (1989) explore many facets of universal domination. Hwang (1985) shows that even δ^+ does not universally dominate X unless the class of loss functions is restricted.

We also note that although the inequality in part (a) may seem reminiscent of the "Pitman closeness" criterion, there is really no relation. The criterion of Pitman closeness suffers from a number of defects not shared by stochastic domination (see Robert et al. 1993).

Section 5

5.1 Show that the estimator δ_c defined by (5.2) with $0 < c = 1 - \Delta < 1$ is dominated by any δ_d with $|d - 1| < \Delta$.

5.2 In the context of Theorem 5.1, show that

$$E_\theta\left[\frac{1}{|\mathbf{X}|^2}\right] \leq E_0\left[\frac{1}{|\mathbf{X}|^2}\right] < \infty.$$

[*Hint*: The chi-squared distribution has monotone likelihood ratio in the noncentrality parameter.]

5.3 Stigler (1990) presents an interesting explanation of the Stein phenomenon using a regression perspective, and also gives an identity that can be used to prove the minimaxity of the James-Stein estimator. For $\mathbf{X} \sim (N_r\boldsymbol{\theta}, I)$, and $\delta^c(\mathbf{x}) = \left(1 - \frac{c}{|\mathbf{x}|^2}\right)\mathbf{x}$:

(a) Show that

$$E_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \delta^c(\mathbf{X})|^2 = r - 2cE_{\boldsymbol{\theta}}\left[\frac{\mathbf{X}'\boldsymbol{\theta} + (c/2)}{|\mathbf{X}|^2} - 1\right].$$

(b) The expression in square brackets is increasing in c. Prove the minimaxity of δ^c for $0 \le c \le 2(r-2)$ by establishing Stigler's identity

$$E_{\boldsymbol{\theta}}\left[\frac{\mathbf{X}'\boldsymbol{\theta} + r - 2}{|\mathbf{X}|^2}\right] = 1.$$

[*Hint*: Part (b) can be established by transforming to polar coordinates and directly integrating, or by writing $\frac{\mathbf{x}'\boldsymbol{\theta}}{|\mathbf{x}|^2} = \frac{\mathbf{x}'(\boldsymbol{\theta} - \mathbf{x}) + |\mathbf{x}|^2}{|\mathbf{x}|^2}$ and using Stein's identity.]

5.4 (a) Prove Theorem 5.5.

(b) Apply Theorem 5.5 to establish conditions for minimaxity of Strawderman's (1971) proper Bayes estimator given by (5.10) and (5.12).

[*Hint*: (a) Use the representation of the risk given in (5.4), with $g(\mathbf{x}) = c(|\mathbf{x}|)(r-2)\mathbf{x}/|\mathbf{x}|^2$. Show that $R(\boldsymbol{\theta}, \delta)$ can be written

$$R(\boldsymbol{\theta}, \delta) = 1 - \frac{(r-2)^2}{r}E_\theta\left[\frac{c(|\mathbf{X}|)(2 - c(|\mathbf{X}|))}{|\mathbf{X}|^2}\right] - \frac{2(r-2)}{r}E_\theta\frac{\Sigma X_i \frac{\partial}{\partial X_i}c(|\mathbf{X}|)}{|\mathbf{X}|^2}$$

and an upper bound on $R(\boldsymbol{\theta}, \delta)$ is obtained by dropping the last term. It is not necessary to assume that $c(\cdot)$ is differentiable everywhere; it can be nondifferentiable on a set of Lebesgue measure zero.]

5.5 For the hierarchical model (5.11) of Strawderman (1971):

(a) Show that the Bayes estimator against squared error loss is given by $E(\boldsymbol{\theta}|\mathbf{x}) = [1 - E(\lambda|\mathbf{x})]\mathbf{x}$ where

$$E(\lambda|\mathbf{x}) = \frac{\int_0^1 \lambda^{r/2-a+1}e^{-1/2\lambda|\mathbf{x}|^2}\,d\lambda}{\int_0^1 \lambda^{r/2-a}e^{-1/2\lambda|\mathbf{x}|^2}\,d\lambda.}.$$

(b) Show that $E(\lambda|\mathbf{x})$ has the alternate representations

$$E(\lambda|\mathbf{x}) = \frac{r - 2a + 2}{|\mathbf{x}|^2}\frac{P(\chi^2_{r-2a+4} \le |\mathbf{x}|^2)}{P(\chi^2_{r-2a+2} \le |\mathbf{x}|^2},$$

$$E(\lambda|\mathbf{x}) = \frac{r - 2a + 2}{|\mathbf{x}|^2} - \frac{2e^{-1/2|\mathbf{x}|^2}}{|\mathbf{x}|^2\int_0^1 \lambda^{r/2-a}e^{-1/2\lambda|\mathbf{x}|^2}}\,d\lambda,$$

and hence that $a = 0$ gives the estimator of (5.12).

(c) Show that $|\mathbf{x}|^2E(\lambda|\mathbf{x})$ is increasing in $|\mathbf{x}|^2$ with maximum $r - 2a + 2$. Hence, the Bayes estimator is minimax if $r - 2a + 2 \le 2(r-2)$ or $r \ge 2(3-a)$. For $0 \le a \le 1$, this requires $r \ge 5$.

[Berger (1976b) considers matrix generalizations of this hierarchical model and derives admissible minimax estimators. Proper Bayes minimax estimators only exist if $r \ge 5$ (Strawderman 1971); however, formal Bayes minimax estimators exist for $r = 3$ and 4.]

5.6 Consider a generalization of the Strawderman (1971) hierarchical model of Problem 5.5:

$$\begin{aligned}\mathbf{X}|\boldsymbol{\theta} &\sim N(\boldsymbol{\theta}, I),\\ \boldsymbol{\theta}|\lambda &\sim N(0, \lambda^{-1}(1-\lambda)I),\\ \lambda &\sim \pi(\lambda).\end{aligned}$$

(a) Show that the Bayes estimator against squared error loss is $[1 - E(\lambda|\mathbf{x})]\mathbf{x}$, where

$$E(\lambda|\mathbf{x}) = \frac{\int_0^1 \lambda^{r/2+1} e^{-1/2\lambda|\mathbf{x}|^2}\pi(\lambda)\,d\lambda}{\int_0^1 \lambda^{r/2} e^{-1/2\lambda|\mathbf{x}|^2}\pi(\lambda)\,d\lambda}.$$

(b) Suppose $\lambda \sim \text{beta}(\alpha, \beta)$, with density

$$\pi(\lambda) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1-\lambda)^{\beta-1}.$$

Show that the Bayes estimator is minimax if $\beta \geq 1$ and $0 \leq a \leq (r-4)/2$.

[*Hint*: Use integration by parts on $E(\lambda|\mathbf{x})$, and apply Theorem 5.5. These estimators were introduced by Faith (1978).]

(c) Let $t = \lambda^{-1}(1-\lambda)$, the prior precision of $\boldsymbol{\theta}$. If $\lambda \sim \text{beta}(\alpha, \beta)$, show that the density of t is proportional to $t^{\alpha-1}/(1+t)^{\alpha+\beta}$, that is, $t \sim F_{2\alpha,2\beta}$, the F-distribution with 2α and 2β degrees of freedom.

[Strawderman's prior of Problem 5.5 corresponds to $\beta = 1$ and $0 < \alpha < 1$. If we take $\alpha = 1/2$ and $\beta = 1$, then $t \sim F_{1,2}$.]

(d) Two interesting limiting cases are $\alpha = 1, \beta = 0$ and $\alpha = 0, \beta = 1$. For each case, show that the resulting prior on t is proper, and comment on the minimaxity of the resulting estimators.

5.7 Faith (1978) considered the hierarchical model

$$\begin{aligned}\mathbf{X}|\boldsymbol{\theta} &\sim N(\boldsymbol{\theta}, I),\\ \boldsymbol{\theta}|t &\sim N\left(0, \frac{1}{t}I\right),\\ t &\sim \text{Gamma}(a, b),\end{aligned}$$

that is,

$$\pi(t) = \frac{1}{\Gamma(a)b^a}t^{a-1}e^{-t/b}.$$

(a) Show that the marginal prior for $\boldsymbol{\theta}$, unconditional on t, is

$$\pi(\boldsymbol{\theta}) \propto (2/b + |\boldsymbol{\theta}|^2)^{-(a+r/2)},$$

a multivariate Student's t-distribution.

(b) Show that $a \leq -1$ is a sufficient condition for $\sum_i \frac{\partial^2\pi(\boldsymbol{\theta})}{\partial\theta_i^2} \geq 0$ and, hence, is a sufficient condition for the minimaxity of the Bayes estimator against squared error loss.

(c) Show, more generally, that the Bayes estimator against squared error loss is minimax if $a \leq (r-4)/2$ and $a \leq 1/b + 3$.

(d) What choices of a and b would produce a multivariate Cauchy prior for $\pi(\boldsymbol{\theta})$? Is the resulting Bayes estimator minimax?

5.8 (a) Let $X \sim N(\boldsymbol{\theta}, \Sigma)$ and consider the estimation of $\boldsymbol{\theta}$ under the loss $L(\boldsymbol{\theta}, \delta) = (\boldsymbol{\theta} - \delta)'(\boldsymbol{\theta} - \delta)$. Show that $R(\boldsymbol{\theta}, \mathbf{X}) = \text{tr } \Sigma$, the minimax risk. Hence, $\mathbf{X}$ is a minimax estimator.

(b) Let $X \sim N(\boldsymbol{\theta}, I)$ and consider estimation of $\boldsymbol{\theta}$ under the loss $L(\boldsymbol{\theta}, \delta) = (\boldsymbol{\theta} - \delta)'Q(\boldsymbol{\theta} - \delta)$, where Q is a known positive definite matrix. Show that $R(\boldsymbol{\theta}, \mathbf{X}) = \text{tr } Q$, the minimax risk. Hence, $\mathbf{X}$ is a minimax estimator.

(c) Show that the calculations in parts (a) and (b) are equivalent.

5.9 In Theorem 5.7, verify

$$E_\theta \frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'(\boldsymbol{\theta} - \mathbf{X}) = E_\theta \left\{ \frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\text{tr}(\Sigma) - 2\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^4}\mathbf{X}'\Sigma\mathbf{X} + 2\frac{c'(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'\Sigma\mathbf{X} \right\}.$$

[*Hint*: There are several ways to do this:

(a) Write

$$\begin{aligned} E_\theta \frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'(\boldsymbol{\theta} - \mathbf{X}) &= E_\theta \frac{c(\mathbf{Y}'\mathbf{Y}}{\mathbf{Y}'\mathbf{Y}}\mathbf{Y}'\Sigma(\boldsymbol{\eta} - \mathbf{Y}) \\ &= \sum_i E_\theta \left\{ \frac{c(\mathbf{Y}'\mathbf{Y}}{\mathbf{Y}'\mathbf{Y}} \sum_j Y_j \sigma_{ji}(\eta_i - Y_i) \right\} \end{aligned}$$

where $\Sigma = \{\sigma_{ij}\}$ and $\mathbf{Y} = \Sigma^{-1/2}\mathbf{X} \sim N(\Sigma^{-1/2}\boldsymbol{\theta}, I) = N(\boldsymbol{\eta}, I)$. Now apply Stein's lemma.

(b) Write $\Sigma = PDP'$, where P is an orthogonal matrix ($P'P = I$) and D=diagonal matrix of eigenvalues of Σ, $D = \text{diagonal}\{d_i\}$. Then, establish that

$$E_\theta \frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'(\boldsymbol{\theta} - \mathbf{X}) = \sum_j E_\theta \frac{c(\sum_i d_i Z_i^2)}{\sum_i d_i Z_i^2} d_j Z_j(\eta_j^* - Z_j)$$

where $\mathbf{Z} = P\Sigma^{-1/2}\mathbf{X}$ and $\boldsymbol{\eta}^* = P\Sigma^{-1/2}\boldsymbol{\theta}$. Now apply Stein's lemma.

5.10 In Theorem 5.7, show that condition (i) allows the most shrinkage when $\Sigma = \sigma^2 I$, for some value of σ^2. That is, show that for all $r \times r$ positive definite Σ,

$$\max_\Sigma \frac{\text{tr } \Sigma}{\lambda_{\max}(\Sigma)} = \frac{\text{tr } \sigma^2 I}{\lambda_{\max}(\sigma^2 I)} = r.$$

[*Hint*: Write $\frac{\text{tr}\,\Sigma}{\lambda_{\max}(\Sigma)} = \sum \lambda_i/\lambda_{\max}$, where the λ_i's are the eigenvalues of Σ.]

5.11 The estimation problem of (5.18),

$$\begin{gathered} \mathbf{X} \sim N(\boldsymbol{\theta}, \Sigma) \\ L(\boldsymbol{\theta}, \delta) = (\boldsymbol{\theta} - \delta)'Q(\boldsymbol{\theta} - \delta), \end{gathered}$$

where both Σ and Q are positive definite matrices, can always be reduced, without loss of generality, to the simpler case,

$$\begin{gathered} \mathbf{Y} \sim N(\boldsymbol{\eta}, I) \\ L(\boldsymbol{\eta}, \delta^*) = (\boldsymbol{\eta} - \delta^*)'D_{q^*}(\boldsymbol{\eta} - \delta^*), \end{gathered}$$

where D_{q^*} is a diagonal matrix with elements $(q_1^*, \dots, q_r^*)$, using the following argument. Define $R = \Sigma^{1/2}B$, where $\Sigma^{1/2}$ is a symmetric square root of Σ (that is, $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$), and B is the matrix of eigenvectors of $\Sigma^{1/2}Q\Sigma^{1/2}$ (that is, $B'\Sigma^{1/2}Q\Sigma^{1/2}B = D_{q^*}$).

(a) Show that R satisfies

$$R'\Sigma^{-1}R = I, \quad R'QR = D_{q^*}$$

(b) Define $\mathbf{Y} = R^{-1}\mathbf{X}$. Show that $\mathbf{Y} \sim N(\boldsymbol{\eta}, I)$, where $\boldsymbol{\eta} = R^{-1}\boldsymbol{\theta}$.

(c) Show that estimations problems are equivalent if we define $\delta^*(\mathbf{Y}) = R^{-1}\delta(R\mathbf{Y})$.

[*Note*: If Σ has the eigenvalue-eigenvector decomposition $P'\Sigma P = D = \text{diagonal}(d_1, \cdots, d_r)$, then we can define $\Sigma^{1/2} = PD^{1/2}P'$, where $D^{1/2}$ is a diagonal matrix with elements $\sqrt{d_i}$. Since Σ is positive definite, the d_i's are positive.]

5.12 Complete the proof of Theorem 5.9.

(a) Show that the risk of $\delta(\mathbf{x})$ is

$$\begin{aligned} R(\boldsymbol{\theta}, \delta) = E_\theta &\left[(\boldsymbol{\theta} - \mathbf{X})'Q(\boldsymbol{\theta} - \mathbf{X})\right] \\ &-2E_\theta\left[\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'Q(\boldsymbol{\theta} - \mathbf{X})\right] \\ &+E_\theta\left[\frac{c^2(|\mathbf{X}|^2)}{|\mathbf{X}|^4}\mathbf{X}'Q\mathbf{X}\right] \end{aligned}$$

where $E_\theta(\boldsymbol{\theta} - \mathbf{X})'Q(\boldsymbol{\theta} - \mathbf{X}) = \text{tr}(Q)$.

(b) Use Stein's lemma to verify

$$\begin{aligned} &E_\theta\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'Q(\boldsymbol{\theta} - \mathbf{X}) \\ &= E_\theta\left\{\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\text{tr}(Q) - 2\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^4}\mathbf{X}'Q\mathbf{X} + 2\frac{c'(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'Q\mathbf{X}\right\}. \end{aligned}$$

Use an argument similar to the one in Theorem 5.7.

[*Hint*: Write

$$E_\theta\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\mathbf{X}'Q(\boldsymbol{\theta} - \mathbf{X}) = \sum_i E_\theta\left\{\frac{c(|\mathbf{X}|^2)}{|\mathbf{X}|^2}\sum_j X_j q_{ji}(\theta_i - X_i)\right\}$$

and apply Stein's lemma.]

5.13 Prove the following "generalization" of Theorem 5.9.

Theorem 8.2 *Let* $\mathbf{X} \sim N(\boldsymbol{\theta}, \Sigma)$. *An estimator of the form (5.13) is minimax against the loss* $L(\boldsymbol{\theta}, \delta) = (\boldsymbol{\theta} - \delta)'Q(\boldsymbol{\theta} - \delta)$, *provided*

(i) $0 \le c(|\mathbf{x}|^2) \le 2[\text{tr}(Q^*)/\lambda_{\max}(Q^*)] - 4$,

(ii) the function $c(\cdot)$ *is nondecreasing,*

where $Q^* = \Sigma^{1/2}Q\Sigma^{1/2}$.

5.14 Brown (1975) considered the performance of an estimator against a class of loss functions

$$\mathcal{L}(C) = \left\{L : L(\boldsymbol{\theta}, \delta) = \sum_{i=1}^r c_i(\theta_i - \delta_i)^2; \quad (c_1, \ldots, c_r) \in C\right\}$$

for a specified set C, and proved the following theorem.

Theorem 8.3 *For* $\mathbf{X} \sim N_r(\boldsymbol{\theta}, I)$, *there exists a spherically symmetric estimator* δ, *that is,* $\delta(\mathbf{x}) = [1 - h(|\mathbf{x}|^2)]\mathbf{x}$, *where* $h(|\mathbf{x}|^2) \neq 0$, *such that* $R(\boldsymbol{\theta}, \delta) \le R(\boldsymbol{\theta}, \mathbf{X})$ *for all* $L \in \mathcal{L}(C)$ *if, for all* $(c_1, \ldots, c_r) \in C$, *the inequality* $\sum_{j=1}^r c_i > 2c_k$ *holds for* $k = 1, \ldots, r$.

Show that this theorem is equivalent to Theorem 5.9 in that the above inequality is equivalent to part (i) of Theorem 5.9, and the estimator (5.13) is minimax.

[*Hint*: Identify the eigenvalues of Q with $c_1, \ldots, c_r$.]

Bock (1975) also establishes this theorem; see also Shinozaki (1980).

5.15 There are various ways to seemingly generalize Theorems 5.5 and 5.9. However, if both the estimator and loss function are allowed to depend on the covariance and loss matrix, then linear transformations can usually reduce the problem.

Let $X \sim N_r(\boldsymbol{\theta}, \Sigma)$, and let the loss function be $L(\boldsymbol{\theta}, \delta) = (\boldsymbol{\theta} - \delta)' Q(\boldsymbol{\theta} - \delta)$, and consider the following "generalizations" of Theorems 5.5 and 5.9.

$$\text{(a)} \qquad \delta(\mathbf{x}) = \left(1 - \frac{c(\mathbf{x}'\Sigma^{-1}\mathbf{x})}{\mathbf{x}'\Sigma^{-1}\mathbf{x}}\right)\mathbf{x}, \quad Q = \Sigma^{-1},$$

$$\text{(b)} \qquad \delta(\mathbf{x}) = \left(1 - \frac{c(\mathbf{x}'Q\mathbf{x})}{\mathbf{x}'Q\mathbf{x}}\right)\mathbf{x}, \quad \Sigma = I \text{ or } \Sigma = Q,$$

$$\text{(c)} \qquad \delta(\mathbf{x}) = \left(1 - \frac{c(\mathbf{x}'\Sigma^{-1/2}Q\Sigma^{-1/2}\mathbf{x})}{\mathbf{x}'\Sigma^{-1/2}Q\Sigma^{-1/2}\mathbf{x}}\right)\mathbf{x}.$$

In each case, use transformations to reduce the problem to that of Theorem 5.5 or 5.9, and deduce the condition for minimaxity of δ.

[*Hint*: For example, in (a) the transformation $Y = \Sigma^{-1/2}\mathbf{X}$ will show that δ is minimax if $0 < c(\cdot) < 2(r-2)$.]

5.16 A natural extension of the estimator (5.10) is to one that shrinks toward an arbitrary known point $\mu = (\mu_1, \ldots, \mu_r)$,

$$\delta_\mu(\mathbf{x}) = \mu + \left[1 - c(S)\frac{r-2}{|\mathbf{x} - \mu|^2}\right](\mathbf{x} - \mu)$$

where $|\mathbf{x} - \mu|^2 = \Sigma(x_i - \mu_i)^2$.

(a) Show that, under the conditions of Theorem 5.5, δ_μ is minimax.

(b) Show that its positive-part version is a better estimator.

5.17 Let $\mathbf{X} \sim N_r(\boldsymbol{\theta}, I)$. Show that the Bayes estimator of $\boldsymbol{\theta}$, against squared error loss, is given by $\delta(\mathbf{x}) = \mathbf{x} + \nabla \log m(\mathbf{x})$ where $m(\mathbf{x})$ is the marginal density function and $\nabla f = \{\partial/\partial x_i f\}$.

5.18 Verify (5.27).

[*Hint*: Show that, as a function of $|\mathbf{x}|^2$, the only possible interior extremum is a minimum, so the maximum must occur either at $|\mathbf{x}|^2 = 0$ or $|\mathbf{x}|^2 = \infty$.]

5.19 The property of superharmonicity, and its relationship to minimaxity, is not restricted to Bayes estimators. For $X \sim N_r(\boldsymbol{\theta}, I)$, a *pseudo-Bayes* estimator (so named, and investigated by Bock, 1988) is an estimator of the form

$$\mathbf{x} + \nabla \log m(\mathbf{x})$$

where $m(\mathbf{x})$ is not necessarily a marginal density.

(a) Show that the positive-part Stein estimator

$$\delta_a^+ = \mu + \left(1 - \frac{a}{|\mathbf{x} - \mu|^2}\right)^+ (\mathbf{x} - \mu)$$

is a pseudo-Bayes estimator with

$$m(\mathbf{x}) = \begin{cases} e^{-(1/2)|\mathbf{x}-\mu|^2} & \text{if } |\mathbf{x} - \mu|^2 < a \\ (|\mathbf{x} - \mu|^2)^{-a/2} & \text{if } |\mathbf{x} - \mu|^2 \geq a. \end{cases}$$

(b) Show that, except at the point of discontinuity, if $a \le r-2$, then $\sum_{i=1}^{r} \frac{\partial^2}{\partial x_i^2} m(\mathbf{x}) \le 0$, so $m(\mathbf{x})$ is superharmonic.

(c) Show how to modify the proof of Corollary 5.11 to accommodate superharmonic functions $m(\mathbf{x})$ with a finite number of discontinuities of measure zero.

This result is adapted from George (1986a, 1986b), who exploits both pseudo-Bayes and superharmonicity to establish minimaxity of an interesting class of estimators that are further investigated in the next problem.

5.20 For $X|\boldsymbol{\theta} \sim N_r(\boldsymbol{\theta}, I)$, George (1986a, 1986b) looked at *multiple shrinkage* estimators, those that can shrink to a number of different targets. Suppose that $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}) = \sum_{j=1}^{k} \omega_i \pi_i(\boldsymbol{\theta})$, where the ω_i are known positive weights, $\sum \omega_i = 1$.

(a) Show that the Bayes estimator against $\pi(\boldsymbol{\theta})$, under squared error loss, is given by $\delta^*(\mathbf{x}) = \mathbf{x} + \nabla \log m^*(\mathbf{x})$ where $m^*(\mathbf{x}) = \sum_{j=1}^{k} \omega_j m_j(\mathbf{x})$ and

$$m_i(\mathbf{x}) = \int_\Omega \frac{1}{(2\pi)^{p/2}} e^{-(1/2)|\mathbf{x}-\boldsymbol{\theta}|^2} \pi_i(\boldsymbol{\theta})\, d\boldsymbol{\theta}.$$

(b) Clearly, δ^* is minimax if $m^*(\mathbf{x})$ is superharmonic. Show that $\delta^*(\mathbf{x})$ is minimax if either (i) $m_i(\mathbf{x})$ is superharmonic, $i = 1, \ldots, k$, or (ii) $\pi_i(\boldsymbol{\theta})$ is superharmonic, $i = 1, \ldots, k$. [*Hint*: Problem 1.7.16]

(c) The real advantage of δ^* occurs when the components specify different targets. For $\rho_j = \omega_j m_j(x)/m^*(x)$, let $\delta^*(\mathbf{x}) = \sum_{j=1}^{k} \rho_j \delta_j^+(\mathbf{x})$ where

$$\delta_j^+(\mathbf{x}) = \mu_j + \left(1 - \frac{r-2}{|\mathbf{x}-\mu_j|^2}\right)^+ (\mathbf{x} - \mu_j)$$

and the μ_j's are target vectors. Show that $\delta^*(\mathbf{x})$ is minimax. [*Hint*: Problem 5.19]

[George (1986a, 1986b) investigated many types of multiple targets, including multiple points, subspaces, and clusters and subvectors. The subvector problem was also considered by Berger and Dey (1983a, 1983b). Multiple shrinkage estimators were also investigated by Ki and Tsui (1990) and Withers (1991).]

5.21 Let X_i, Y_j be independent $N(\xi_i, 1)$ and $N(\eta_j, 1)$, respectively ($i = 1, \ldots, r$; $j = 1, \ldots, s$).

(a) Find an estimator of $(\xi_1, \ldots, \xi_r; \eta_1, \ldots, \eta_s)$ that would be good near $\xi_i = \cdots = \xi_r = \xi$, $\eta_1 = \cdots = \eta_s = \eta$, with ξ and η unknown, if the variability of the ξ's and η's is about the same.

(b) When the loss function is (4.17), determine the risk function of your estimator.

[*Hint*: Consider the Bayes situation in which $\xi_i \sim N(\xi, A)$ and $\eta_j \sim N(\eta, A)$. See Berger 1982b for further development of such estimators].

5.22 The early proofs of minimaxity of Stein estimators (James and Stein 1961, Baranchik 1970) relied on the representation of a noncentral χ^2-distribution as a Poisson sum of central χ^2 (TSH2, Problem 6.7). In particular, if $\chi_r^2(\lambda)$ is a noncentral χ^2 random variable with noncentrality parameter λ, then

$$E_\lambda h(\chi_r^2(\lambda)) = E[Eh(\chi_{r+2K}^2)|K]$$

where $K \sim \text{Poisson}(\lambda)$ and χ_{r+2k}^2 is a central χ^2 random variable with $r + 2k\, df$. Use this representation, and the properties of the central χ^2-distribution, to establish the following identities for $X \sim N_r(\boldsymbol{\theta}, I)$ and $\lambda = |\boldsymbol{\theta}|^2$.

(a) $E_\theta \frac{\mathbf{X}'\boldsymbol{\theta}}{|\mathbf{X}|^2} = |\boldsymbol{\theta}|^2 E \frac{1}{\chi^2_{r+2}(\lambda)}$.

(b) $(r-2)E\frac{1}{\chi^2_r(\lambda)} + |\boldsymbol{\theta}|^2 E \frac{1}{\chi^2_{r+2}(\lambda)} = 1$.

(c) For $\delta(\mathbf{x}) = (1 - c/|\mathbf{x}|^2)\mathbf{x}$, use the identities (a) and (b) to show that for $L(\boldsymbol{\theta}, \delta) = |\boldsymbol{\theta} - \delta|^2$,

$$\begin{aligned} R(\boldsymbol{\theta}, \delta) &= r + 2c|\boldsymbol{\theta}|^2 E \frac{1}{\chi^2_{r+2}(\lambda)} - 2c + c^2 E \frac{1}{\chi^2_r(\lambda)} \\ &= r + 2c\left[1 - (r-2)E\frac{1}{\chi^2_r(\lambda)}\right] - 2c + c^2 E \frac{1}{\chi^2_r(\lambda)} \end{aligned}$$

and, hence, that $\delta(\mathbf{x})$ is minimax if $0 \le c \le 2(r-2)$.

[See Bock 1975 or Casella 1980 for more identities involving noncentral χ^2 expectations.]

5.23 Let $\chi^2_r(\lambda)$ be a χ^2 random variable with r degrees of freedom and noncentrality parameter λ.

(a) Show that $E\frac{1}{\chi^2_r(\lambda)} = E\left[E\frac{1}{\chi^2_{r+2K}}|K\right] = E\left[\frac{1}{r-2+2K}\right]$, where $K \sim$ Poisson$(\lambda/2)$.

(b) Establish (5.32).

5.24 For the most part, the risk function of a Stein estimator increases as $|\boldsymbol{\theta}|$ moves away from zero (if zero is the shrinkage target). To guarantee that the risk function is monotone increasing in $|\boldsymbol{\theta}|$ (that is, that there are no "dips" in the risk as in Berger's 1976a tail minimax estimators) requires a somewhat stronger assumption on the estimator (Casella 1990). Let $X \sim N_r(\boldsymbol{\theta}, I)$ and $L(\boldsymbol{\theta}, \delta) = |\boldsymbol{\theta} - \delta|^2$, and consider the Stein estimator

$$\delta(\mathbf{x}) = \left(1 - c(|\mathbf{x}|^2)\frac{(r-2)}{|\mathbf{x}|^2}\right)\mathbf{x}.$$

(a) Show that if $0 \le c(\cdot) \le 2$ and $c(\cdot)$ is concave and twice differentiable, then $\delta(\mathbf{x})$ is minimax. [*Hint*: Problem 1.7.7.]

(b) Under the conditions in part (a), the risk function of $\delta(\mathbf{x})$ is nondecreasing in $|\boldsymbol{\theta}|$. [*Hint*: The conditions on $c(\cdot)$, together with the identity

$$(d/d\lambda)E_\lambda[h(\chi^2_p(\lambda))] = E_\lambda\{[\partial/\partial\chi^2_{p+2}(\lambda)]h(\chi^2_{p+2}(\lambda))\},$$

where $\chi^2_p(\lambda)$ is a noncentral χ^2 random variable with p degrees of freedom and noncentrality parameter λ, can be used to show that $(\partial/\partial|\boldsymbol{\theta}|^2)R(\boldsymbol{\theta}, \delta) > 0$.]

5.25 In the spirit of Stein's "large r and $|\boldsymbol{\theta}|$" argument, Casella and Hwang (1982) investigated the limiting risk ratio of $\delta^{JS}(\mathbf{x}) = (1-(r-2)/|\mathbf{x}|^2)\mathbf{x}$ to that of $\mathbf{x}$. If $X \sim N_r(\boldsymbol{\theta}, I)$ and $L(\boldsymbol{\theta}, \delta) = |\boldsymbol{\theta} - \delta|^2$, they showed

$$\lim_{\substack{r\to\infty \\ \frac{|\boldsymbol{\theta}|^2}{r} \to c}} \frac{R(\boldsymbol{\theta}, \delta^{JS})}{R(\boldsymbol{\theta}, \mathbf{X})} = \frac{c}{c+1}.$$

To establish this limit we can use the following steps.

(a) Show that $\frac{R(\boldsymbol{\theta}, \delta^{JS})}{R(\boldsymbol{\theta}, x)} = 1 - \frac{(r-2)^2}{r} E_\theta \frac{1}{|\mathbf{X}|^2}$.

(b) Show that $\frac{1}{p-2+|\boldsymbol{\theta}|^2} \le E_\theta \frac{1}{|\mathbf{X})|^2} \le \frac{1}{p-2}\left(\frac{p}{p+|\boldsymbol{\theta}|^2}\right)$.

(c) Show that the upper and lower bounds on the risk ratio both have the same limit.

[*Hint*: (b) The upper bound is a consequence of Problem 5.22(b). For the lower bound, show $E_\theta(1/|\mathbf{X})|^2) = E(1/p - 2 + K)$, where $K \sim$ Poisson($|\boldsymbol{\theta}|^2$) and use Jensen's inequality.]

Section 6

6.1 Referring to Example 6.1, this problem will establish the validity of the expression (6.2) for the risk of the estimator δ^L of (6.1), using an argument similar to that in the proof of Theorem 5.7.

(a) Show that

$$\begin{aligned} R(\boldsymbol{\theta}, \delta^L) &= \sum_i E_\theta[\theta_i - \delta_i^L(\mathbf{X})]^2 \\ &= \sum_i E_\theta \left\{(\theta_i - X_i)^2 + \frac{2c(r-3)}{S}(\theta_i - X_i)(X_i - \bar{\mathbf{X}}) \right. \\ &\quad \left. + \frac{[c(r-3)]^2}{S^2}(X_i - \bar{\mathbf{X}})^2 \right\} \end{aligned}$$

where $S = \sum_j (X_j - \bar{\mathbf{X}})^2$.

(b) Use integration by parts to show

$$E_\theta \frac{(\theta_i - X_i)(X_i - \bar{\mathbf{X}})}{S} = -E_\theta \frac{\frac{r-1}{r} S + 2(X_i - \bar{\mathbf{X}})^2}{S^2}.$$

[*Hint*: Write the cross-term as $-E_\theta \left[\frac{(X_i - \bar{\mathbf{X}})}{S}\right](X_i - \theta_i)$ and adapt Stein's identity (Lemma 1.5.15).]

(c) Use the results of parts (a) and (b) to establish (6.2).

6.2 In Example 6.1, show that:

(a) The estimator δ^L is minimax if $r \geq 4$ and $c \leq 2$.

(b) The risk of δ^L is infinite if $r \leq 3$

(c) The minimum risk is equal to $3/r$, and is attained at $\theta_1 = \theta_2 = \cdots = \theta$.

(d) The estimator δ^L is dominated in risk by its positive-part version

$$\delta^{L^+} = \bar{x}\mathbf{1} + \left(1 - \frac{c(r-3)}{|\mathbf{x} - \bar{x}\mathbf{1}|^2}\right)^+ (\mathbf{x} - \bar{x}\mathbf{1}).$$

6.3 In Example 6.2:

(a) Show that $k\mathbf{x}$ is the MLE if $\boldsymbol{\theta} \in \mathcal{L}_k$.

(b) Show that $\delta^k(\mathbf{x})$ of (6.8) is minimax under squared error loss.

(c) Verify that θ_i of the form (6.4) satisfy $T(T'T)^{-1}T'\boldsymbol{\theta} = \boldsymbol{\theta}$ for T of (6.5), and construct a minimax estimator that shrinks toward this subspace.

6.4 Consider the problem of estimating the mean based on $X \sim N_r(\boldsymbol{\theta}, I)$, where it is thought that $\theta_i = \sum_{j=1}^s \beta_j t_i^j$ where $(t_i, \ldots, t_r)$ are known, $(\beta_1, \ldots, \beta_s)$ are unknown, and $r - s > 2$.

(a) Find the MLE of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}_R$, if $\boldsymbol{\theta}$ is assumed to be in the linear subspace

$$\mathcal{L} = \left\{\boldsymbol{\theta} : \sum_{j=1}^s \beta_j t_i^j = \theta_i, \quad i = 1, \ldots, r\right\}.$$

(b) Show that $\mathcal{L}$ can be written in the form (6.7), and find K.

(c) Construct a Stein estimator that shrinks toward the MLE of part (a) and prove that it is minimax.

6.5 For the situation of Example 6.3:

(a) Show that $\delta^c(\mathbf{x}, \mathbf{y})$ is minimax if $0 \le c \le 2$.

(b) Show that if $\boldsymbol{\xi} = 0$, $R(\boldsymbol{\theta}, \delta^1) = 1 - \frac{\sigma^2}{\sigma^2+\tau^2}\frac{r-2}{r}$, $R(\boldsymbol{\theta}, \delta^{\text{comb}}) = 1 - \frac{\sigma^2}{\sigma^2+\tau^2}$, and, hence, $R(\boldsymbol{\theta}, \delta^1) > R(\boldsymbol{\theta}, \delta^{\text{comb}})$.

(c) For $\boldsymbol{\xi} \neq 0$, show that $R(\boldsymbol{\theta}, \delta^{\text{comb}}) = 1 - \frac{\sigma^2}{\sigma^2+\tau^2} + \frac{|\boldsymbol{\xi}|^2\sigma^2}{r(\sigma^2+\tau^2)}$ and hence is unbounded as $|\boldsymbol{\xi}| \to \infty$.

6.6 The Green and Strawderman (1991) estimator $\delta^c(\mathbf{x}, \mathbf{y})$ can be derived as an empirical Bayes estimator.

(a) For $X|\boldsymbol{\theta} \sim N_r(\boldsymbol{\theta}, \sigma^2 I)$, $Y|\boldsymbol{\theta}, \boldsymbol{\xi} \sim N_r(\boldsymbol{\theta} + \boldsymbol{\xi}, \tau^2 I)$, $\boldsymbol{\xi} \sim N(0, \gamma^2 I)$, and $\theta_i \sim$ Uniform$(-\infty, \infty)$, with σ^2 and τ^2 assumed to be known, show how to derive $\delta^{r-2}(\mathbf{x}, \mathbf{y})$ as an empirical Bayes estimator.

(b) Calculate the Bayes estimator, δ^π, against squared error loss.

(c) Compare $r(\pi, \delta^\pi)$ and $r(\pi, \delta^{r-2})$.

[*Hint*: For part (a), Green and Strawderman suggest starting with $\boldsymbol{\theta} \sim N(0, \kappa^2 I)$ and let $\kappa^2 \to \infty$ get the uniform prior.]

6.7 In Example 6.4:

(a) Verify the risk function (6.13).

(b) Verify that for unknown σ^2, the risk function of the estimator (6.14) is given by (6.15).

(c) Show that the minimum risk of the estimator (6.14) is $1 - \frac{\nu}{\nu+2}\frac{r-2}{r}$.

6.8 For the situation of Example 6.4, the analogous modification of the Lindley estimator (6.1) is

$$\delta^L = \bar{x}\mathbf{1} + \left(1 - \frac{r-3}{\Sigma(x_i - \bar{x})^2/\hat{\sigma}^2}\right)(\mathbf{x} - \bar{x}\mathbf{1}),$$

where $\hat{\sigma}^2 = S^2/(\nu + 2)$ and $S^2/\sigma^2 \sim \chi^2_\nu$, independent of $\mathbf{X}$.

(a) Show that $R(\boldsymbol{\theta}, \delta^L) = 1 - \frac{\nu}{\nu+2}\frac{(r-3)^2}{r}E_\theta\frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}$.

(b) Show that both δ^L and δ of (6.14) can be improved by using their positive-part versions.

6.9 The major application of Example 6.4 is to the situation

$$Y_{ij} \sim N(\theta_i, \sigma^2), \; i = 1, \ldots, s, \; j = 1, \ldots, n, \text{ independent}$$

with $\bar{Y}_i = (1/n)\Sigma_j Y_{ij}$ and $\hat{\sigma}^2 = \Sigma_{ij}(Y_{ij} - \overline{Y}_i)^2/s(n-1)$. Show that the estimator

$$\delta_i = \bar{\bar{y}} + \left(1 - c\frac{(s-3)\hat{\sigma}^2}{\Sigma(\bar{y}_i - \bar{\bar{y}})^2}\right)^+ (\bar{y}_i - \bar{\bar{y}})$$

is a minimax estimator, where $\bar{y} = \Sigma_{ij}y_{ij}/sn$, as long as $0 \le c \le 2$.

[The case of unequal sample sizes n_i is not covered by what we have done so far. See Efron and Morris 1973b, Berger and Bock 1976, and Morris 1983 for approaches to this problem. The case of totally unknown covariance matrix is considered by Berger et al. (1977) and Gleser (1979, 1986).]

6.10 The positive-part Lindley estimator of Problem 6.9 has an interesting interpretation in the one-way analysis of variance, in particular with respect to the usual test performed, that of $H_0 : \theta_1 = \theta_2 = \cdots = \theta_s$. This hypothesis is tested with the statistic

$$F = \frac{\Sigma(\bar{y}_i - \bar{\bar{y}})^2/(s-1)}{\Sigma(y_{ij} - \bar{y}_i)^2/s(n-1)},$$

which, under H_0, has an F-distribution with $s-1$ and $s(n-1)$ degrees of freedom.

(a) Show that the positive-part Lindley estimator can be written as

$$\delta_i = \bar{\bar{y}} + \left(1 - c\frac{s-3}{s-1}\frac{1}{F}\right)^+ (\bar{y}_i - \bar{\bar{y}}).$$

(b) The null hypothesis is rejected if F is large. Show that this corresponds to using the MLE under H_0 if F is small, and a Stein estimator if F is large.

(c) The null hypothesis is rejected at level α if $F > F_{s-1,s(n-1),\alpha}$. For $s = 8$ and $n = 6$:

(i) What is the level of the test that corresponds to choosing $c = 1$, the optimal risk choice?

(ii) What values of c correspond to choosing $\alpha = .05$ or $\alpha = .01$, typical α levels. Are the resulting estimators minimax?

6.11 Prove the following extension of Theorem 5.5 to the case of unknown variance, due to Strawderman (1973).

Theorem 8.4 *Let $X \sim N_r(\boldsymbol{\theta}, \sigma^2 I)$ and let $S^2/\sigma^2 \sim \chi^2_\nu$, independent of X. The estimator*

$$\delta^c(\mathbf{x}) = \left(1 - \frac{c(F, S^2)}{S^2}\frac{r-2}{\nu+2}\right)\mathbf{x},$$

where $F = \Sigma x_i^2/S^2$, is a minimax estimator of $\boldsymbol{\theta}$, provided

(i) for each fixed S^2, $c(\cdot, S^2)$ is nondecreasing,

(ii) for each fixed F, $c(F, \cdot)$ is nonincreasing,

(iii) $0 \le c(\cdot, \cdot) \le 2$.

[Note that, here, the loss function is taken to be scaled by σ^2, $L(\boldsymbol{\theta}, \delta) = |\boldsymbol{\theta} - \delta|^2/\sigma^2$, otherwise the minimax risk is not finite. Strawderman (1973) went on to derive proper Bayes minimax estimators in this case.]

6.12 For the situation of Example 6.5:

(a) Show that $E_\sigma \frac{1}{\sigma^2} = E_0 \frac{r-2}{|\mathbf{X}|^2}$.

(b) If $1/\sigma^2 \sim \chi^2_\nu/\nu$, then $f(|\mathbf{x} - \boldsymbol{\theta}|)$ of (6.19) is the multivariate t-distribution, with ν degrees of freedom and $E_0|\mathbf{X}|^{-2} = (r-2)^{-1}$.

(c) If $1/\sigma^2 \sim Y$, where χ^2_ν/ν is stochastically greater than Y, then $\delta(\mathbf{x})$ of (6.20) is minimax for this mixture as long as $0 \le c \le 2(r-2)$.

6.13 Prove Lemma 6.2.

6.14 For the situation of Example 6.7:

(a) Verify that the estimator (6.25) is minimax if $0 \le c \le 2$. (Theorem 5.5 will apply.)

(b) Referring to (6.27), show that

$$E|\delta^\pi(\mathbf{X}) - \delta^c(\mathbf{X})|^2 I[|\mathbf{X}|^2 \geq c(r-2)(\sigma^2+\tau^2)]$$
$$= \frac{\sigma^4}{\sigma^2+\tau^2} E \frac{1}{Y}[Y - c(r-2)]^2 I[Y \geq c(r-2)]$$

where $Y \sim \chi^2_r$.

(c) If χ^2_ν denotes a chi-squared random variable with ν degrees of freedom, establish the identity

$$Eh(\chi^2_\nu) = \nu E \frac{h(\chi^2_{\nu+2})}{\chi^2_{\nu+2}}$$

to show that

$$r(\pi, \delta^R) = r(\pi, \delta^\pi) + \frac{1}{r-2} \frac{\sigma^4}{\sigma^2+\tau^2} E[Y - c(r-2)]^2 I[Y \geq c(r-2)]$$

where, now, $Y \sim \chi^2_{r-2}$.

(d) Verify (6.29), hence showing that $r(\pi, \delta^R) \leq r(\pi, \delta^c)$.

(e) Show that $E(Y-a)^2 I(Y > a)$ is a decreasing function of a, and hence the maximal Bayes risk improvement, while maintaining minimaxity, is obtained at $c = 2$.

6.15 For $X_i \sim \text{Poisson}(\lambda_i)$ $i = 1, \ldots, r$, independent, and loss function $L(\boldsymbol{\lambda}, \delta) = \Sigma(\lambda_i - \delta_i)^2/\lambda_i$:

(a) For what values of a, α, and β are the estimators of (4.6.29) minimax? Are they also proper Bayes for these values?

(b) Let $\Lambda = \Sigma\lambda_i$ and define $\theta_i = \lambda_i/\Lambda, i = 1, \ldots, r$. For the prior distribution $\pi(\boldsymbol{\theta}, \Lambda) = m(\Lambda)d\Lambda \prod_{i=1}^r d\theta_i$, show that the Bayes estimator is

$$\delta^\pi(\mathbf{x}) = \frac{\psi_\pi(z)}{z+r-1}\mathbf{x},$$

where $z = \Sigma x_i$ and

$$\psi_\pi(z) = \frac{\int \Lambda^z e^{-\Lambda} m(\Lambda)\, d\Lambda}{\int \Lambda^{z-1} e^{-\Lambda} m(\Lambda)\, d\Lambda}.$$

(c) Show that the choice $m(\Lambda) = 1$, yields the estimator $\delta(\mathbf{x}) = [1-(r-1)/(z+r-1)]\mathbf{x}$, which is minimax.

(d) Show that the choice $m(\Lambda) = (1+\Lambda)^{-\beta}$, $1 \leq \beta \leq r-1$ yields an estimator that is proper Bayes minimax for $r > 2$.

(e) The estimator of part (d) is difficult to evaluate. However, for the prior choice

$$m(\Lambda) = \int_0^\infty \frac{t^{-r}e^{-1/t}}{(1+\Lambda t)^\beta}\, dt, \quad 1 \leq \beta \leq r-1,$$

show that the generalized Bayes estimator is

$$\delta^\pi(\mathbf{x}) = \frac{z}{z+\beta+r-1}\mathbf{x},$$

and determine conditions for its minimaxity. Show that it is proper Bayes if $\beta > 1$.

6.16 Let $X_i \sim \text{binomial}(p, n_i), i = 1, \ldots, r$, where n_i are unknown and p is known. The estimation target is $\mathbf{n} = (n_1, \ldots, n_r)$ with loss function

$$L(\mathbf{n}, \delta) = \sum_{i=1}^r \frac{1}{n_i}(n_i - \delta_i)^2.$$

(a) Show that the usual estimator $\mathbf{x}/p$ has constant risk $r(1-p)/p$.

(b) For $r \geq 2$, show that the estimator

$$\delta(\mathbf{x}) = \left(1 - \frac{a}{z + r - 1}\right) \frac{\mathbf{x}}{p}$$

dominates $\mathbf{x}/p$ in risk, where $z = \Sigma x_i$ and $0 < a < 2(r-1)(1-p)$.

[*Hint*: Use an argument similar to Theorem 6.8, but here $X_i|Z$ is hypergeometric, with $E(X_i|z) = z\frac{n_i}{N}$ and $\text{var}(X_i|z) = z\frac{n_i}{N}\left(1 - \frac{n_i}{N}\right)\frac{N-z}{N-1}$, where $N = \Sigma n_i$.]

(c) Extend the argument from part (b) and find conditions on the function $c(\cdot)$ and constant b so that

$$\delta(\mathbf{x}) = \left(1 - \frac{c(z)}{z + b}\right) \frac{\mathbf{x}}{p}$$

dominates $\mathbf{x}/p$ in risk.

Domination of the usual estimator of $\mathbf{n}$ was looked at by Feldman and Fox (1968), Johnson (1987), and Casella and Strawderman (1994). The problem of $\mathbf{n}$ estimation for the binomial has some interesting practical applications; see Olkin et al. 1981, Carroll and Lombard 1985, Casella 1986. Although we have made the unrealistic assumption that p is known, these results can be adapted to the more practical unknown p case (see Casella and Strawderman 1994 for details).

6.17 (a) Prove Lemma 6.9. [*Hint*: Change variables from $\mathbf{x}$ to $\mathbf{x} - e_i$, and note that h_i must be defined so that $\delta^0(0) = 0$.]

(b) Prove that for $X \sim p_i(x|\theta)$, where $p_i(x|\theta)$ is given by (6.36), $\delta^0(x) = h_i(x-1)/h_i(x)$ is the UMVU estimator of θ (Roy and Mitra 1957).

(c) Prove Theorem 6.10.

6.18 For the situation of Example 6.11:

(a) Establish that $\mathbf{x} + g(\mathbf{x})$, where $g(\mathbf{x})$ is given by (6.42), satisfies $\mathcal{D}(\mathbf{x}) \leq 0$ for the loss $L_0(\boldsymbol{\theta}, \delta)$ of (6.38), and hence dominates $\mathbf{x}$ in risk.

(b) Derive $\mathcal{D}(\mathbf{x})$ for $X_i \sim \text{Poisson}(\lambda_i)$, independent, and loss $L_{-1}(\boldsymbol{\lambda}, \delta)$ of (6.38). Show that $\mathbf{x} + g(\mathbf{x})$, for $g(\mathbf{x})$ given by (6.43), satisfies $\mathcal{D}(\mathbf{x}) \leq 0$ and hence is a minimax estimator of $\boldsymbol{\lambda}$.

6.19 For the situation of Example 6.12:

(a) Show that the estimator $\delta_0(\mathbf{x}) + g(\mathbf{x})$, for $g(\mathbf{x})$ of (6.45) dominates δ^0 in risk under the loss $L_{-1}(\boldsymbol{\theta}, \delta)$ of (6.38) by establishing that $\mathcal{D}(\mathbf{x}) \leq 0$.

(b) For the loss $L_0(\boldsymbol{\theta}, \delta)$ of (6.38), show that the estimator $\delta^0(\mathbf{x}) + g(\mathbf{x})$, where

$$g_i(\mathbf{x}) = \frac{c(\mathbf{x}) k_i(x_i)}{\sum_{j=1}^r [k_j^2(x_j) + \left(\frac{1+t_j}{2}\right) k_j(x_j)]},$$

with $k_i(x) = \Sigma_{\ell=1}^x (t_i - 1 + \ell)/\ell$ and $c(\cdot)$ nondecreasing with $0 \leq c(\cdot) \leq 2[(\#x_i s > 1) - 2]$ has $\mathcal{D}(\mathbf{x}) \leq 0$ and hence dominates $\delta^0(\mathbf{x})$ in risk.

6.20 In Example 6.12, we saw improved estimators for the success probability of negative binomial distributions. Similar results hold for estimating the means of the negative binomial distributions, with some added features of interest. Let $X_1, \ldots, X_r$ be independent negative binomial random variables with mass function (6.44), and suppose we want to estimate $\mu = \{\mu_i\}$, where $\mu_i = t_i\theta_i/(1-\theta_i)$, the mean of the ith distribution, using the loss $L(\mu, \delta) = \Sigma(\mu_i - \delta_i)^2/\mu_i$.

(a) Show that the MLE of μ is X, and the risk of an estimator $\delta(\mathbf{x}) = \mathbf{x} + g(\mathbf{x})$ can be written

$$R(\mu, \delta) = R(\mu, \mathbf{X}) + E_\mu[\mathcal{D}_1(\mathbf{X}) + \mathcal{D}_2(\mathbf{X})]$$

where

$$\mathcal{D}_1(\mathbf{x}) = \sum_{i=1}^r \left\{ 2[g_i(\mathbf{x} + e_i) - g_i(\mathbf{x})] + \frac{g_i^2(\mathbf{x} + e_i)}{x_i + 1} \right\}$$

and

$$\mathcal{D}_2(\mathbf{x}) = \sum_{i=1}^r \left\{ 2\frac{x_i}{t_i}[g_i(\mathbf{x} + e_i) - g_i(\mathbf{x})] + \frac{g_i^2(\mathbf{x} + e_i)}{t_i} \left[\frac{x_i}{x_{i+1}} - 1 \right] \right\}$$

so that a sufficient condition for domination of the MLE is $\mathcal{D}_1(\mathbf{x}) + \mathcal{D}_2(\mathbf{x}) \le 0$ for all $\mathbf{x}$. [Use Lemma 6.9 in the form $Ef(\mathbf{X})/\theta_i = E\left[\frac{t_i + X_i}{X_i + 1} f(\mathbf{X} + e_i)\right]$.]

(b) Show that if X_i are Poisson(θ_i) (instead of negative binomial), then $\mathcal{D}_2(\mathbf{x}) = 0$. Thus, any estimator that dominates the MLE in the negative binomial case also dominates the MLE in the Poisson case.

(c) Show that the Clevenson-Zidek estimator

$$\delta_{cz}(\mathbf{x}) = \left(1 - \frac{c(r-1)}{\Sigma x_i + r - 1}\right) \mathbf{x}$$

satisfies $\mathcal{D}_1(\mathbf{x}) \le 0$ and $\mathcal{D}_2(\mathbf{x}) \le 0$ and, hence, dominates the MLE under both the Poisson and negative binomial model.

This robustness property of Clevenson-Zidek estimators was discovered by Tsui (1984) and holds for more general forms of the estimator. Tsui (1984, 1986) also explores other estimators of Poisson and negative binomial means and their robustness properties.

Section 7

7.1 Establish the claim made in Example 7.2. Let X_1 and X_2 be independent random variables, $X_i \sim N(\theta_i, 1)$, and let $L((\theta_1, \theta_2), \delta) = (\theta_1 - \delta)^2$. Show that $\delta = \text{sign}(X_2)$ is an admissible estimate of θ_1, even though its distribution does not depend on θ_1.

7.2 Efron and Morris (1973a) give the following derivation of the positive-part Stein estimator as a *truncated* Bayes estimator. For $X \sim N_r(\boldsymbol{\theta}, \sigma^2 I), r \ge 3$, and $\boldsymbol{\theta} \sim N(0, \tau^2 I)$, where σ^2 is known and τ^2 is unknown, define $t = \sigma^2/(\sigma^2 + \tau^2)$ and put a prior $h(t), 0 < t < 1$ on t.

(a) Show that the Bayes estimator against squared error loss is given by $E(\boldsymbol{\theta}|\mathbf{x}) = [1 - E(t|\mathbf{x})]\mathbf{x}$ where

$$\pi(t|\mathbf{x}) = \frac{t^{r/2} e^{-t|\mathbf{x}|^2/2} h(t)}{\int_0^1 t^{r/2} e^{-t|\mathbf{x}|^2/2} h(t)\, dt}.$$

(b) For estimators of the form $\delta^\tau(\mathbf{x}) = \left(1 - \tau(|\mathbf{x}|^2)\frac{r-2}{|\mathbf{x}|^2}\right)\mathbf{x}$, the estimator that satisfies

(i) $\tau(\cdot)$ is nondecreasing,
(ii) $\tau(\cdot) \le c$,
(iii) δ^τ minimizes the Bayes risk against $h(t)$

has $\tau(|\mathbf{x}|^2) = \tau^*(|\mathbf{x}|^2) = \min\{c, \frac{|\mathbf{x}|^2}{r-2} E(t|\mathbf{x})\}$. (This is a truncated Bayes estimator, and is minimax if $c \leq 2$.)

(c) Show that if $h(t)$ puts all of its mass on $t = 1$, then

$$\tau^*(|\mathbf{x}|^2) = \min\left\{c, \frac{|\mathbf{x}|^2}{r-2}\right\}$$

and the resulting truncated Bayes estimator is the positive-part estimator.

7.3 Fill in the details of the proof of Lemma 7.5.

7.4 For the situation of Example 7.8, show that if δ_0 is any estimator of θ, then the class of all estimators with $\delta(x) < \delta_0(x)$ for some x is complete.

7.5 A decision problem is *monotone* (as defined by Karlin and Rubin 1956; see also Brown, Cohen and Strawderman 1976 and Berger 1985, Section 8.4) if the loss function $L(\theta, \delta)$ is, for each θ, minimized at $\delta = \theta$ and is an increasing function of $|\delta - \theta|$. An estimator δ is *monotone* if it is a nondecreasing function of x.

(a) Show that if $L(\theta, \delta)$ is convex, then the monotone estimators form a complete class.

(b) If $\delta(x)$ is not monotone, show that the monotone estimator δ' defined implicitly by

$$P_t(\delta'(X) \leq t) = P_t(\delta(X) \leq t) \quad \text{for every } t$$

satisfies $R(\theta, \delta') \leq R(\theta, \delta)$ for all θ.

(c) If $X \sim N(\theta, 1)$ and $L(\theta, \delta) = (\theta - \delta)^2$, construct a monotone estimator that dominates

$$\delta^a(x) = \begin{cases} -2a - x & \text{if } x < -a \\ x & \text{if } |x| \leq a \\ 2a - x & \text{if } x > a. \end{cases}$$

7.6 Show that, in the following estimation problems, all risk functions are continuous.

(a) Estimate θ with $L(\theta, \delta(x)) = [\theta - \delta(x)]^2$, $X \sim N(\theta, 1)$.

(b) Estimate θ with $L(\theta, \delta(\mathbf{x})) = |\theta - \delta(\mathbf{x})|^2$, $X \sim N_r(\theta, I)$.

(c) Estimate $\boldsymbol{\lambda}$ with $L(\boldsymbol{\lambda}, \delta(\mathbf{x})) = \sum_{i=1}^r \lambda_i^{-m}(\lambda_i - \delta_i(\mathbf{x}))^2$, $X_i \sim$ Poisson(λ_i), independent.

(d) Estimate β with $L(\beta, \delta(\mathbf{x})) = \sum_{i=1}^r \beta_i^{-m}(\beta_i - \delta_i(\mathbf{x}))^2$, $X_i \sim$ Gamma(α_i, β_i), independent, α_i known.

7.7 Prove the following theorem, which gives sufficient conditions for estimators to have continuous risk functions.

Theorem 8.5 (Ferguson 1967, Theorem 3.7.1) *Consider the estimation of θ with loss $L(\theta, \delta)$, where $X \sim f(x|\theta)$. Assume*

(i) the loss function $L(\theta, \delta)$ is bounded and continuous in θ uniformly in δ (so that $\lim_{\theta \to \theta_0} \sup_\delta |L(\theta, \delta) - L(\theta_0, \delta)| = 0$);

(ii) for any bounded function φ, $\int \varphi(x) f(x|\theta) d\mu(x)$ is continuous in θ.

Then, the risk function $R(\theta, \delta) = E_\theta L(\theta, \delta)$ is continuous in θ.

[*Hint*: Show that

$$|R(\theta', \delta) - R(\theta, \delta)| \leq \int |L(\theta', \delta(x)) - L(\theta, \delta(x))| f(x|\theta')\, dx + \int L(\theta, \delta(x)) |f(x|\theta') - f(x|\theta)|\, dx,$$

and use (i) and (ii) to make the first integral $< \varepsilon/2$, and (i) and (iii) to make the second integral $< \varepsilon/2$.]

7.8 Referring to Theorem 8.5, show that condition (iii) is satisfied by

(a) the exponential family,

(b) continuous densities in which θ is a one-dimensional location or scale parameter.

7.9 A family of functions $\mathcal{F}$ is *equicontinuous at the point* x_0 if, given $\varepsilon > 0$, there exists δ such that $|f(x) - f(x_0)| < \varepsilon$ for all $|x - x_0| < \delta$ and all $f \in \mathcal{F}$. (The same δ works for all f.) The family is *equicontinuous* if it is equicontinuous at each x_0.

Theorem 8.6 *(Communicated by L. Gajek) Consider estimation of θ with loss $L(\theta, \delta)$, where $X \sim f(x|\theta)$ is continuous in θ for each x. If*

(i) The family $L(\theta, \delta(x))$ is equicontinuous in θ for each δ.

(ii) For all $\theta, \theta' \in \Omega$,

$$\sup_x \frac{f(x|\theta')}{f(x|\theta)} < \infty.$$

Then, any finite-valued risk function $R(\theta, \delta) = E_\theta L(\theta, \delta)$ is continuous in θ and, hence, the estimators with finite, continuous risks form a complete class.

(a) Prove Theorem 8.6.

(b) Give an example of an equicontinuous family of loss functions. [*Hint*: Consider squared error loss with a bounded sample space.]

7.10 Referring to Theorem 7.11, this problem shows that the assumption of continuity of $f(x|\theta)$ in θ cannot be relaxed. Consider the density $f(x|\theta)$ that is $N(\theta, 1)$ if $\theta \leq 0$ and $N(\theta + 1, 1)$ if $\theta > 0$.

(a) Show that this density has monotone likelihood ratio, but is not continuous in θ.

(b) Show that there exists a bounded continuous loss function $L(\theta - \delta)$ for which the risk $R(\theta, X)$ is discontinuous.

7.11 For $\mathbf{X} \sim f(\mathbf{x}|\theta)$ and loss function $L(\theta, \delta) = \sum_{i=1}^r \theta_i^m (\theta_i - \delta_i)^2$, show that condition (iii) of Theorem 7.11 holds.

7.12 Prove the following (equivalent) version of Blyth's Method (Theorem 7.13).

Theorem 8.7 *Suppose that the parameter space $\Omega \in \Re^r$ is open, and estimators with continuous risks are a complete class. Let δ be an estimator with a continuous risk function, and let $\{\pi_n\}$ be a sequence of (possibly improper) prior measures such that*

(i) $r(\pi_n, \delta) < \infty$ for all n,

(ii) for any nonempty open set $\Theta_0 \in \Omega$,

$$\frac{r(\pi_n, \delta) - r(\pi_n, \delta^{\pi_n})}{\int_{\Theta_0} \pi_n(\theta)\, d\theta} \to 0 \quad \text{as } n \to \infty.$$

Then, δ is an admissible estimator.

7.13 Fill in some of the gaps in Example 7.14:

(i) Verify the expressions for the posterior expected losses of δ^0 and δ^π in (7.7).

(ii) Show that the normalized beta priors will not satisfy condition (b) of Theorem 7.13, and then verify (7.9).

(iii) Show that the marginal distribution of X is given by (7.10).

(iv) Show that

$$\sum_{x=1}^{\infty} D(x) \le \max\{a^2, b^2\} \sum_{x=1}^{\infty} \frac{1}{x^2} \to 0,$$

and hence that δ^0 is admissible.

7.14 Let $X \sim$ Poisson(λ). Use Blyth's method to show that $\delta^0 = X$ is an admissible estimator of λ under the loss function $L(\lambda, \delta) = (\lambda - \delta)^2$ with the following steps:

(a) Show that the unnormalized gamma priors $\pi_n(\lambda) = \lambda^{a-1}e^{-\lambda/n}$ satisfy condition (b) of Theorem 7.13 by verifying that for any c,

$$\lim_{n\to\infty} \int_0^c \pi_n(\lambda)\, d\lambda = \text{ constant.}$$

Also show that the normalized gamma priors will not work.

(b) Show that under the priors $\pi_n(\lambda)$, the Bayes risks of δ^0 and $\delta^{\pi_n'}$, the Bayes estimator, are given by

$$r(\pi_n', \delta^0) = n^a \Gamma(a),$$
$$r(\pi_n', \delta^{\pi_n'}) = \frac{n}{n+1} n^a \Gamma(a).$$

(c) The difference in risks is

$$r(\pi_n', \delta^0) - r(\pi_n', \delta^{\pi_n'}) = \Gamma(a) n^a \left(1 - \frac{n}{n+1}\right),$$

which, for fixed $a > 0$, goes to infinity as $n \to \infty$ (Too bad!). However, show that if we choose $a = a(n) = 1/\sqrt{n}$, then $\Gamma(a)n^a\left(1 - \frac{n}{n+1}\right) \to 0$ as $n \to \infty$. Thus, the difference in risks goes to zero.

(d) Unfortunately, we must go back and verify condition (b) of Theorem 7.13 for the sequence of priors with $a = 1/\sqrt{n}$, as part (a) no longer applies. Do this, and conclude that $\delta^0(x) = x$ is an admissible estimator of λ.

[*Hint*: For large n, since $t \le c/n$, use Taylor's theorem to write $e^{-t} = 1 - t +$ error, where the error can be ignored.]

(Recall that we have previously considered the admissibility of $\delta^0 = X$ in Corollaries 2.18 and 2.20, where we saw that δ^0 is admissible.)

7.15 Use Blyth's method to establish admissibility in the following situations.

(a) If $X \sim$ Gamma(α, β), α known, then x/α is an admissible estimator of β using the loss function $L(\beta, \delta) = (\beta - \delta)^2/\beta^2$.

(b) If $X \sim$ Negative binomial(k, p), then X is an admissible estimator of $\mu = k(1-p)/p$ using the loss function $L(\mu, \delta) = (\mu - \delta)^2/(\mu + \frac{1}{k}\mu^2)$.

7.16 (i) Show that, in general, if δ^π is the Bayes estimator under squared error loss, then

$$r(\pi, \delta^\pi) - r(\pi, \delta^g) = E\left|\delta^\pi(\mathbf{X}) - \delta^g(\mathbf{X})\right|^2,$$

thus establishing (7.13).

(ii) Prove (7.15).

(iii) Use (7.15) to prove the admissibility of X in one dimension.

7.17 The identity (7.14) can be established in another way. For the situation of Example 7.18, show that

$$r(\pi, \delta^g) = r - 2\int [\nabla \log m_\pi(\mathbf{x})][\nabla \log m_g(\mathbf{x})] m_\pi(\mathbf{x})\, d\mathbf{x} + \int |\nabla \log m_g(\mathbf{x})|^2\, m_\pi(\mathbf{x})\, d\mathbf{x},$$

which implies

$$r(\pi, \delta^\pi) = r - \int |\nabla \log m_\pi(\mathbf{x})|^2\, m_\pi(\mathbf{x})\, d\mathbf{x},$$

and hence deduce (7.14).

7.18 This problem will outline the argument needed to prove Theorem 7.19:

(a) Show that $\nabla m_g(\mathbf{x}) = m_{\nabla g}(\mathbf{x})$, that is,

$$\nabla \int g(\boldsymbol{\theta}) e^{-|\mathbf{x}-\boldsymbol{\theta}|^2}\, d\boldsymbol{\theta} = \int [\nabla g(\boldsymbol{\theta})]\, e^{-|\mathbf{x}-\boldsymbol{\theta}|^2}\, d\boldsymbol{\theta}.$$

(b) Using part (a), show that

$$\begin{aligned} r(\pi, \delta^g) - r(g_n, \delta^{g_n}) &= \int \left|\nabla \log m_g(\mathbf{x}) - \nabla \log m_{g_n}(\mathbf{x})\right|^2 m_{g_n}(\mathbf{x})\, d\mathbf{x} \\ &= \int \left| \frac{\nabla m_g(\mathbf{x})}{m_g(\mathbf{x})} - \frac{\nabla m_{g_n}(\mathbf{x})}{m_{g_n}(\mathbf{x})} \right|^2 m_{g_n}(\mathbf{x})\, d\mathbf{x} \\ &\le 2\int \left| \frac{\nabla m_g(\mathbf{x})}{m_g(\mathbf{x})} - \frac{m_{h_n^2 \nabla g}(\mathbf{x})}{m_{g_n}(\mathbf{x})} \right|^2 m_{g_n}(\mathbf{x})\, d\mathbf{x} \\ &\quad + 2\int \left| \frac{m_{g\nabla h_n^2}(\mathbf{x})}{m_{g_n}(\mathbf{x})} \right|^2 m_{g_n}(\mathbf{x})\, d\mathbf{x} \\ &= B_n + A_n. \end{aligned}$$

(c) Show that

$$A_n = 4\int \left| \frac{m_{gh_n\nabla h_n}(\mathbf{x})}{m_{gh_n^2}(\mathbf{x})} \right|^2 m_{g_n}(\mathbf{x})\, d\mathbf{x} \le 4\int m_{g(\nabla h_n)^2}(\mathbf{x})\, d\mathbf{x}$$

and this last bound $\to 0$ by condition (a).

(d) Show that the integrand of $B_n \to 0$ as $n \to \infty$, and use condition (b) together with the dominated convergence theorem to show $B_n \to 0$, proving the theorem.

7.19 Brown and Hwang (1982) actually prove Theorem 7.19 for the case $f(\mathbf{x}|\boldsymbol{\theta}) = e^{\boldsymbol{\theta}'\mathbf{x}-\psi(\boldsymbol{\theta})}$, where we are interested in estimating $\tau(\boldsymbol{\theta}) = E_\theta(\mathbf{X}) = \nabla\psi(\boldsymbol{\theta})$ under the loss $L(\boldsymbol{\theta}, \delta) = |\tau(\boldsymbol{\theta}) - \delta|^2$. Prove Theorem 7.19 for this case. [The proof is similar to that outlined in Problem 7.18.]

7.20 For the situation of Example 7.20:

(a) Using integration by parts, show that

$$\begin{aligned} \frac{\partial}{\partial x_i} \int g(\boldsymbol{\theta}) e^{-|\mathbf{x}-\boldsymbol{\theta}|^2}\, d\boldsymbol{\theta} &= -\int (x_i - \theta_i) g(\boldsymbol{\theta}) e^{-|\mathbf{x}-\boldsymbol{\theta}|^2}\, d\boldsymbol{\theta} \\ &= \int \left[\frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta})\right] e^{-|\mathbf{x}-\boldsymbol{\theta}|^2}\, d\boldsymbol{\theta} \end{aligned}$$

and hence

$$\frac{\nabla m_g(\mathbf{x})}{m_g(\mathbf{x})} = \frac{\int [\nabla g(\boldsymbol{\theta})] e^{-|\mathbf{x}-\boldsymbol{\theta}|^2} d\boldsymbol{\theta}}{\int g(\boldsymbol{\theta}) e^{-|\mathbf{x}-\boldsymbol{\theta}|^2} d\boldsymbol{\theta}}.$$

(b) Use the Laplace approximation (4.6.33) to show that

$$\frac{\int [\nabla g(\boldsymbol{\theta})] e^{-|\mathbf{x}-\boldsymbol{\theta}|^2} d\boldsymbol{\theta}}{\int g(\boldsymbol{\theta}) e^{-|\mathbf{x}-\boldsymbol{\theta}|^2} d\boldsymbol{\theta}} \approx \frac{\nabla g(\mathbf{x})}{g(\mathbf{x})},$$

and that

$$\delta^g(\mathbf{x}) \approx \mathbf{x} + \frac{\nabla g(\mathbf{x})}{g(\mathbf{x})}.$$

(c) If $g(\boldsymbol{\theta}) = 1/|\boldsymbol{\theta}|^k$, show that

$$\delta^g(\mathbf{x}) \approx \left(1 - \frac{k}{\mathbf{x}^2}\right)\mathbf{x}.$$

7.21 In Example 7.20, if $g(\boldsymbol{\theta}) = 1/|\boldsymbol{\theta}|^k$ is a proper prior, then δ^g is admissible. For what values of k is this the case?

7.22 Verify that the conditions of Theorem 7.19 are satisfied for $g(\boldsymbol{\theta}) = 1/|\boldsymbol{\theta}|^k$ if (a) $k > r - 2$ and (b) $k = r - 2$.

7.23 Establish conditions for the admissibility of Strawderman's estimator (Example 5.6)

(a) using Theorem 7.19,

(b) using the results of Brown (1971), given in Example 7.21.

(c) Give conditions under which Strawderman's estimator is an admissible minimax estimator.

(See Berger 1975, 1976b for generalizations).

7.24 (a) Verify the Laplace approximation of (7.23).

(b) Show that, for $h(|\mathbf{x}|) = k/|x|^{2\alpha}$, (7.25) can be written as (7.26) and that $\alpha = 1$ is needed for an estimator to be both admissible and minimax.

7.25 Theorem 7.17 also applies to the Poisson($\boldsymbol{\lambda}$) case, where Johnstone (1984) obtained the following characterization of admissible estimators for the loss $L(\boldsymbol{\lambda}, \delta) = \sum_{i=1}^r (\lambda_i - \delta_i)^2/\lambda_i$.

A generalized Bayes estimator of the form $\delta(\mathbf{x}) = [1 - h(\Sigma x_i)]\mathbf{x}$ is

(i) *inadmissible* if there exists $\varepsilon > 0$ and $M < \infty$ such that

$$h(\Sigma x_i) < \frac{r - 1 - \varepsilon}{\Sigma x_i} \quad \text{for } \Sigma x_i > M,$$

(ii) *admissible* if $h(\Sigma x_i)(\Sigma x_i)^{1/2}$ is bounded and there exits $M < \infty$ such that

$$h(\Sigma x_i) \geq \frac{r-1}{\Sigma x_i} \quad \text{for } \Sigma x_i > M.$$

(a) Use Johnstone's characterization of admissible Poisson estimators (Example 7.22) to find an admissible Clevenson-Zidek estimator (6.31).

(b) Determine conditions under which the estimator is both admissible and minimax.

7.26 For the situation of Example 7.23:

(a) Show that X/n and $(n/n+1)(X/n)(1 - X/n)$ are admissible for estimating p and $p(1-p)$, respectively.

(b) Show that $\alpha(X/n) + (1 - \alpha)(a/(a + b))$ is an admissible estimator of p, where $\alpha = n/(n + a + b)$. Compare the results here to that of Theorem 2.14 (Karlin's theorem). [Note that the results of Diaconis and Ylvisaker (1979) imply that $\pi(\cdot) =$ uniform are the only priors that give linear Bayes estimators.]

7.27 Fill in the gaps in the proof that estimators δ^π of the form (7.27) are a complete class.

(a) Show that δ^π is admissible when $r = -1, s = n + 1$, and $r + 1 = s$.

(b) For any other estimator $\delta'(x)$ for which $\delta'(x) = h(0)$ for $x \leq r'$ and $\delta'(x) = h(1)$ for $x \geq s'$, show that we must have $r' \geq r$ and $s' \leq s$.

(c) Show that $R(p, \delta') \leq R(p, \delta^\pi)$ for all $p \in [0, 1]$ if and only if $R_{r,s}(p, \delta') \leq R_{r,s}(p, \delta^\pi)$ for all $p \in [0, 1]$.

(d) Show that $\int_0^1 R_{r,s}(p, \delta)k(p)d\pi(p)$ is uniquely minimized by $[\delta^\pi(r+1), \ldots, \delta^\pi(s - 1)]$, and hence deduce the admissibility of δ^π.

(e) Use Theorem 7.17 to show that any admissible estimator of $h(p)$ is of the form (7.27), and hence that (7.27) is a minimal complete class.

7.28 For $i = 1, 2, \ldots, k$, let $X_i \sim f_i(x|\theta_i)$ and suppose that $\delta_i^*(x_i)$ is a unique Bayes estimator of θ_i under the loss $L_i(\theta_i, \delta)$, where L_i satisfies $L_i(a, a) = 0$ and $L_i(a, a') > 0, \quad a \neq a'$. Suppose that for some $j, 1 \leq j \leq k$, there is a value θ^* such that if $\theta_j = \theta^*$,

(i) $X_j = x^*$ with probability 1,

(ii) $\delta_j^*(x^*) = \theta^*$.

Show that $(\delta_1^*(x_1), \delta_2^*(x_2), \ldots, \delta_k^*(x_k))$ is admissible for $(\theta_1, \theta_2, \cdots, \theta_k)$ under the loss $\sum_i L_i(\theta_i, \delta)$; that is, there is no Stein effect.

7.29 Suppose we observe $X_1, X_2, \ldots$ sequentially, where $X_i \sim f_i(x|\theta_i)$. An estimator of $\boldsymbol{\theta}_j = (\theta_1, \theta_2, \ldots, \theta_j)$ is called *nonanticipative* (Gutmann 1982b) if it only depends on $(X_1, X_2, \ldots, X_j)$. That is, we cannot use information that comes later, with indices $> j$. If $\delta_i^*(x_i)$ is an admissible estimator of θ_i, show that it cannot be dominated by a nonanticipative estimator. Thus, this is again a situation in which there is no Stein effect.

[*Hint*: It is sufficient to consider $j = 2$. An argument similar to that of Example 7.24 will work.]

7.30 For $X \sim N_r(\boldsymbol{\theta}, I)$, consider estimation of $\varphi'\boldsymbol{\theta}$ where $\varphi_{r\times 1}$ is known, using the estimator $a'X$ with loss function $L(\varphi'\boldsymbol{\theta}, \delta) = (\varphi'\boldsymbol{\theta} - \delta)^2$.

(a) Show that if a lies outside the sphere (7.31), then $a'X$ is inadmissible.

(b) Show that the Bayes estimator of $\varphi'\boldsymbol{\theta}$ against the prior $\boldsymbol{\theta} \sim N(0, V)$ is given by

$$E(\varphi'\boldsymbol{\theta}|\mathbf{x}) = (I + V)^{-1}\varphi\mathbf{x}.$$

(c) Find a covariance matrix V such that $E(\varphi'\boldsymbol{\theta}|\mathbf{x})$ lies inside the sphere (7.31) [V will be of rank one, hence of the form vv' for some $r \times 1$ vector v].

Parts (a)–(c) show that all linear estimators inside the sphere (7.31) are admissible, and those outside are inadmissible. It remains to consider the boundary, which is slightly more involved. See Cohen 1966 for details.

7.31 *Brown's ancillarity paradox.* Let $\mathbf{X} \sim N_r(\mu, I), r > 2$, and consider the estimation of $\mathbf{w}'\mu = \Sigma_{i=1}^r w_i\mu_i$, where $\mathbf{w}$ is a known vector with $\Sigma w_i^2 > 0$, using loss function $L(\mu, d) = (\mathbf{w}'\mu - \mathbf{w}'d)^2$.

(a) Show that the estimator $\mathbf{w}'\mathbf{X}$ is minimax and admissible.

(b) Assume no that $\mathbf{w}$ is the realized value of a random variable $\mathbf{W}$, with distribution independent of $\mathbf{X}$, where $\mathbf{V} = E(\mathbf{W}'\mathbf{W})$ is known. Show that the estimator $\mathbf{w}'d^*$, where

$$d^*(\mathbf{x}) = \left(I - \frac{c\mathbf{V}^{-1}}{\mathbf{x}\mathbf{V}^{-1}\mathbf{x}}\right)\mathbf{x},$$

with $0 < c < 2(r-2)$, dominates $\mathbf{w}'\mathbf{X}$ in risk.

[*Hint*: Establish and use the fact that $E[L(\mu, d)] = E[(d-\mu)'\mathbf{V}(d-\mu)]$. This is a special case of results established by Brown (1990a). It is referred to as a *paradox* because the distribution of the ancillary, which should not affect the estimation of μ, has an enormous effect on the properties of the standard estimator. Brown showed how these results affect the properties of coefficient estimates in multiple regression when the assumption of random regressors is made. In that context, the ancillarity paradox also relates to Shaffer's (1991) work on best linear unbiased estimation (see Theorem 3.4.14 and Problems 3.4.16-3.4.18.]

7.32 Efron (1990), in a discussion of Brown's (1990a) ancillarity paradox, proposed an alternate version.

Suppose $\mathbf{X} \sim N_r(\mu, I)$, $r > 2$, and with probability $1/r$, independent of $\mathbf{X}$, the value of the random variable $J = j$ is observed, $j = 1, 2, \ldots, r$. The problem is to estimate θ_j using the loss function $L(\theta_j, d) = (\theta_j - d)^2$. Show that, conditional on $J = j$, X_j is a minimax and admissible estimator of θ_j. However, unconditionally, X_j is dominated by the jth coordinate of the James-Stein estimator. This version of the paradox may be somewhat more transparent. It more clearly shows how the presence of the ancillary random variable forces the problem to be considered as a multivariate problem, opening the door for the Stein effect.

9 Notes

9.1 History

Deliberate efforts to develop statistical inference and decision making not based on "inverse probability" (i.e., without assuming prior distributions) were mounted by R.A. Fisher (for example, 1922, 1930, and 1935; see also Lane 1980), by Neyman and Pearson (for example, 1933ab), and by Wald (1950). The latter's general decision theory introduced, as central notions, the minimax principle and least favorable distributions in close parallel to the corresponding concepts of the theory of games. Many of the examples of Section 5.2 were first worked out by Hodges and Lehmann (1950). Admissibility is another basic concept of Wald's decision theory. The admissibility proofs in Example 2.8 are due to Blyth (1951) and Hodges and Lehmann (1951). A general necessary and sufficient condition for admissibility was obtained by Stein (1955). Theorem 2.14 is due to Karlin (1958), and the surprising inadmissibility results of Section 5.5 had their origin in Stein's seminal paper (1956b). The relationship between equivariance and the minimax property was foreshadowed in Wald (1939) and was developed for point estimation by Peisakoff (1950), Girshick and Savage (1951), Blackwell and Girshick (1954), Kudo (1955), and Kiefer (1957).

Characterizations of admissible estimators and complete classes have included techniques such as Blyth's method and the information inequality. The pathbreaking paper of Brown (1971) was influential in shaping the understanding of admissibility problems, and motivated further study of differential inequalities (Brown 1979, 1988) and associated stochastic processes and Markov chains (Brown 1971, Johnstone 1984, Eaton 1992).

Combining this fact with (8.26), we see that the right side of (8.28) is $\leq C'e^{-t^2 I(\theta_0)/4}$ for all t satisfying (ii), with probability arbitrarily close to 1, and this establishes (8.27).

(iii) $|t| \geq \delta\sqrt{n}$. As in (ii), the second term in the integrand of (8.13) can be neglected, and it is enough to show that for all δ,

$$\int_{|t|\geq\delta\sqrt{n}} \exp[\omega(t)]\pi\left(T_n + \frac{t}{\sqrt{n}}\right) dt$$

$$= \sqrt{n}\int_{|\theta-T_n|\geq\delta} \pi(\theta)\exp\left\{l(\theta) - l(\theta_0) - \frac{1}{2nI(\theta_0)}[l'(\theta_0)]^2\right\} d\theta \xrightarrow{P} 0. \tag{8.29}$$

From (8.24) and (B3), it is seen that given δ, there exists ε such that

$$\sup_{|\theta-T_n|\geq\delta} e^{[l(\theta)-l(\theta_0)]} \leq e^{-n\varepsilon}$$

with probability tending to 1. By (8.26), the right side of (8.29) is therefore bounded above by

$$C\sqrt{n}\,e^{-n\varepsilon}\int \pi(\theta)\,d\theta = C\sqrt{n}\,e^{-n\varepsilon} \tag{8.30}$$

with probability tending to 1, and this completes the proof of (iii).

To prove (8.13), let us now combine (i)-(iii). Given $\varepsilon > 0$ and $\delta > 0$, choose M so large that

$$\int_M^\infty \left[C\exp\left[-\frac{t^2}{2}I(\theta_0)\right] + \exp\left[-\frac{t^2}{2}I(\theta_0)\right]\pi(\theta_0)\right] dt \leq \frac{\varepsilon}{3}, \tag{8.31}$$

and, hence, that for sufficiently large n, the integral (8.13) over (ii) is $\leq \varepsilon/3$ with probability $\geq 1-\varepsilon$. Next, choose n so large that the integrals (8.13) over (i) and over (iii) are also $\leq \varepsilon/3$ with probability $\geq 1-\varepsilon$. Then, $P[J_1 \leq \varepsilon] \geq 1-3\varepsilon$, and this completes the proof of (8.13).

The proof for J_1' requires only trivial changes. In part (i), the factor $[1+|t|]$ is bounded, so that the proof continues to apply. In part (ii), multiplication of the integrand of (8.31) by $[1+|t|]$ does not affect its integrability, and the proof goes through as before. Finally, in part (iii), the integral in (8.30) must be replaced by $Cne^{-n\varepsilon}\int|\theta|\pi(\theta)\,d\theta$, which is finite by (B5).

9 Problems

Section 1

1.1 Let $X_1, \ldots, X_n$ be iid with $E(X_i) = \xi$.

(a) If the X_is have a finite fourth moment, establish (1.3)

(b) For k a positive integer, show that $E(\bar{X}-\xi)^{2k-1}$ and $E(\bar{X}-\xi)^{2k}$, if they exist, are both $O(1/n^k)$.

[*Hint*: Without loss of generality, let $\xi = 0$ and note that $E(X_{i_1}^{r_1}X_{i_2}^{r_2}\cdots) = 0$ if any of the r's is equal to 1.]

1.2 For fixed n, describe the relative error in Example 1.3 as a function of p.

1.3 Prove Theorem 1.5.

1.4 Let $X_1, \ldots, X_n$ be iid as $N(\xi, \sigma^2)$, σ^2 known, and let $g(\xi) = \xi^r$, $r = 2, 3, 4$. Determine, up to terms of order $1/n$,

(a) the variance of the UMVU estimator of $g(\xi)$;

(b) the bias of the MLE of $g(\xi)$.

1.5 Let $X_1, \ldots, X_n$ be iid as $N(\xi, \sigma^2)$, ξ known. For even r, determine the variance of the UMVU estimator (2.2.4) of σ^r up to terms of order r.

1.6 Solve the preceding problem for the case that ξ is unknown.

1.7 For estimating p^m in Example 3.3.1, determine, up to order $1/n$,

(a) the variance of the UMVU estimator (2.3.2);

(b) the bias of the MLE.

1.8 Solve the preceding problem if p^m is replaced by the estimand of Problem 2.3.3.

1.9 Let $X_1, \ldots, X_n$ be iid as Poisson $P(\theta)$.

(a) Determine the UMVU estimator of $P(X_i = 0) = e^{-\theta}$.

(b) Calculate the variance of the estimator of (a) up to terms of order $1/n$.

[*Hint*: Write the estimator in the form (1.15) where $h(\bar{X})$ is the MLE of $e^{-\theta}$.]

1.10 Solve part (b) of the preceding problem for the estimator (2.3.22).

1.11 Under the assumptions of Problem 1.1, show that $E|\bar{X} - \xi|^{2k-1} = O(n^{-k+1/2})$. [*Hint*: Use the fact that $E|\bar{X} - \xi|^{2k-1} \leq [E(\bar{X} - \xi)^{4k-2}]^{1/2}$ together with the result of Problem 1.1.]

1.12 Obtain a variant of Theorem 1.1, which requires existence and boundedness of only h''' instead of $h^{(iv)}$, but where R_n is only $O(n^{-3/2})$.

[*Hint*: Carry the expansion (1.6) only to the second instead of the third derivative, and apply Problem 1.11.]

1.13 To see that Theorem 1.1 is not necessarily valid without boundedness of the fourth (or some higher) derivative, suppose that the X's are distributed as $N(\xi, \sigma^2)$ and let $h(X) = e^{x^4}$. Then, all moments of the X's and all derivatives of h exist.

(a Show that the expectation of $h(\bar{X})$ does not exist for any n, and hence that $E\{\sqrt{n}[h(\bar{X}) - h(\xi)]\}^2 = \infty$ for all values of n.

(b On the other hand, show that $\sqrt{n}\,[h(\bar{X}) - h(\xi)]$ has a normal limit distribution with finite variance, and determine that variance.

1.14 Let $X_1, \ldots, X_n$ be iid from the exponential distribution with density $(1/\theta)e^{-x/\theta}$, $x > 0$, and $\theta > 0$.

(a) Use Theorem 1.1 to find approximations to $E(\sqrt{\bar{X}})$ and $\text{var}(\sqrt{\bar{X}})$.

(b) Verify the exact calculation

$$\text{var}(\sqrt{\bar{X}}) = \left[1 - \frac{1}{n}\left(\frac{\Gamma(n+1/2)}{\Gamma(n)}\right)^2\right]\theta$$

and show that $\lim_{n\to\infty} n\,\text{var}(\sqrt{\bar{X}}) = \theta/4$.

(c) Reconcile the results in parts (a) and (b). Explain why, even though Theorem 1.1 did not apply, it gave the correct answer.

(d) Show that a similar conclusion holds for $h(x) = 1/x$.

[*Hint*: For part (b), use the fact that $T = \Sigma X_i$ has a gamma distribution. The limit can be evaluated with Stirling's formula. It can also be evaluated with a computer algebra program.]

1.15 Let $X_1, \ldots, X_n$ be iid according to $U(0, \theta)$. Determine the variance of the UMVU estimator of θ^k, where k is an integer, $k > -n$.

1.16 Under the assumptions of Problem 1.15, find the MLE of θ^k and compare its expected squared error with the variance of the UMVU estimator.

1.17 Let $X_1, \ldots, X_n$ be iid according to $U(0, \theta)$, let $T = \max(X_1, \ldots, X_n)$, and let h be a function satisfying the conditions of Theorem 1.1. Show that

$$E[h(T)] = h(\theta) - \frac{\theta}{n}h'(\theta) + \frac{1}{n^2}[\theta h'(\theta) + \theta^2 h''(\theta)] + O\left(\frac{1}{n^3}\right)$$

and

$$\operatorname{var}[h(T)] = \frac{\theta^2}{n^2}[h'(\theta)]^2 + O\left(\frac{1}{n^3}\right).$$

1.18 Apply the results of Problem 1.17 to obtain approximate answers to Problems 1.15 and 1.16, and compare the answers with the exact solutions.

1.19 If the X's are as in Theorem 1.1 and if the first five derivatives of h exist and the fifth derivative is bounded, show that

$$E[h(\bar{X})] = h(\xi) + \frac{1}{2}h''\frac{\sigma^2}{n} + \frac{1}{24n^2}[4h'''\mu_3 + 3h^{(iv)}\sigma^4] + O(n^{-5/2})$$

and if the fifth derivative of h^2 is also bounded

$$\operatorname{var}[h(\bar{X})] = (h'^2)\frac{\sigma^2}{n} + \frac{1}{n^2}[h'h''\mu_3 + (h'h''' + \frac{1}{2}h''^2)\sigma^4] + O(n^{-5/2})$$

where $\mu_3 = E(X - \xi)^3$.

[*Hint*: Use the facts that $E(\bar{X} - \xi)^3 = \mu_3/n^2$ and $E(\bar{X} - \xi)^4 = 3\sigma^4/n^2 + O(1/n^3)$.]

1.20 Under the assumptions of the preceding problem, carry the calculation of the variance (1.16) to terms of order $1/n^2$, and compare the result with that of the preceding problem.

1.21 Carry the calculation of Problem 1.4 to terms of order $1/n^2$.

1.22 For the estimands of Problem 1.4, calculate the expected squared error of the MLE to terms of order $1/n^2$, and compare it with the variance calculated in Problem 1.21.

1.23 Calculate the variance (1.18) to terms of order $1/n^2$ and compare it with the expected squared error of the MLE carried to the same order.

1.24 Find the variance of the estimator (2.3.17) up to terms of the order $1/n^3$.

1.25 For the situation of Example 1.12, show that the UMVU estimator δ_{1n} is the bias-corrected MLE, where the MLE is δ_{3n}.

1.26 For the estimators of Example 1.13:

(a) Calculate their exact variances.

(b) Use the result of part (a) to verify (1.27).

1.27 (a) Under the assumptions of Theorem 1.5, if all fourth moments of the $X_{i\nu}$ are finite, show that $E(\bar{X}_i - \xi_i)(\bar{X}_j - \xi_j) = \sigma_{ij}/n$ and that all third and fourth moments $E(\bar{X}_i - \xi_i)(\bar{X}_j - \xi_j)(\bar{X}_k - \xi_k)$, and so on are of the order $1/n^2$.

(b) If, in addition, all derivatives of h of total order ≤ 4 exist and those of order 4 are uniformly bounded, then

$$E[h(\bar{X}_1, \ldots, \bar{X}_s)] = h(\xi_1, \ldots, \xi_s) + \frac{1}{2n} \sum_{i=1}^{s} \sum_{j=1}^{s} \sigma_{ij} \frac{\partial^2 h(\xi_1, \ldots, \xi_s)}{\partial \xi_i \partial \xi_j} + R_n,$$

and if the derivatives of h^2 of order 4 are also bounded,

$$\operatorname{var}[h(\bar{X}_1, \ldots, \bar{X}_s)] = \frac{1}{n} \Sigma \Sigma \sigma_{ij} \frac{\partial h}{\partial \xi_i} \frac{\partial h}{\partial \xi_j} + R_n$$

where the remainder R_n in both cases is $O(1/n^2)$.

1.28 On the basis of a sample from $N(\xi, \sigma^2)$, let $P_n(\xi, \sigma)$ be the probability that the UMVU estimator $\bar{X}^2 - \sigma^2/n$ of ξ^2 (σ known) is negative.

(a) Show that $P_n(\xi, \sigma)$ is a decreasing function of $\sqrt{n}\,|\xi|/\sigma$.

(b) Show that $P_n(\xi, \sigma) \to 0$ as $n \to \infty$ for any fixed $\xi \neq 0$ and σ.

(c) Determine the value of $P_n(0, \sigma)$.

[*Hint*: $P_n(\xi, \sigma) = P[-1 - \sqrt{n}\,\xi/\sigma < Y < 1 - \sqrt{n}\,\xi/\sigma]$, where $Y = \sqrt{n}\,(\bar{X} - \xi)/\sigma$ is distributed as $N(0, 1)$.]

1.29 Use the t-distribution to find the value of $P_n(0, \sigma)$ in the preceding problem for the UMVU estimator of ξ^2 when σ is unknown for representative values of n.

1.30 Fill in the details of the proof of Theorem 1.9. (See also Problem 1.8.8.)

1.31 In Example 8.13 with $\theta = 0$, show that δ_{2n} is not exactly distributed as $\sigma^2(\chi_1^2 - 1)/n$.

1.32 In Example 8.13, let $\delta_{4n} = \max(0, \bar{X}^2 - \sigma^2/n)$, which is an improvement over δ_{1n}.

(a) Show that $\sqrt{n}(\delta_{4n} - \theta^2)$ has the same limit distribution as $\sqrt{n}(\delta_{1n} - \theta^2)$ when $\theta \neq 0$.

(b) Describe the limit distribution of $n\delta_{4n}$ when $\theta = 0$.

[*Hint*: Write $\delta_{4n} = \delta_{1n} + R_n$ and study the behavior of R_n.]

1.33 Let X have the binomial distribution $b(p, n)$, and let $g(p) = pq$. The UMVU estimator of $g(p)$ is $\delta = X(n - X)/n(n-1)$. Determine the limit distribution of $\sqrt{n}(\delta - pq)$ and $n(\delta - pq)$ when $g'(p) \neq 0$ and $g'(p) = 0$, respectively.

[*Hint*: Consider first the limit behavior of $\delta' = X(n - X)/n^2$.]

1.34 Let $X_1, \ldots, X_n$ be iid as $N(\xi, 1)$. Determine the limit behavior of the distribution of the UMVU estimator of $p = P[|X_i| \leq u]$.

1.35 Determine the limit behavior of the estimator (2.3.22) as $n \to \infty$.

[*Hint*: Consider first the distribution of $\log \delta(T)$.]

1.36 Let $X_1, \ldots, X_n$ be iid with distribution P_θ, and suppose δ_n is UMVU for estimating $g(\theta)$ on the basis of $X_1, \ldots, X_n$. If there exists n_0 and an unbiased estimator $\delta_0(X_1, \ldots, X_{n_0})$ which has finite variance for all θ, then δ_n is consistent for $g(\theta)$.

[*Hint*: For $n = kn_0$ (with k an integer), compare δ_n with the estimator

$$\frac{1}{k}\{\delta_0(X_1, \ldots, X_{n_0}) + \delta_0(X_{n_0+1}, \ldots, X_{2n_0}) + \ldots\}].$$

1.37 Let Y_n be distributed as $N(0, 1)$ with probability π_n and as $N(0, \tau_n^2)$ with probability $1 - \pi_n$. If $\tau_n \to \infty$ and $\pi_n \to \pi$, determine for what values of π the sequence $\{Y_n\}$ does and does not have a limit distribution.

1.38 (a) In Problem 1.37, determine to what values $\operatorname{var}(Y_n)$ can tend as $n \to \infty$ if $\pi_n \to 1$ and $\tau_n \to \infty$ but otherwise both are arbitrary.

(b) Use (a) to show that the limit of the variance need not agree with the variance of the limit distribution.

1.39 Let $b_{m,n}$, $m, n = 1, 2, \ldots$, be a double sequence of real numbers, which for each fixed m is nondecreasing in n. Show that $\lim_{n\to\infty} \lim_{m\to\infty} b_{m,n} = \lim_{m,n\to\infty} \inf b_{m,n}$ and $\lim_{m\to\infty} \lim_{n\to\infty} b_{m,n} = \lim_{m,n\to\infty} \sup b_{m,n}$ provided the indicated limits exist (they may be infinite) and where $\lim\inf b_{m,n}$ and $\lim\sup b_{m,n}$ denote, respectively, the smallest and the largest limit points attainable by a sequence b_{m_k,n_k}, $k = 1, 2, \ldots$, with $m_k \to \infty$ and $n_k \to \infty$.

Section 2

2.1 Let $X_1, \ldots, X_n$ be iid as $N(0, 1)$. Consider the two estimators

$$T_n = \begin{cases} \bar{X}_n & \text{if } S_n \le a_n \\ n & \text{if } S_n > a_n, \end{cases}$$

where $S_n = \Sigma(X_i - \bar{X})^2$, $P(S_n > a_n) = 1/n$, and $T_n' = (X_1 + \cdots + X_{k_n})/k_n$ with k_n the largest integer $\le \sqrt{n}$.

(a) Show that the asymptotic efficiency of T_n' relative to T_n is zero.

(b) Show that for any *fixed* $\varepsilon > 0$, $P[|T_n - \theta| > \varepsilon] = \frac{1}{n} + o\left(\frac{1}{n}\right)$, but $P[|T_n' - \theta| > \varepsilon] = o\left(\frac{1}{n}\right)$.

(c) For large values of n, what can you say about the two probabilities in part (b) when ε is replaced by $a/\sqrt{n}$? (Basu 1956).

2.2 If $k_n[\delta_n - g(\theta)] \xrightarrow{\mathcal{L}} H$ for some sequence k_n, show that the same result holds if k_n is replaced by k_n', where $k_n/k_n' \to 1$.

2.3 Assume that the distribution of $Y_n = \sqrt{n}(\delta_n - g(\theta))$ converges to a distribution with mean 0 and variance $v(\theta)$. Use Fatou's lemma (Lemma 1.2.6) to establish that $\text{var}_\theta(\delta_n) \to 0$ for all θ.

2.4 If $X_1, \ldots, X_n$ are a sample from a one-parameter exponential family (1.5.2), then $\Sigma T(X_i)$ is minimal sufficient and $E[(1/n)\Sigma T(X_i)] = (\partial/\partial\eta)A(\eta) = \tau$. Show that for any function $g(\cdot)$ for which Theorem 1.8.12 holds, $g((1/n)\Sigma T(X_i))$ is asymptotically unbiased for $g(\tau)$.

2.5 If $X_1, \ldots, X_n$ are iid $n(\mu, \sigma^2)$, show that $S^r = [1/(n-1)\Sigma(x_i - \bar{x})^2]^{r/2}$ is an asymptotically unbiased estimator of σ^r.

2.6 Let $X_1, \ldots, X_n$ be iid as $U(0, \theta)$. From Example 2.1.14, $\delta_n = (n+1)X_{(n)}/n$ is the UMVU estimator of θ, whereas the MLE is $X_{(n)}$. Determine the limit distribution of (a) $n[\theta - \delta_n]$ and (b) $n[\theta - X_{(n)}]$. Comment on the asymptotic bias of these estimators. [*Hint*: $P(X_{(n)} \le y) = y^n/\theta^n$ for any $0 < y < \theta$.]

2.7 For the situation of Problem 2.6:

(a) Calculate the mean squared errors of both δ_n and $X_{(n)}$ as estimators of θ.

(b) Show

$$\lim_{n\to\infty} \frac{E(X_{(n)} - \theta)^2}{E(\delta_n - \theta)^2} = 2.$$

2.8 Verify the asymptotic distribution claimed for δ_n in Example 2.5.

2.9 Let δ_n be any estimator satisfying (2.2) with $g(\theta) = \theta$. Construct a sequence δ_n' such that $\sqrt{n}(\delta_n' - \theta) \xrightarrow{\mathcal{L}} N[0, w^2(\theta)]$ with $w(\theta) = v(\theta)$ for $\theta \ne \theta_0$ and $w(\theta_0) = 0$.

2.10 In the preceding problem, construct δ_n' such that $w(\theta) = v(\theta)$ for all $\theta \neq \theta_0$ and θ_1 and $< v(\theta)$ for $\theta = \theta_0$ and θ_1.

2.11 Construct a sequence $\{\delta_n\}$ satisfying (2.2) but for which the bias $b_n(\theta)$ does not tend to zero.

2.12 In Example 2.7 with $R_n(\theta)$ given by (2.11), show that $R_n(\theta) \to 1$ for $\theta \neq 0$ and that $R_n(0) \to a^2$.

2.13 Let $b_n(\theta) = E_\theta(\delta_n) - \theta$ be the bias of the estimator δ_n of Example 2.5.

(a) Show that

$$b_n(\theta) = \frac{-(1-a)}{\sqrt{n}} \int_{-\sqrt[4]{n}}^{\sqrt[4]{n}} x\phi(x - \sqrt{n}\theta)\,dx;$$

(b) Show that $b_n'(\theta) \to 0$ for any $\theta \neq 0$ and $b_n'(0) \to (1-a)$.

(c) Use (b) to explain how the Hodges estimator δ_n can violate (2.7) without violating the information inequality.

2.14 In Example 2.7, show that if $\theta_n = c/\sqrt{n}$, then $R_n(\theta_n) \to a^2 + c^2(1-a)^2$.

Section 3

3.1 Let X have the binomial distribution $b(p, n), 0 \leq p \leq 1$. Determine the MLE of p

(a) by the usual calculus method determining the maximum of a function;

(b) by showing that $p^x q^{n-x} \leq (x/n)^x[(n-x)/n]^{n-x}$.

[*Hint*: (b) Apply the fact that the geometric mean is equal to or less than the arithmetic mean to n numbers of which x are equal to np/x and $n - x$ equal to $nq/(n-x)$.]

3.2 In the preceding problem, show that the MLE does not exist when p is restricted to $0 < p < 1$ and when $x = 0$ or $= n$.

3.3 Let $X_1, \ldots, X_n$ be iid according to $N(\xi, \sigma^2)$. Determine the MLE of (a) ξ when σ is known, (b) σ when ξ is known, and (c) (ξ, σ) when both are unknown.

3.4 Suppose $X_1, \ldots, X_n$ are iid as $N(\xi, 1)$ with $\xi > 0$. Show that the MLE is $\bar{X}$ when $\bar{X} > 0$ and does not exist when $\bar{X} \leq 0$.

3.5 Let X take on the values 0 and 1 with probabilities p and q, respectively. When it is known that $1/3 \leq p \leq 2/3$, (a) find the MLE and (b) show that the expected squared error of the MLE is uniformly larger than that of $\delta(x) = 1/2$.

[A similar estimation problem arises in *randomized response* surveys. See Example 5.2.2.]

3.6 When Ω is finite, show that the MLE is consistent if and only if it satisfies (3.2).

3.7 Show that Theorem 3.2 remains valid if assumption A1 is relaxed to A1$'$: There is a nonempty set $\Omega_0 \in \Omega$ such that $\theta_0 \in \Omega_0$ and Ω_0 is contained in the support of each P_θ.

3.8 Prove the existence of unique $0 < a_k < a_{k-1}, k = 1, 2, \ldots,$ satisfying (3.4).

3.9 Prove (3.9).

3.10 In Example 3.6 with $0 < c < 1/2$, determine a consistent estimator of k.

[*Hint*: (a) The smallest value K of j for which I_j contains at least as many of the X's as any other I is consistent. (b) The value of j for which I_j contains the median of the X's is consistent since the median of f_k is in I_k.]

3.11 Verify the nature of the roots in Example 3.9.

3.12 Let X be distributed as $N(\theta, 1)$. Show that conditionally given $a < X < b$, the variable X tends in probability to b as $\theta \to \infty$.

3.13 Consider a sample $X_1, \ldots, X_n$ from a Poisson distribution conditioned to be positive, so that $P(X_i = x) = \theta^x e^{-\theta}/x!(1 - e^{-\theta})$ for $x = 1, 2, \ldots$. Show that the likelihood equation has a unique root for all values of x.

3.14 Let X have the negative binomial distribution (2.3.3). Find an ELE of p.

3.15 (a) A density function is *strongly unimodal*, or equivalently *log concave*, if $\log f(x)$ is a concave function. Show that such a density function has a unique mode.

(b) Let $X_1, \ldots, X_n$ be iid with density $f(x - \theta)$. Show that the likelihood function has a unique root if $f'(x)/f(x)$ is monotone, and the root is a maximum if $f'(x)/f(x)$ is decreasing. Hence, densities that are log concave yield unique MLEs.

(c) Let $X_1, \ldots, X_n$ be positive random variables (or symmetrically distributed about zero) with joint density $a^n \Pi f(ax_i)$, $a > 0$. Show that the likelihood equation has a unique maximum if $xf'(x)/f(x)$ is strictly decreasing for $x > 0$.

(d) If $X_1, \ldots, X_n$ are iid with density $f(x_i - \theta)$ where f is unimodal and if the likelihood equation has a unique root, show that the likelihood equation also has a unique root when the density of each X_i is $af[a(x_i - \theta)]$, with a known.

3.16 For each of the following densities, $f(\cdot)$, determine if (a) it is strongly unimodal and (b) $xf'(x)/f(x)$ is strictly decreasing for $x > 0$. Hence, comment on whether the respective location and scale parameters have unique MLEs:

(a) $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty$ (normal)

(b) $f(x) = \dfrac{1}{\sqrt{2\pi}} \dfrac{1}{x} e^{-\frac{1}{2}(\log x)^2}, \quad 0 \le x < \infty$ (lognormal)

(c) $f(x) = e^{-x}/(1 + e^{-x})^2, \quad -\infty < x < \infty$ (logistic)

(d) $f(x) = \dfrac{\Gamma(\nu + 1/2)}{\Gamma(\nu/2)} \dfrac{1}{\sqrt{\nu\pi}} \dfrac{1}{[1 + (x/\nu)^2]^{\frac{\nu+1}{2}}}, \quad -\infty < x < \infty$ (t with ν df)

3.17 If $X_1, \ldots, X_n$ are iid with density $f(x_i - \theta)$ or $af(ax_i)$ and f is the logistic density $L(0, 1)$, the likelihood equation has unique solutions $\hat{\theta}$ and $\hat{a}$ both in the location and the scale case. Determine the limit distribution of $\sqrt{n}(\hat{\theta} - \theta)$ and $\sqrt{n}(\hat{a} - a)$.

3.18 In Problem 3.15(b), with f the Cauchy density $C(0, a)$, the likelihood equation has a unique root $\hat{a}$ and $\sqrt{n}(\hat{a} - a) \xrightarrow{\mathcal{L}} N(0, 2a^2)$.

3.19 If $X_1, \ldots, X_n$ are iid as $C(\theta, 1)$, then for any fixed n there is positive probability (a) that the likelihood equation has $2n - 1$ roots and (b) that the likelihood equation has a unique root.

[*Hint*: (a) If the x's are sufficiently widely separated, the value of $L'(\theta)$ in the neighborhood of x_i is dominated by the term $(x_i - \theta)/[1 + (x_i - \theta)^2]$. As θ passes through x_i, this term changes signs so that the log likelihood has a local maximum near x_i. (b) Let the x's be close together.]

3.20 If $X_1, \ldots, X_n$ are iid according to the gamma distribution $\Gamma(\theta, 1)$, the likelihood equation has a unique root.

[*Hint*: Use Example 3.12. Alternatively, write down the likelihood and use the fact that $\Gamma'(\theta)/\Gamma(\theta)$ is an increasing function of θ.]

3.21 Let $X_1, \ldots, X_n$ be iid according to a Weibull distribution with density

$$f_\theta(x) = \theta x^{\theta-1} e^{-x^\theta}, \quad x > 0, \theta > 0,$$

which is not a member of the exponential, location, or scale family. Nevertheless, show that there is a unique interior maximum of the likelihood function.

3.22 Under the assumptions of Theorem 3.2, show that

$$\left[L\left(\theta_0 + \frac{1}{\sqrt{n}}\right) - L(\theta_0) + \frac{1}{2}I(\theta_0)\right] / \sqrt{I(\theta_0)}$$

tends in law to $N(0, 1)$.

3.23 Let $X_1, \ldots, X_n$ be iid according to $N(\theta, a\theta^2)$, $\theta > 0$, where a is a known positive constant.

(a) Find an explicit expression for an ELE of θ.

(b) Determine whether there exists an MRE estimator under a suitable group of transformations.

[This case was considered by Berk (1972).]

3.24 Check that the assumptions of Theorem 3.10 are satisfied in Example 3.12.

3.25 For $X_1, \ldots, X_n$ iid as $DE(\theta, 1)$, show that (a) the sample median is an MLE of θ and (b) the sample median is asymptotically normal with variance $1/n$, the information inequality bound.

3.26 In Example 3.12, show directly that $(1/n)\Sigma T(X_i)$ is an asymptotically efficient estimator of $\theta = E_\eta[T(X)]$ by considering its limit distribution.

3.27 Let $X_1, \ldots, X_n$ be iid according to $\theta g(x) + (1-\theta)h(x)$, where (g, h) is a pair of specified probability densities with respect to μ, and where $0 < \theta < 1$.

(a) Give one example of (g, h) for which the assumptions of Theorem 3.10 are satisfied and one for which they are not.

(b) Discuss the existence and nature of the roots of the likelihood equation for $n = 1, 2, 3$.

3.28 Under the assumptions of Theorem 3.7, suppose that $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are two consistent sequences of roots of the likelihood equation. Prove that $P_{\theta_0}(\hat{\theta}_{1n} = \hat{\theta}_{2n}) \to 1$ as $n \to \infty$.

[*Hint*:

(a) Let $S_n = \{\mathbf{x} : \mathbf{x} = (x_1, \ldots, x_n)$ such that $\hat{\theta}_{1n}(\mathbf{x}) \neq \hat{\theta}_{2n}(\mathbf{x})\}$. For all $\mathbf{x} \in S_n$, there exists θ_n^* between $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ such that $L''(\theta_n^*) = 0$. For all $\mathbf{x} \notin S_n$, let θ_n^* be the common value of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$. Then, θ_n^* is a consistent sequence of roots of the likelihood equation.

(b) $(1/n)L''(\theta_n^*) - (1/n)L''(\theta_0) \to 0$ in probability and therefore $(1/n)L''(\theta_n^*) \to -I(\theta_0)$ in probability.

(c) Let $0 < \varepsilon < I(\theta_0)$ and let

$$S_n' = \left\{\mathbf{x} : \frac{1}{n}L''(\theta_n^*) < -I(\theta_0) + \varepsilon\right\}.$$

Then, $P_{\theta_0}(S_n') \to 1$. On the other hand, $L''(\theta_n^*) = 0$ on S_n so that S_n is contained in the complement of S_n' (Huzurbazar 1948).]

3.29 To establish the measurability of the sequence of roots $\hat{\theta}_n^*$ of Theorem 3.7, we can follow the proof of Serfling (1980, Section 4.2.2) where the measurability of a similar sequence is proved.

(a) For definiteness, define $\hat{\theta}_n(a)$ as the value that minimizes $|\hat{\theta} - \theta_0|$ subject to

$$\theta_0 - a \le \hat{\theta} \le \theta_0 + a \quad \text{and} \quad \frac{\partial}{\partial\theta} l(\theta|\mathbf{x})|_{\theta=\hat{\theta}} = 0.$$

Show that $\tilde{\theta}_n(a)$ is measurable.

(b) Show that θ_n^*, the root closest to θ^*, is measurable.

[*Hint*: For part (a), write the set $\{\hat{\theta}_n(a) > t\}$ as countable unions and intersections of measurable sets, using the fact that $(\partial/\partial\theta)\log(\theta|\mathbf{x})$ is continuous, and hence measurable.]

Section 4

4.1 Let

$$u(t) = \begin{cases} c\int_0^t e^{-1/x(1-x)}dx & \text{for } 0 < t < 1 \\ 0 & \text{for } t \le 0 \\ 1 & \text{for } t \ge 1. \end{cases}$$

Show that for a suitable c, the function u is continuous and infinitely differentiable for $-\infty < t < \infty$.

4.2 Show that the density (4.1) with $\Omega = (0, \infty)$ satisfies all conditions of Theorem 3.10 with the exception of (d) of Theorem 2.6.

4.3 Show that the density (4.4) with $\Omega = (0, \infty)$ satisfies all conditions of Theorem 3.10.

4.4 In Example 4.5, evaluate the estimators (4.8) and (4.14) for the Cauchy case, using for $\tilde{\theta}_n$ the sample median.

4.5 In Example 4.7, show that $l(\theta)$ is concave.

4.6 In Example 4.7, if $\eta = \xi$, show how to obtain a $\sqrt{n}$-consistent estimator by equating sample and population second moments.

4.7 In Theorem 4.8, show that $\sigma_{11} = \sigma_{12}$.

4.8 Without using Theorem 4.8, in Example 4.13 show that the EM sequence converges to the MLE.

4.9 Consider the following 12 observations from a bivariate normal distribution with parameters $\mu_1 = \mu_2 = 0, \sigma_1^2, \sigma_2^2, \rho$:

x_1	1	1	-1	-1	2	2	-2	-2	*	*	*	*
x_2	1	-1	1	-1	*	*	*	*	2	2	-2	-2

where "$*$" represents a missing value.

(a) Show that the likelihood function has global maxima at $\rho = \pm 1/2, \sigma_1^2 = \sigma_2^2 = 8/3$, and a saddlepoint at $\rho = 0, \sigma_1^2 = \sigma_2^2 = 5/2$.

(b) Show that if an EM sequence starts with $\rho = 0$, then it remains with $\rho = 0$ for all subsequent iterations.

(c) Show that if an EM sequence starts with ρ bounded away from zero, it will converge to a maximum.

[This problem is due to Murray (1977), and is discussed by Wu (1983).]

4.10 Show that if the EM complete-data density $f(\mathbf{y}, \mathbf{z}|\theta)$ of (4.21) is in a curved exponential family, then the hypotheses of Theorem 4.12 are satisfied.

4.11 In the EM algorithm, calculation of the E-step, the expectation calculation, can be complicated. In such cases, it may be possible to replace the E-step by a Monte Carlo evaluation, creating the MCEM algorithm (Wei and Tanner 1990). Consider the following MCEM evaluation of $Q(\theta|\hat{\theta}_{(j)}, \mathbf{y})$:

Given $\hat{\theta}^{(k)}_{(j)}$

(1) Generate $Z_1, \cdots, Z_k$, iid, from $k(\mathbf{z}|\hat{\theta}^{(k)}_{(j)}, \mathbf{y})$,

(2) Let $\hat{Q}(\theta|\hat{\theta}^{(k)}_{(j)}, \mathbf{y}) = \frac{1}{k}\sum_{i=1}^k \log L(\theta|\mathbf{y}, \mathbf{z})$

and then calculate $\hat{\theta}^{(k)}_{(j+1)}$ as the value that maximizes $\hat{Q}(\theta|\hat{\theta}^{(k)}_{(j)}, \mathbf{y})$.

(a) Show that $\hat{Q}(\theta|\hat{\theta}^{(k)}_{(j)}, \mathbf{y}) \to \hat{Q}(\theta|\hat{\theta}_{(j)}, \mathbf{y})$ as $k \to \infty$.

(b) What conditions will ensure that $L(\hat{\theta}^{(k)}_{(j+1)}|\mathbf{y}) \ge L(\hat{\theta}^{(k)}_{(j)}|\mathbf{y})$ for sufficiently large k? Are the hypotheses of Theorem 4.12 sufficient?

4.12 For the mixture distribution of Example 4.7, that is,

$$X_i \sim \theta g(x) + (1-\theta)h(x), \quad i = 1, \ldots, n, \text{ independent}$$

where $g(\cdot)$ and $h(\cdot)$ are known, an EM algorithm can be used to find the ML estimator of θ. Let $Z_1, \cdots, Z_n$, where Z_i indicates from which distribution X_i has been drawn, so

$$X_i|Z_i = 1 \sim g(x)$$
$$X_i|Z_i = 0 \sim h(x).$$

(a) Show that the complete-data likelihood can be written

$$L(\theta|\mathbf{x}, \mathbf{z}) = \prod_{i=1}^n [z_i g(x_i) + (1-z_i)h(x_i)]\,\theta^{z_i}(1-\theta)^{1-z_i}.$$

(b) Show that $E(Z_i|\theta, x_i) = \theta g(x_i)/[\theta g(x_i) + (1-\theta)h(x_i)]$ and hence that the EM sequence is given by

$$\hat{\theta}_{(j+1)} = \frac{1}{n}\sum_{i=1}^n \frac{\hat{\theta}_{(j)} g(x_i)}{\hat{\theta}_{(j)} g(x_i) + (1-\hat{\theta}_{(j)})h(x_i)}.$$

(c) Show that $\hat{\theta}_{(j)} \to \hat{\theta}$, the ML estimator of θ.

4.13 For the situation of Example 4.10:

(a) Show that the M-step of the EM algorithm is given by

$$\hat{\mu} = \left(\sum_{i=1}^4 \sum_{j=1}^{n_i} y_{ij} + z_1 + z_2\right)/12,$$

$$\hat{\alpha}_i = \left(\sum_{j=1}^2 y_{ij} + z_i\right)/3 - \hat{\mu}, \quad i = 1, 3$$
$$= \left(\sum_{j=1}^3 y_{ij}\right)/3 - \hat{\mu}, \quad i = 2, 4.$$

(b) Show that the E-step of the EM algorithm is given by

$$z_i = E\left[Y_{i3}|\mu = \hat{\mu}, \alpha_i = \hat{\alpha}_i\right] = \hat{\mu} + \hat{\alpha}_i \quad i = 1, 3.$$

(c) Under the restriction $\sum_i \alpha_i = 0$, show that the EM sequence converges to $\hat{\alpha}_i = \bar{y}_{i\cdot} - \hat{\mu}$, where $\hat{\mu} = \sum_i \bar{y}_{i\cdot}/4$.

(d) Under the restriction $\sum_i n_i \alpha_i = 0$, show that the EM sequence converges to $\hat{\alpha}_i = \bar{y}_{i\cdot} - \hat{\mu}$, where $\hat{\mu} = \sum_{ij} y_{ij}/10$.

(e) For a general one-way layout with a treatments and n_{ij} observations per treatment, show how to use the EM algorithm to augment the data so that each treatment has n observation. Write down the EM sequence, and show what it converges to under the restrictions of parts (c) and (d).

[The restrictions of parts (c) and (d) were encountered in Example 3.4.9, where they led, respectively, to an *unweighted* means analysis and a *weighted* means analysis.]

4.14 In the two-way layout (see Example 3.4.11), the EM algorithm can be very helpful in computing ML estimators in the unbalanced case. Suppose that we observe

$$Y_{ijk} : N(\xi_{ij}, \sigma^2), \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, n_{ij},$$

where $\xi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$. The data will be augmented so that the complete data have n observations per cell.

(a) Show how to compute both the E-step and the M-step of the EM algorithm.

(b) Under the restriction $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$, show that the EM sequence converges to the ML estimators corresponding to an unweighted means analysis.

(c) Under the restriction $\sum_i n_{i\cdot}\alpha_i = \sum_j n_{\cdot j}\beta_j = \sum_i n_{i\cdot}\gamma_{ij} = \sum_j \cdot j\gamma_{ij} = 0$, show that the EM sequence converges to the ML estimators corresponding to a weighted means analysis.

4.15 For the one-way layout with random effects (Example 3.5.1), the EM algorithm is useful for computing ML estimates. (In fact, it is very useful in many mixed models; see Searle et al. 1992, Chapter 8.) Suppose we have the model

$$X_{ij} = \mu + A_i + U_{ij} \quad (j = 1, \ldots, n_i, \ i = 1, \ldots, s)$$

where A_i and U_{ij} are independent normal random variables with mean zero and known variance. To compute the ML estimates of μ, σ_U^2, and σ_U^2 it is typical to employ an EM algorithm using the unobservable A_i's as the augmented data. Write out both the E-step and the M-step, and show that the EM sequence converges to the ML estimators.

4.16 Maximum likelihood estimation in the *probit model* of Section 3.6 can be implemented using the EM algorithm. We observe independent Bernoulli variables $X_1, \ldots, X_n$, which depend on unobservable variables Z_i distributed independently as $N(\zeta_i, \sigma^2)$, where

$$X_i = \begin{cases} 0 & \text{if } Z_i \le u \\ 1 & \text{if } Z_i > u. \end{cases}$$

Assuming that u is known, we are interested in obtaining ML estimates of ζ and σ^2.

(a) Show that the likelihood function is $p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$, where

$$p = P(Z_i > u) = \Phi\left(\frac{\zeta - u}{\sigma}\right).$$

(b) If we consider $Z_1, \ldots, Z_n$ to be the complete data, the complete-data likelihood is

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(z_i-\zeta)^2}$$

and the expected complete-data log likelihood is

$$-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[E(Z_i^2|X_i) - 2\zeta E(Z_i|X_i) + \zeta^2\right].$$

(c) Show that the EM sequence is given by

$$\hat{\zeta}_{(j+1)} = \frac{1}{n}\sum_{i=1}^{n} t_i(\hat{\zeta}_{(j)}, \hat{\sigma}^2_{(j)})$$

$$\hat{\sigma}^2_{(j+1)} = \frac{1}{n}\left[\sum_{i=1}^{n} v_i(\hat{\zeta}_{(j)}, \hat{\sigma}^2_{(j)}) - \frac{1}{n}\left(\sum_{i=1}^{n} t_i(\hat{\zeta}_{(j)}, \hat{\sigma}^2_{(j)})\right)^2\right]$$

where

$$t_i(\zeta, \sigma^2) = E(Z_i|X_i, \zeta, \sigma^2) \quad \text{and} \quad v_i(\zeta, \sigma^2) = E(Z_i^2|X_i, \zeta, \sigma^2).$$

(d) Show that

$$E(Z_i|X_i, \zeta, \sigma^2) = \zeta + \sigma H_i\left(\frac{u-\zeta}{\sigma}\right),$$

$$E(Z_i^2|X_i, \zeta, \sigma^2) = \zeta^2 + \sigma^2 + \sigma(u+\zeta)H_i\left(\frac{u-\zeta}{\sigma}\right)$$

where

$$H_i(t) = \begin{cases} \dfrac{\varphi(t)}{1-\Phi(t)} & \text{if } X_i = 1 \\[2ex] -\dfrac{\varphi(t)}{\Phi(t)} & \text{if } X_i = 0. \end{cases}$$

(e) Show that $\hat{\zeta}_{(j)} \to \hat{\zeta}$ and $\hat{\sigma}^2_{(j)} \to \hat{\sigma}^2$, the ML estimates of ζ and σ^2.

4.17 Verify (4.30).

4.18 The EM algorithm can also be implemented in a Bayesian hierarchical model to find a posterior mode. Recall the model (4.5.5.1),

$$X|\theta \sim f(x|\theta),$$
$$\Theta|\lambda \sim \pi(\theta|\lambda),$$
$$\Lambda \sim \gamma(\lambda),$$

where interest would be in estimating quantities from $\pi(\theta|x)$. Since

$$\pi(\theta|x) = \int \pi(\theta, \lambda|x)d\lambda,$$

where $\pi(\theta, \lambda|x) = \pi(\theta|\lambda, x)\pi(\lambda|x)$, the EM algorithm is a candidate method for finding the mode of $\pi(\theta|x)$, where λ would be used as the augmented data.

(a) Define $k(\lambda|\theta, x) = \pi(\theta, \lambda|x)/\pi(\theta|x)$, and show that

$$\log \pi(\theta|x) = \int \log \pi(\theta, \lambda|x)k(\lambda|\theta^*, x)d\lambda - \int \log k(\lambda|\theta, x)k(\lambda|\theta^*, x)d\lambda.$$

(b) If the sequence $\{\hat{\theta}_{(j)}\}$ satisfies

$$\max_\theta \int \log \pi(\theta, \lambda|x)k(\lambda|\theta_{(j)}, x)d\lambda = \int \log \pi(\theta_{(j+1)}, \lambda|x)k(\lambda|\theta_{(j)}, x)d\lambda,$$

show that $\log \pi(\theta_{(j+1)}|x) \geq \log \pi(\theta_{(j)}|x)$. Under what conditions will the sequence $\{\hat{\theta}_{(j)}\}$ converge to the mode of $\pi(\theta|x)$?

(c) For the hierarchy

$$X|\theta \sim N(\theta), 1),$$
$$\Theta|\lambda \sim N(\lambda, 1)),$$
$$\Lambda \sim \text{Uniform}(-\infty, \infty),$$

show how to use the EM algorithm to calculate the posterior mode of $\pi(\theta|x)$.

4.19 There is a connection between the EM algorithm and Gibbs sampling, in that both have their basis in Markov chain theory. One way of seeing this is to show that the incomplete-data likelihood is a solution to the integral equation of successive substitution sampling (see Problems 4.5.9-4.5.11), and that Gibbs sampling can then be used to calculate the likelihood function. If $L(\theta|\mathbf{y})$ is the incomplete-data likelihood and $L(\theta|\mathbf{y}, \mathbf{z})$ is the complete-data likelihood, define

$$L^*(\theta|\mathbf{y}) = \frac{L(\theta|\mathbf{y})}{\int L(\theta|\mathbf{y})d\theta},$$
$$L^*(\theta|\mathbf{y}, \mathbf{z}) = \frac{L(\theta|\mathbf{y}, \mathbf{z})}{\int L(\theta|\mathbf{y}, \mathbf{z})d\theta}.$$

.

(a) Show that $L^*(\theta|\mathbf{y})$ is the solution to

$$L^*(\theta|\mathbf{y}) = \int \left[\int L^*(\theta|\mathbf{y}, \mathbf{z})k(\mathbf{z}|\theta', \mathbf{y})d\mathbf{z}\right] L^*(\theta'|\mathbf{y})d\theta'$$

where, as usual, $k(\mathbf{z}|\theta, \mathbf{y}) = L(\theta|\mathbf{y}, \mathbf{z})/L(\theta|\mathbf{y})$.

(b) Show how the sequence $\theta_{(j)}$ from the Gibbs iteration,

$$\theta_{(j)} \sim L^*(\theta|\mathbf{y}, \mathbf{z}_{(j-1)}),$$
$$\mathbf{z}_{(j)} \sim k(\mathbf{z}|\theta_{(j)}, \mathbf{y}),$$

will converge to a random variable with density $L^*(\theta|\mathbf{y})$ as $j \to \infty$. How can this be used to compute the likelihood function $L(\theta|\mathbf{y})$?

[Using the functions $L(\theta|\mathbf{y}, \mathbf{z})$ and $k(\mathbf{z}|\theta, \mathbf{y})$, the EM algorithm will get us the ML estimator from $L(\theta|\mathbf{y})$, whereas the Gibbs sampler will get us the entire function. This likelihood implementation of the Gibbs sampler was used by Casella and Berger (1994) and is also described by Smith and Roberts (1993). A version of the EM algorithm, where the Markov chain connection is quite apparent, was given by Baum and Petrie (1966) and Baum et al. (1970).]

Section 5

5.1 (a) If a vector $\mathbf{Y}_n$ in E_s converges in probability to a constant vector $\mathbf{a}$, and if h is a continuous function defined over E_s, show that $h(\mathbf{Y}_n) \to h(\mathbf{a})$ in probability.

(b) Use (a) to show that the elements of $||A_{jkn}||^{-1}$ tend in probability to the elements of B as claimed in the proof of Lemma 5.2.

[*Hint*: (a) Apply Theorem 1.8.19 and Problem 1.8.13.]

5.2 (a) Show that (5.26) with the remainder term neglected has the same form as (5.15) and identify the A_{jkn}.

(b) Show that the resulting a_{jk} of Lemma 5.2 are the same as those of (5.23).

(c) Show that the remainder term in (5.26) can be neglected in the proof of Theorem 5.3.

5.3 Let $X_1, \ldots, X_n$ be iid according to $N(\xi, \sigma^2)$.

(a) Show that the likelihood equations have a unique root.

(b) Show directly (i.e., without recourse to Theorem 5.1) that the MLEs $\hat{\xi}$ and $\hat{\sigma}$ are asymptotically efficient.

5.4 Let $(X_0, \ldots, X_s)$ have the multinomial distribution $M(p_0, \ldots, p_s; n)$.

(a) Show that the likelihood equations have a unique root.

(b) Show directly that the MLEs $\hat{p}_i$ are asymptotically efficient.

5.5 Prove Corollary 5.4.

5.6 Show that there exists a function f of two variables for which the equations $\partial f(x, y)/\partial x = 0$ and $\partial f(x, y)/\partial y = 0$ have a unique solution, and this solution is a local but not a global maximum of f.

Section 6

6.1 In Example 6.1, show that the likelihood equations are given by (6.2) and (6.3).

6.2 In Example 6.1, verify Equation (6.4).

6.3 Verify (6.5).

6.4 If $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r, \theta_{r+1}, \ldots, \theta_s)$ and if

$$\operatorname{cov}\left[\frac{\partial}{\partial\theta_i} L(\boldsymbol{\theta}), \frac{\partial}{\partial\theta_j} L(\boldsymbol{\theta})\right] = 0 \quad \text{for any } i \le r < j,$$

then the asymptotic distribution of $(\hat{\theta}_1, \ldots, \hat{\theta}_r)$ under the assumptions of Theorem 5.1 is unaffected by whether or not $\theta_{r+1}, \ldots, \theta_s$ are known.

6.5 Let $X_1, \ldots, X_n$ be iid from a $\Gamma(\alpha, \beta)$ distribution with density $1/(\Gamma(\alpha)\beta^\alpha) \times x^{\alpha-1} e^{-x/\beta}$.

(a) Calculate the information matrix for the usual (α, β) parameterization.

(b) Write the density in terms of the parameters $(\alpha, \mu) = (\alpha, \alpha/\beta)$. Calculate the information matrix for the (α, μ) parameterization and show that it is diagonal, and, hence, the parameters are orthogonal.

(c) If the MLE's in part (a) are $(\hat{\alpha}, \hat{\beta})$, show that $(\hat{\alpha}, \hat{\mu}) = (\hat{\alpha}, \hat{\alpha}/\hat{\beta})$. Thus, either model estimates the mean equally well.

(For the theory behind, and other examples of, parameter orthogonality, see Cox and Reid 1987.)

6.6 In Example 6.4, verify the MLEs $\hat{\xi}_i$ and $\hat{\sigma}_{jk}$ when the ξ's are unknown.

6.7 In Example 6.4, show that the S_{jk} given by (6.15) are independent of $(X_1, \ldots, X_p)$ and have the same joint distribution as the statistics (6.13) with n replaced by $n - 1$. [*Hint*: Subject each of the p vectors $(X_{i1}, \ldots, X_{in})$ to the same orthogonal transformation, where the first row of the orthogonal matrix is $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$.]

6.8 Verify the matrices (a) (6.17) and (b) (6.18).

6.9 Consider the situation leading to (6.20), where $(X_i, Y_i), i = 1, \ldots, n$, are iid according to a bivariate normal distribution with $E(X_i) = E(Y_i) = 0$, $\operatorname{var}(X_i) = \operatorname{var}(Y_i) = 1$, and unknown correlation coefficient ρ.

(a) Show that the likelihood equation is a cubic for which the probability of a unique root tends to 1 as $n \to \infty$. [*Hint*: For a cubic equation $ax^3 + 3bx^2 + 3cx + d = 0$, let $G = a^2d - 3abc + 2b^3$ and $H = ac - b^2$. Then the condition for a unique real root is $G^2 + 4H^3 > 0$.]

(b) Show that if $\hat{\rho}_n$ is a consistent solution of the likelihood equation, then it satisfies (6.20).

(c) Show that $\delta = \Sigma X_i Y_i / n$ is a consistent estimator of ρ and that $\sqrt{n}(\delta - \rho) \xrightarrow{\mathcal{L}} N(0, 1 + \rho^2)$ and, hence, that δ is less efficient than the MLE of ρ.

6.10 Verify the limiting distribution asserted in (6.21).

6.11 Let $X_, \ldots, X_n$ be iid according to the Poisson distribution $P(\lambda)$. Find the ARE of $\delta_{2n} = [\text{No. of } X_i = 0]/n$ to $\delta_{1n} = e^{-\bar{X}_n}$ as estimators of $e^{-\lambda}$.

6.12 Show that the efficiency (6.27) tends to 0 as $|a - \theta| \to \infty$.

6.13 For the situation of Example 6.9, consider as another family of distributions, the contaminated normal mixture family suggested by Tukey (1960) as a model for observations which usually follow a normal distribution but where occasionally something goes wrong with the experiment or its recording, so that the resulting observation is a gross error. Under the *Tukey model*, the distribution function takes the form

$$F_{\tau,\epsilon}(t) = (1 - \epsilon)\Phi(t) + \epsilon\Phi\left(\frac{t}{\tau}\right).$$

That is, in the gross error cases, the observations are assumed to be normally distributed with the same mean θ but a different (larger) variance τ^2.[9]

(a) Show that if the X_i's have distribution $F_{\tau,\epsilon}(x - \theta)$, the limiting distribution of δ_{2n} is unchanged.

(b) Show that the limiting distribution of δ_{1n} is normal with mean zero and variance $\frac{n}{n-1}\left\{\phi\left[\sqrt{\frac{n}{n-1}}(a - \theta)\right]\right\}^2 (1 - \epsilon + \epsilon\tau^2)$.

(c) Compare the asymptotic relative efficiency of δ_{1n} and δ_{2n}.

6.14 Let $X_1, \ldots, X_n$ be iid as $N(0, \sigma^2)$.

(a) Show that $\delta_n = k\Sigma|X_i|/n$ is a consistent estimator of σ if and only if $k = \sqrt{\pi/2}$.

(b) Determine the ARE of δ with $k = \sqrt{\pi/2}$ with respect to the MLE $\sqrt{\Sigma X_i^2/n}$.

6.15 Let $X_1, \ldots, X_n$ be iid with $E(X_i) = \theta$, $\text{var}(X_i) = 1$, and $E(X_i - \theta)^4 = \mu_4$, and consider the unbiased estimators $\delta_{1n} = (1/n)\Sigma X_i^2 - 1$ and $\delta_{2n} = \bar{X}_n^2 - 1/n$ of θ^2.

(a) Determine the ARE $e_{2,1}$ of δ_{2n} with respect to δ_{1n}.

(b) Show that $e_{2,1} \geq 1$ if the X_i are symmetric about θ.

(c) Find a distribution for the X_i for which $e_{2,1} < 1$.

6.16 The property of asymptotic relative efficiency was defined (Definition 6.6) for estimators that converged to normality at rate $\sqrt{n}$. This definition, and Theorem 6.7, can be generalized to include other distributions and rates of convergence.

[9] As has been pointed out by Stigler (1973) such models for heavy-tailed distributions had already been proposed much earlier in a forgotten work by Newcomb (1882, 1886).

Theorem 9.1 *Let $\{\delta_{in}\}$ be two sequences of estimators of $g(\theta)$ such that*

$$n^\alpha[\delta_{in} - g(\theta)] \xrightarrow{\mathcal{L}} \tau_i T, \ \alpha > 0, \ \tau_i > 0, \ i = 1, 2,$$

where the distribution H of T has support on an interval $-\infty \le A < B \le \infty$ with strictly increasing cdf on (A, B). Then, the ARE of $\{\delta_{2n}\}$ with respect to $\{\delta_{1n}\}$ exists and is

$$e_{21} = \lim_{n_2 \to \infty} \frac{n_1(n_2)}{n_2} = \left[\frac{\tau_1}{\tau_2}\right]^{1/\alpha}.$$

6.17 In Example 6.10, show that the conditions of Theorem 5.1 are satisfied.

Section 7

7.1 Prove Theorem 7.1.

7.2 For the situation of Example 7.3 with $m = n$:

(a) Show that a necessary condition for (7.5) to converge to $N(0, 1)$ is that $\sqrt{n}(\hat{\lambda} - \lambda) \to 0$, where $\hat{\lambda} = \hat{\sigma}^2/\hat{\tau}^2$ and $\lambda = \sigma^2/\tau^2$, for $\hat{\sigma}^2$ and $\hat{\tau}^2$ of (7.4).

(b) Use the fact that $\hat{\lambda}/\lambda$ has an F-distribution to show that $\sqrt{n}(\hat{\lambda} - \lambda) \not\to 0$.

(c) Show that the full MLE is given by the solution to

$$\xi = \frac{(m/\sigma^2)\bar{X} + (n/\tau^2)\bar{Y}}{m/\sigma^2 + n/\tau^2}, \quad \sigma^2 = \frac{1}{m}\Sigma(X_i - \xi)^2, \quad \tau^2 = \frac{1}{n}\Sigma(Y_j - \xi)^2,$$

and deduce its asymptotic efficiency from Theorem 5.1.

7.3 In Example 7.4, determine the joint distribution of (a) $(\hat{\sigma}^2, \hat{\tau}^2)$ and (b) $(\hat{\sigma}^2, \hat{\sigma}_A^2)$.

7.4 Consider samples $(X_1, Y_1), \ldots, (X_m, Y_m)$ and $(X_1', Y_1'), \ldots, (X_n', Y_n')$ from two bivariate normal distributions with means zero and variance-covariances $(\sigma^2, \tau^2, \rho\sigma\tau)$ and $(\sigma'^2, \tau'^2, \rho'\sigma'\tau')$, respectively. Use Theorem 7.1 and Examples 6.5 and 6.8 to find the limit distribution

(a) of $\hat{\sigma}^2$ and $\hat{\tau}^2$ when it is known that $\rho' = \rho$

(b) of $\hat{\rho}$ when it is known that $\sigma' = \sigma$ and $\tau' = \tau$.

7.5 In the preceding problem, find the efficiency gain (if any)

(a) in part (a) resulting from the knowledge that $\rho' = \rho$

(b) in part (b) resulting from the knowledge that $\sigma' = \sigma$ and $\tau' = \tau$.

7.6 Show that the likelihood equations (7.11) have at most one solution.

7.7 In Example 7.6, suppose that $p_i = 1 - F(\alpha + \beta t_i)$ and that both $\log F(x)$ and $\log[1 - F(x)]$ are strictly concave. Then, the likelihood equations have at most one solution.

7.8 (a) If the cdf F is symmetric and if $\log F(x)$ is strictly concave, so is $\log[1 - F(x)]$.

(b) Show that $\log F(x)$ is strictly concave when F is strongly unimodal but not when F is Cauchy.

7.9 In Example 7.7, show that Y_n is less informative than Y.

[*Hint*: Let Z_n be distributed as $P(\lambda\Sigma_{i=n+1}^\infty \gamma_i)$ independently of Y_n. Then, $Y_n + Z_n$ is a sufficient statistic for λ on the basis of (Y_n, Z_n) and $Y_n + Z_n$ has the same distribution as Y.]

7.10 Show that the estimator δ_n of Example 7.7 satisfies (7.14).

7.11 Find suitable normalizing constants for δ_n of Example 7.7 when (a) $\gamma_i = i$, (b) $\gamma_i = i^2$, and (c) $\gamma_i = 1/i$.

7.12 Let X_i $(i = 1, \ldots, n)$ be independent normal with variance 1 and mean βt_i (with t_i known). Discuss the estimation of β along the lines of Example 7.7.

7.13 Generalize the preceding problem to the situation in which (a) $E(X_i) = \alpha + \beta t_i$ and $\text{var}(X_i) = 1$ and (b) $E(X_i) = \alpha + \beta t_i$ and $\text{var}(X_i) = \sigma^2$ where α, β, and σ^2 are unknown parameters to be estimated.

7.14 Let X_j $(j = 1, \ldots, n)$ be independently distributed with densities $f_j(x_j|\theta)$ (θ real-valued), let $I_j(\theta)$ be the information X_j contains about θ, and let $T_n(\theta) = \Sigma_{j=1}^n I_j(\theta)$ be the total information about θ in the sample. Suppose that $\hat{\theta}_n$ is a consistent root of the likelihood equation $L'(\theta) = 0$ and that, in generalization of (3.18)-(3.20),

$$\frac{1}{\sqrt{T_n(\theta_0)}} L'(\theta_0) \xrightarrow{\mathcal{L}} N(0, 1)$$

and

$$-\frac{L''(\theta_0)}{T_n(\theta_0)} \xrightarrow{P} 1 \quad \text{and} \quad \frac{L'''(\theta_n^*)}{T_n(\theta_0)} \quad \text{is bounded in probability.}$$

Show that

$$\sqrt{T_n(\theta_0)}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, 1).$$

7.15 Prove that the sequence $X_1, X_2, \ldots$ of Example 7.8 is stationary provided it satisfies (7.17).

7.16 (a) In Example 7.8, show that the likelihood equation has a unique solution, that it is the MLE, and that it has the same asymptotic distribution as $\delta_n' = \sum_{i=1}^n X_i X_{i+1} / \sum_{i=1}^n X_i^2$.

(b) Show directly that δ_n' is a consistent estimator of β.

7.17 In Example 7.8:

(a) Show that for $j > 1$ the expected value of the conditional information (given X_{j-1}) that X_j contains about β is $1/(1 - \beta^2)$.

(b) Determine the information X_1 contains about β.

7.18 When $\tau = \sigma$ in (7.21), show that the MLE exists and is consistent.

7.19 Suppose that in (7.21), the ξ's are themselves random variables, which are iid as $N(\mu, \gamma^2)$.

(a) Show that the joint density of the (X_i, Y_i) is that of a sample from a bivariate normal distribution, and identify the parameters of that distribution.

(b) In the model of part (a), find asymptotically efficient estimators of the parameters $\mu, \gamma, \beta, \sigma$, and τ.

7.20 Verify the roots (7.22).

7.21 Show that the likelihood (7.21) is unbounded.

7.22 Show that if ρ s defined by (7.24), then ρ and ρ' are everywhere continuous.

7.23 Let F have a differentiable density f and let $\int \psi^2 f < \infty$.

(a) Use integration by parts to write the denominator of (7.27) as $[\int \psi(x) f'(x) dx]^2$.

(b) Show that $\sigma^2(F, \psi) \geq [\int (f'/f)^2 f]^{-1} = I_f^{-1}$ by applying the Schwarz inequality to part (a).

The following three problems will investigate the technical conditions required for the consistency and asymptotic normality of M*-estimators, as noted in (7.26).*

7.24 To have consistency of M-estimators, a sufficient condition is that the root of the estimating function be unique and isolated. Establish the following theorem.

Theorem 9.2 *Assume that conditions (A0)-(A3) hold. Let t_0 be an isolated root of the equation $E_{\theta_0}[\psi(X,t)] = 0$, where $\psi(\cdot, t)$ is monotone in t and continuous in a neighborhood of t_0. If $T_0(\mathbf{x})$ is a solution to $\sum_{i=1}^n \psi(x_i, t) = 0$, then T_0 converges to t_0 in probability.*

[*Hint:* The conditions on ψ imply that $E_{\theta_0}[\psi(X,t)]$ is monotone, so t_0 is a unique root. Adapt the proofs of Theorems 3.2 and 3.7 to complete this proof.]

7.25 **Theorem 9.3** *Under the conditions of Theorem 9.2, if, in addition*

(i) $E_{\theta_0}\left[\frac{\partial}{\partial t}\psi(X,t)|_{t=t_0}\right]$ *is finite and nonzero,*

(ii) $E_{\theta_0}\left[\psi^2(X,t_0)\right] < \infty$,

then

$$\sqrt{n}(T_0 - t_0) \xrightarrow{\mathcal{L}} N(0, \sigma^2_{T_0}),$$

where $\sigma^2_{T_0} = E_{\theta_0}\left[\psi^2(X,t_0)\right]/(E_{\theta_0}\left[\frac{\partial}{\partial t}\psi(X,t)|_{t=t_0}\right])^2$.

[Note that this is a slight generalization of (7.27).]

[*Hint:* The assumptions on ψ are enough to adapt the Taylor series argument of Theorem 3.10, where ψ takes the place of l'.]

7.26 For each of the following estimates, write out the ψ function that determines it, and show that the estimator is consistent and asymptotically normal under the conditions of Theorems 9.2 and 9.3.

(a) The *least squares estimate*, the minimizer of $\sum(x_i - t)^2$.

(b) The *least absolute value estimate*, the minimizer of $\sum |x_i - t|$.

(c) The *Huber trimmed mean*, the minimizer of (7.24).

7.27 In Example 7.12, compare (a) the asymptotic distributions of $\hat{\xi}$ and δ_n; (b) the normalized expected squared error of $\hat{\xi}$ and δ_n.

7.28 In Example 7.12, show that (a) $\sqrt{n}(\hat{\hat{b}} - b) \xrightarrow{\mathcal{L}} N(0, b^2)$ and (b) $\sqrt{n}(\hat{b} - b) \xrightarrow{\mathcal{L}} N(0, b^2)$.

7.29 In Example 7.13, show that

(a) $\hat{c}$ and $\hat{a}$ are independent and have the stated distributions;

(b) $X_{(1)}$ and $\Sigma \log[X_i/X_{(1)}]$ are complete sufficient statistics on the basis of a sample from (7.33).

7.30 In Example 7.13, determine the UMVU estimators of a and c, and the asymptotic distributions of these estimators.

7.31 In the preceding problem, compare (a) the asymptotic distribution of the MLE and the UMVU estimator of c; (b) the normalized expected squared error of these two estimators.

7.32 In Example 7.15, (a) verify equation (7.39), (b) show that the choice $a = -2$ produces the estimator with the best second-order efficiency, (c) show that the limiting risk ratio of the MLE ($a = 0$) to $\delta_n (a = -2)$ is 2, and (d) discuss the behavior of this estimator in small samples.

7.33 Let $X_1, \ldots, X_n$ be iid according to the three-parameter lognormal distribution (7.37). Show that

(a)

$$p^*(\mathbf{x}|\xi) = \sup_{\gamma,\sigma^2} p(\mathbf{x}|\xi,\gamma,\sigma^2) = c/[\hat{\sigma}(\xi)]^n \Pi[1/(x_i - \xi)]$$

where

$$p(\mathbf{x}|\xi,\gamma,\sigma^2) = \prod_{i=1}^{n} f(x_i|\xi,\sigma^2),$$

$$\hat{\sigma}^2(\xi) = \frac{1}{n}\Sigma[\log(x_i - \xi) - \hat{\gamma}(\xi)]^2 \text{ and } \hat{\gamma}(\xi) = \frac{1}{n}\Sigma\log(x_i - \xi).$$

(b) $p^*(\mathbf{x}|\xi) \to \infty$ as $\xi \to x_{(1)}$.

[*Hint*: (b) For ξ sufficiently near $x_{(1)}$,

$$\hat{\sigma}^2(\xi) \leq \frac{1}{n}\Sigma[\log(x_i - \xi)]^2 \leq [\log(x_{(1)} - \xi)]^2$$

and hence

$$p^*(\mathbf{x}|\xi) \geq |\log(x_{(1)} - \xi)|^{-n}\Pi(x_{(i)} - \xi)^{-1}.$$

The right side tends to infinity as $\xi \to x_{(1)}$ (Hill 1963.]

7.34 The derivatives of all orders of the density (7.37) tend to zero as $x \to \xi$.

Section 8

8.1 Determine the limit distribution of the Bayes estimator corresponding to squared error loss, and verify that it is asymptotically efficient, in each of the following cases:

(a) The observations $X_1, \ldots, X_n$ are iid $N(\theta, \sigma^2)$, with σ known, and the estimand is θ. The prior distribution for Θ is a conjugate normal distribution, say $N(\mu, b^2)$. (See Example 4.2.2.)

(b) The observations Y_i have the gamma distribution $\Gamma(\gamma, 1/\tau)$, the estimand is $1/\tau$, and τ has the conjugate prior density $\Gamma(g, \alpha)$.

(c) The observations and prior are as in Problem 4.1.9 and the estimand is λ.

(d) The observations Y_i have the negative binomial distribution (4.3), p has the prior density $B(a, b)$, and the estimand is (a) p and (b) $1/b$.

8.2 Referring to Example 8.1, consider, instead, the minimax estimator δ_n of p given by (1.11) which corresponds to the sequence of beta priors with $a = b = \sqrt{n}/2$. Then,

$$\sqrt{n}[\delta_n - p] = \sqrt{n}\left(\frac{X}{n} - p\right) + \frac{\sqrt{n}}{1+\sqrt{n}}\left(\frac{1}{2} - \frac{X}{n}\right).$$

(a) Show that the limit distribution of $\sqrt{n}[\delta_n - p]$ is $N[\frac{1}{2} - p, p(1-p)]$, so that δ_n has the same asymptotic variance as X/n, but that for $p \neq \frac{1}{2}$, it is asymptotically biased.

(b) Show that ARE of δ_n relative to X/n does not exist except in the case $p = \frac{1}{2}$ when it is 1.

8.3 The assumptions of Theorem 2.6 imply (8.1) and (8.2).

8.4 In Example 8.5, the posterior density of θ after one observation is $f(x_1 - \theta)$; it is a proper density, and it satisfies (B5) provided $E_\theta|X_1| < \infty$.

8.5 Let $X_1, \ldots, X_n$ be independent, positive variables, each with density $(1/\tau)f(x_i/\tau)$, and let τ have the improper density $\pi(\tau) = 1/\tau$ $(\tau > 0)$. The posterior density after one observation is a proper density, and it satisfies (B5), provided $E_\tau(1/X_1) < \infty$.

8.6 Give an example in which the posterior density is proper (with probability 1) after two observations but not after one.

[*Hint*: In the preceding example, let $\pi(\tau) = 1/\tau^2$.]

8.7 Prove the result stated preceding Example 8.6.

8.8 Let $X_1, \ldots, X_n$ be iid as $N(\theta, 1)$ and consider the improper density $\pi(\theta) = e^{\theta^4}$. Then, the posterior will be improper for all n.

8.9 Prove Lemma 8.7.

8.10 (a) If $\sup|Y_n(t)| \xrightarrow{P} 0$ and $\sup|X_n(t) - c| \xrightarrow{P} 0$ as $n \to \infty$, then $\sup|X_n(t) - ce^{Y_n(t)}| \xrightarrow{P} 0$, where the sup is taken over a common set $t \in T$.

(b) Use (a) to show that (8.22) and (8.23) imply (8.21).

8.11 Show that (B1) implies (a) (8.24) and (b) (8.26).

10 Notes

10.1 Origins

The origins of the concept of maximum likelihood go back to the work of Lambert, Daniel Bernoulli, and Lagrange in the second half of the eighteenth century, and of Gauss and Laplace at the beginning of the nineteenth. (For details and references, see Edwards 1974 or Stigler 1986.) The modern history begins with Edgeworth (1908, 1909) and Fisher (1922, 1925), whose contributions are discussed by Savage (1976) and Pratt (1976).

Fisher's work was followed by a euphoric belief in the universal consistency and asymptotic efficiency of maximum likelihood estimators, at least in the iid case. The true situation was sorted out only gradually. Landmarks are Cramér (1946a, 1946b), who shifted the emphasis from the global to a local maximum and defined the "regular" case in which the likelihood equation has a consistent asymptotically efficient root; Wald (1949), who provided fairly general conditions for consistency; the counterexamples of Hodges (Le Cam, 1953) and Bahadur (1958); and Le Cam's resulting theorem on superefficiency (1953).

Convergence (under suitable restrictions and appropriately normalized) of the posterior distribution of a real-valued parameter with a prior distribution to its normal limit was first discovered by Laplace (1820) and later reobtained by Bernstein (1917) and von Mises (1931). More general versions of this result are given in Le Cam (1958). The asymptotic efficiency of Bayes solutions was established by Le Cam (1958), Bickel and Yahav (1969), and Ibragimov and Has'minskii (1972). (See also Ibragimov and Has'minskii 1981.)

Computation of likelihood estimators was influenced by the development of the *EM Algorithm* (Dempster, Laird, and Rubin 1977). This algorithm grew out of work done on iterative computational methods that were developed in the 1950s and 1960s, and can be traced back at least as far as Hartley (1958). The EM algorithm has enjoyed widespread use as a computational tool for obtaining likelihood estimators in complex problems (see Little and Rubin 1987, Tanner 1996, or McLachlan and Krishnan 1997).

10.2 Alternative Conditions for Asymptotic Normality

The Cramér conditions for asymptotic normality and efficiency that are given in Theorems 3.10 and 5.1 are not the most general; for those, see Strasser 1985, Pfanzagl 1985, or LeCam 1986. They were chosen because they have fairly wide applicability, yet are