CHAPTER 4. APPLICATIONS

This chapter describes three different general applications of the theory developed so far. The first part of the chapter contains a proof of the information inequality and a proof based on this inequality of Karlin' theorem on admissibility of linear estimators.

The second part of the chapter describes Stein's unbiased estimat of the risk and proves the minimaxity of the James-Stein estimator as a specific application of this unbiased estimate.

The third part of the chapter describes generalized Bayes estimat and contains two principle theorems describing situations in which all admiss ble estimators are generalized Bayes -- or at least have a representation similar to that of a generalized Bayes procedure. This part of the chapter deals with two basic situations. The first is estimation of the natural parameter under squared error loss, and the second is estimation of the expectation parameter under squared error loss. The so-called conjugate prio play a natural role in this second situation.

The exercises at the end of the chapter contain a non-systematic selection of some of the specific results derivable from the more general development in the body of the chapter.

INFORMATION INEQUALITY

The information inequality -- also known as the Cramer-Rao inequality -- is an easy consequence of Corollary 2.6.

The version to be proved below applies to vector-valued as well a real-valued statistics. For vector-valued statistics one needs the multivariate Cauchy-Schwarz inequality, as described in the following theorem.

If A,B are symmetric $(m \times m)$ matrices, write $A \geq B$ to mean that $A - B$ is positive semi-definite.

4.1 Theorem

Let T_1, T_2 be, respectively $(\ell \times 1)$ and $(m \times 1)$ vector-valued random variables on some probability space. Let

$$B_{11} = E(T_1 T_1') \qquad (\ell \times \ell)$$

$$B_{12} = E(T_1 T_2') \qquad (\ell \times m)$$

$$B_{22} = E(T_2 T_2') \qquad (m \times m)$$

and suppose B_{11} exists and B_{22} exists and is non-singular. Then

(1) $$B_{11} \geq B_{12} B_{22}^{-1} B_{21} \quad .$$

Remarks. If $\ell = m = 1$ this is the usual Cauchy-Schwarz inequality:

(2) $$E(T_1^2)E(T_2^2) \geq E^2(T_1 T_2) \quad .$$

If B_{22} is singular the inequality (1) remains true with generalized inverses in place of true inverses. See Exercise 4.1.1.

If 4.1(1) is applied to the random vectors $T_1 - E(T_1)$, $T_2 - E(T_2)$ it yields the covariance form of the inequality:

(3) $$\Sigma_{11} \geq \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad .$$

Proof. Consider the $((\ell + m) \times 1)$ random vector $U = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$. Then

$$0 \leq E(U U') = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

Let $W = \begin{pmatrix} I & -B_{12}B_{22}^{-1} \\ 0 & B_{22}^{-1} \end{pmatrix}$. Then

$$0 \le E(WUU'W') = W\,E(UU')W'$$

$$= \begin{pmatrix} B_{11} - B_{12}B_{22}^{-1}B_{21} & 0 \\ 0 & B_{22}^{-1} \end{pmatrix} .$$

It follows that $0 \le B_{11} - B_{12}B_{22}^{-1}B_{21}$, as desired. ||

One further preparatory lemma is needed for the form of the information inequality which appears below.

4.2 Proposition

Let $\{p_\theta\}$ be a standard k-parameter exponential family. Let T be a statistic taking values in R^ℓ. Suppose $\theta_0 \in N^\circ$ and the covariance matrix $\Sigma_{\theta_0}(T)$ of T exists at θ_0. Then $E_\theta(T)$ exists on a neighborhood of θ_0. ($\theta \in N^\circ(||T||)$ in the notation of 2.6.)

Proof. For some $\varepsilon > 0$, $||\theta - \theta_0|| < \varepsilon$ implies $\theta \in N$. Let $||\theta - \theta_0|| < \varepsilon/2$. Then, by the ordinary Cauchy-Schwarz inequality,

$$(1) \quad E_\theta(||T||) = \int ||T(x)|| \exp(\theta \cdot x - \psi(\theta))\nu(dx)$$

$$= \int ||T(x)|| \exp((\theta - \theta_0) \cdot x - \psi(\theta) + \psi(\theta_0)) \exp(\theta_0 \cdot x - \psi(\theta_0))\nu(dx)$$

$$\le [\int ||T(x)||^2 \exp(\theta_0 \cdot x - \psi(\theta_0))\, \nu(dx)$$

$$\int \exp(2(\theta - \theta_0) \cdot x - 2\psi(\theta) + 2\psi(\theta_0))\exp(\theta_0 \cdot x - \psi(\theta_0))\nu(dx)]^{\frac{1}{2}}$$

$$= E_{\theta_0}^{\frac{1}{2}}(||T(x)||^2)[\exp \psi(2(\theta - \theta_0) + \theta_0) - 2\psi(\theta) + \psi(\theta_0)]^{\frac{1}{2}}$$

$$< \infty$$

since $E_{\theta_0}(||T(x)||^2) < \infty$ by assumption and since $2(\theta - \theta_0) + \theta_0 \in N$. ||

4.3 Setting

The following version of the information inequality applies to differentiable exponential subfamilies, as defined at the end of Chapter 3.

Let $\{p_\theta : \theta \in \Theta\}$ be such a family with Θ m-dimensional. Let $\theta_0 \in \Theta$. For N a neighborhood in R^m let $\theta : N \to \Theta \subset R^k$, with $\theta(\rho_0) = \theta_0$ be a parametrization of Θ in a neighborhood of θ_0. By definition $\nabla\theta(\rho)$ is the m×k matrix with elements

$$(1) \qquad (\nabla\theta(\rho))_{ij} = \frac{\partial\theta_j(\rho)}{\partial\rho_i} \qquad 1 \le i \le m, \qquad 1 \le j \le k \quad .$$

The parametrization can always be chosen so that $\nabla\theta(\rho)$ is of rank m, and we assume this is so.

Define the *information matrix* $J(\rho)$ at $\rho = \rho_0$ by

$$(2) \qquad J(\rho_0) = (\nabla\theta(\rho_0))(\Sigma(\theta_0)(\nabla\theta(\rho_0))' \quad .$$

If $\{p_\theta\}$ is a minimal exponential family then $\Sigma(\theta_0)$ is non-singular, and so $J(\rho_0)$ is then a positive definite m×m symmetric matrix. The chain rule and the basic differentiation formula 2.3(2) yield two alternate expressions for J; namely

$$(3) \qquad (J(\rho_0))_{ij} = E_{\theta_0}\left(\frac{\partial \log p_{\theta(\rho_0)}(X)}{\partial\rho_i} \; \frac{\partial \log p_{\theta(\rho_0)}(X)}{\partial\rho_j}\right)$$

$$= -E_{\theta_0}\left(\frac{\partial^2 \log p_{\theta(\rho_0)}(X)}{\partial\rho_i \, \partial\rho_j}\right) \quad .$$

The first expression of (3) is, of course, the usual definition of J in contexts more general than differentiable subfamilies.

If T is a statistic taking values in R^ℓ let

$$(4) \qquad e(\rho) = e_T(\rho) = E_{\theta(\rho)}(T) \quad .$$

Suppose $\theta_0 \in N^\circ(||T||)$. Then $E_\theta(T)$ and its derivatives exists at θ_0 by Corollary 2.6. The chain rule then yields

(5) $$\nabla e(\rho_0) = (\nabla\theta(\rho_0))(\nabla E_{\theta_0}(T)) \quad .$$

(The preceding formulation of course includes the case where $\{p_\theta\}$ is a full exponential family. Simply set $\rho = \theta$ so that $\theta(\rho) \equiv \theta$. In that case $J(\rho_0) = \mathcal{I}(\theta_0)$ and $\nabla e(\rho_0) = \nabla E_{\theta_0}(T)$.)

4.4 Theorem (Information inequality)

Let $\{p_\theta: \theta \in \Theta\}$ be a differentiable subfamily of a canonical exponential family with $\theta_0 = \theta(\rho_0)$, as above. Let T be an ℓ-dimensional statistic. Suppose $\mathcal{I}_{\theta_0}(T)$ exists. Then $e(\rho) = E_{\theta(\rho)}(T)$ exists and is differentiable on a neighborhood of ρ_0, and the covariance matrix of T satisfies

(1) $$\mathcal{I}_{\theta_0}(T) \geq (\nabla e(\rho_0))' J^{-1}(\rho_0)(\nabla e(\rho_0)) \quad .$$

Proof. $\theta_0 \in N^\circ(||T||)$ by Proposition 4.2. Now apply the Cauchy-Schwarz inequality 4.1(1) with $T_1 = T - E_{\theta_0}(T)$ and

(2) $$T_2(X) = \nabla \ln p_{\theta(\rho_0)}(X) = (\nabla\theta(\rho_0))\ (X - \xi(\theta_0)) \quad .$$

Then $B_{11} = \mathcal{I}_{\theta_0}(T)$,

(3) $$B_{22} = E(T_2\, T_2') = (\nabla\theta(\rho_0))\ \mathcal{I}(\theta_0)(\nabla\theta(\rho_0))' = J(\rho_0) \ ,$$

and

(4) $$B_{12} = E(T_1\, T_2') = (\nabla\theta(\rho_0))(\nabla E_{\theta_0}(T)) = \nabla e(\rho_0)$$

by 2.6(3) and 4.3(5). The Cauchy-Schwarz inequality says $B_{11} \geq B_{12}B_{22}^{-1}\, B_{21}$ which is the same as (1). ||

A useful feature of the form of Theorem 4.4 is the absence of any regularity condition on T other than the existence of $\mathcal{I}_{\theta_0}(T)$. Many other versions of the information inequality contain further assumptions about T (See e.g. Lehmann (1983, Theorem 7.3).) but these are superfluous here.

An information inequality like Theorem 4.4 is needed for applications of the following type.

4.5 Application (Karlin's Theorem on Admissibility of Linear Estimates)

The information inequality can sometimes be used to prove admissibility. In these situations other, more flexible, proofs can also be used, but the information inequality proof is nevertheless easy and revealing. The following result is due to Karlin (1958). The information inequality proof, due to Ping (1964), is a generalization of the first proof of this sort in Hodges and Lehmann (1951). See Lehmann (1983, p.271) for further references and details of the proof.

Theorem. Let $\{p_\theta\}$ be a full regular one-dimensional exponential family with $N = (\underline{\theta}, \bar{\theta})$, $-\infty \leq \underline{\theta} < \bar{\theta} \leq \infty$. Consider the problem of estimating $\xi(\theta) = E_\theta(X)$ under squared error loss. The risk of any (non-randomized) estimator δ is thus $R(\theta, \delta) = E_\theta((\delta(x) - \xi(\theta))^2)$. Then the linear estimator

$$(1) \qquad \delta_{\alpha,\beta}(x) = \alpha x + \beta$$

is admissible if $0 < \alpha \leq 1$ and if

$$(2) \qquad \int \exp(-\gamma\theta + \lambda\psi(\theta))\, d\theta$$

diverges at both $\underline{\theta}$ and $\bar{\theta}$, where γ, λ are defined by

$$(3) \qquad \alpha = \frac{1}{1+\lambda}, \qquad \beta = \frac{\gamma}{1+\lambda} \quad .$$

Proof. We consider here only the case $\beta = 0 = \gamma$. (See Exercise 4.5.1.) Fix α. Let δ be any estimator with finite risk. Let $b(\theta) = E_\theta(\delta(X)) - \alpha\xi(\theta)$. The information inequality yields

$$(4) \qquad R(\theta, \delta) \geq \frac{[(\alpha\xi(\theta) + b(\theta))']^2}{\xi'(\theta)} + (\xi(\theta)(1-\alpha) - b(\theta))^2$$

$$\geq \alpha^2\xi'(\theta) + 2\alpha b'(\theta) + (\xi(\theta)(1-\alpha) - b(\theta))^2$$

since $\xi(\theta) = E_\theta(X)$ and $\xi'(\theta) = J(\theta) = \mathrm{Var}_\theta\, X$. For $\delta_{\alpha,0}$

$$R(\theta,\ \delta_{\alpha,0}) \ = \ \alpha^2\xi'(\theta) + (1 - \alpha)^2\ \xi^2(\theta)\ . \tag{5}$$

Hence, if

$$R(\theta,\ \delta) \ \leq \ R(\theta,\ \delta_{\alpha,0}) \tag{6}$$

then

$$2b'(\theta) - 2\lambda\xi(\theta)\ b(\theta) + (1 + \lambda)\ b^2(\theta) \ \leq \ 0 \quad . \tag{7}$$

Let

$$K(\theta) \ = \ e^{\lambda\psi(\theta)}\ b(\theta) \quad .$$

Then (7) becomes

$$2K'(\theta) + (1 + \lambda)\ K^2(\theta)e^{\lambda\psi(\theta)} \ \leq \ 0 \quad . \tag{8}$$

Now, let $\theta_0 \in (a,\ b)$ and make the change of variables

$$t(\theta) \ = \ \int_{\theta_0}^{\theta} \exp(\lambda\psi(t))dt.$$

Correspondingly, define $k(t)$ by $k(t(\theta)) = K(\theta)$, so that (8) becomes

$$2k'(t) + (1 + \lambda)\ k^2(t) \ \leq \ 0 \tag{9}$$

where $-\infty < t < \infty$ by (2). The only solution of (9) for $t \in (-\infty,\ \infty)$ is $k \equiv 0$ since integration of (9) shows that for $t > t_1$ k is non-increasing and

$$k^{-1}(t) - k^{-1}(t_1) \ \geq \ (1 + \lambda)(t - t_1)/2 \quad ;$$

and hence $k(t_1) < 0$ is impossible. A similar inequality for $t < t_1$ shows that $k(t_1) > 0$ is also impossible. It follows that (6) implies $b \equiv 0$, which in turn implies $\delta = \delta_{\alpha,0}$ (a.e.(ν)) by completeness. This proves admissibility of $\delta_{\alpha,0}$. ||

It is generally conjectured that the condition 4.5(2) is necessary as well as sufficient for admissibility of $\delta_{\alpha,\beta}$. However only partial results are known in this connection. See Joshi (1969) and also Exercises 4.5.4, 4.5.5.

4.6 Further Developments

It is useful in considering asymptotic theory to have available a few further results concerning the information inequality. These results are sketched below; the proofs are left for exercises. These results have nothing to do specifically with exponential families but only require a setting in which the information inequality is valid. Nevertheless, for precision assume below the setting of Theorem 4.4, and let $S \subset R^m$ denote a (possibly large) open set on which $\Sigma_{\theta(\rho)}(T)$ exists. For convenience we consider below only estimation of ρ under the quadratic type loss function

$$(1) \qquad L(\rho, \delta) = (\delta - \rho)' J(\rho)(\delta - \rho) ,$$

and under a truncated version of this loss. (See (3) below.) For proof of the following assertions see Exercises 4.6.1 - 4.6.7 and Brown (1986).

Let h be an absolutely continuous probability density on S, supported on a compact subset $H \subset S$. Then the expected risk satisfies

$$(2) \qquad \int_H R(\rho, \delta)h(\rho)d\rho \geq m - \int_H \left(\frac{\nabla h(\rho)}{h(\rho)}\right)' J^{-1}(\rho)\left(\frac{\nabla h(\rho)}{h(\rho)}\right) h(\rho)\, d\rho \quad .$$

Note that the right side of this inequality is independent of δ, and thus provides a lower bound for the Bayes risk under the prior density h.

A natural truncation of the loss (1) is the function $\min(L(\rho, \delta), K)$. Generalizations of the information inequality and of (2), like those to be described below, can be stated for this natural truncation; however the statements and proofs are easier under a different truncation which is equally useful in asymptotics. This truncation will now be described.

Let $K > 0$. For $v \in R$ define

$$v_K = \begin{cases} -K & v < -K \\ |v| & v \le K \\ K & v > K \end{cases} .$$

For $v \in R^k$ define v_K to be the vector with coordinates $(v_K)_i = (v_i)_K$, $i=1,\ldots,k$. Now let

$$L_K(\rho, \delta) = (\delta - \rho)_K' J^{-1}(\rho)(\delta - \rho)_K \quad . \tag{3}$$

Let R_K denote the risk function corresponding to this truncated loss function.

If δ is an estimator of ρ, let

$$\delta_{(K)}(x; \rho) = \rho + (\delta(x) - \rho)_K \quad , \tag{4}$$

and

$$b_{(K)}(\rho) = E_\theta(\delta_{(K)}(X, \rho)) - \rho = e_{(K)}(\rho) - \rho \quad . \tag{5}$$

Let $\lambda_1(\rho) \ge \ldots \ge \lambda_m(\rho) > 0$ denote the ordered eigenvalues of $J(\rho)$. Let α be any number satisfying $0 < \alpha < 1$. Then

$$\left(1 + \frac{\alpha}{(1 - \alpha)\lambda_m K^2}\right) R_K(\rho, \delta) \tag{6}$$

$$\ge \alpha \operatorname{Tr}(J(\rho)(\nabla e_{(K)}(\rho))' J^{-1}(\rho)(\nabla e_{(K)}(\rho))) + \operatorname{Tr}(J(\rho)b_{(K)}(\rho)b_{(K)}'(\rho)) \ .$$

(Note: $\nabla e_{(K)}$ exists except possibly for a countable number of values of ρ. At these values interpret the right side of (6) as its lim sup; or use right (or left) partial derivatives in place of $\nabla e_{(K)}$, for these always exist.)

This inequality becomes more interesting as K gets large relative to $1/\lambda_m$, for then α can be chosen near 1 but so that $\frac{\alpha}{(1-\alpha)\lambda_m K^2}$ is small.

The inequality (6) leads to an inequality concerning the Bayes risk just as the usual information inequality leads to (2). With h as in (2)

(7) $$\left(1 + \frac{\alpha}{(1-\alpha)\lambda_m K^2}\right) \int_H R_K(\rho,\ \delta)h(\rho)d\rho$$

$$\geq\ \alpha m - \alpha^2 \int_H \left(\frac{\nabla h(\rho)}{h(\rho)}\right)' J^{-1}(\rho)\left(\frac{\nabla h(\rho)}{h(\rho)}\right) h(\rho)d\rho \quad .$$

The above bound, unlike (6), does not involve δ (through $e_{(K)}$).

UNBIASED ESTIMATES OF THE RISK

An unbiased estimate of the risk as a tool for proving inadmissibility of estimators first appears in Stein (1973), and has been widely exploited since then. The basic technique is embarassingly simple. It involves merely an integration by parts which succeeds because of the term $e^{\theta\cdot x}$ appearing in the exponential density. Here we describe the method and a few of the easier applications. For further (more complex) applications, see, for example, Berger (1980b), Berger and Haff (1981), and Haff (1983). Here is the heart of the method.

A function $t : R^k \to R$ is called *absolutely continuous* if $t(x_1,\ldots,x_k)$, is absolutely continuous in x_i, $i=1,\ldots,k$, when all x_j, $j\neq i$ are held fixed. Let $t_i' = \frac{\partial t}{\partial x_i}$.

4.7 Theorem

Let $s : R^k \to R$ be absolutely continuous. Assume

(1) $$\int |s(x)|e^{\theta\cdot x}\, dx < \infty \quad , \text{ and}$$

(2) $$\int |s_i'(x)|e^{\theta\cdot x}\, dx < \infty, \quad i = 1,\ldots,k.$$

Then

(3) $$\theta_i \int s(x)e^{\theta\cdot x}\, dx = -\int s_i'(x)e^{\theta\cdot x}\, dx \quad .$$

Proof. Set $i = 1$ for convenience. For almost every $(x_2,\ldots,x_k)$

(4) $$\int |s(x_1,\ x_2,\ldots,x_k)|e^{\theta\cdot x}\, dx_1 < \infty$$

and

(5) $$\int |s_1'(x_1,\ x_2,\ldots,x_k)|e^{\theta\cdot x}\, dx_1 < \infty$$

because of (1), (2). For any such $(x_2,\ldots,x_k)$ integration by parts yields

(6) $$\theta_1 \int s(x_1, x_2,\ldots,x_k)e^{\theta\cdot x}\, dx_1$$

$$= \lim_{B\to\infty} \theta_1 \int_{-B}^{B} s(x_1,x_2,\ldots,x_k)e^{\theta\cdot x}\, dx_1$$

$$= \lim_{B\to\infty} \left\{ -\int_{-B}^{B} s_1'(x_1,x_2,\ldots,x_k)e^{\theta\cdot x}dx_1 + \left[s(x_1,x_2,\ldots,x_k)e^{\theta\cdot x}\right]_{x_1=-B}^{B} \right\}$$

$$= -\int s_1'(x_1,x_2,\ldots,x_k)e^{\theta\cdot x}dx_1 + \liminf_{B\to\infty} \left[s(x_1,x_2,\ldots,x_k)e^{\theta\cdot x}\right]_{x_1=-B}^{B}$$

$$= -\int s_1'(x_1,x_2,\ldots,x_k)e^{\theta\cdot x}\, dx$$

by (2) and then (1). Integration over $x_2,\ldots,x_k$ then yields (3). ||

The assumptions (1) and (2) are slightly more stringent than necessary, and also can be given alternate forms. For example the assumption (5) together with

(7) $$\lim_{x_1\to\pm\infty} s(x_1, x_2,\ldots,x_k)e^{\theta\cdot x} = 0$$

for almost every $x_2,\ldots,x_k$ implies (4), and hence (3) when $i = 1$. Or, for example, when $k = 1$ a potentially useful result is the equality

(8) $$\int_0^\infty \theta s(x)e^{\theta x}\, dx = -\int_0^\infty s'(x)e^{\theta x} - s(0^+)$$

for absolutely continuous functions s having $\int|s'(x)|e^{\theta x}dx < \infty$ and $\lim_{x\to\infty} s(x)e^{\theta x} = 0$. However, the version of the theorem given above suffices for the usual applications.

Theorem 4.6 can be expressed in other forms which are more suggestive of its applications, as in the following two corollaries.

4.8 Corollary

Let $p_\theta(x)$ be a probability density on R^k (relative to Lebesgue measure) of the form

$$(1)\qquad p_\theta(x) = h(x)\exp(\theta\cdot x - \psi(\theta))$$

where $h \geq 0$ is absolutely continuous. Let $t : R^k \to R$ be absolutely continuous. Let $t_i' = \frac{\partial t}{\partial x_i}$. Then

$$(2)\qquad \theta_i E_\theta(t) = -E_\theta((t_i' + \frac{h_i'}{h} t))$$

provided both expectations in (2) exist.

Let $t : R^k \to R^k$ be absolutely continuous. Then

$$(3)\qquad E_\theta(\theta\cdot t) = -E_\theta(\nabla\cdot t + \frac{\nabla h}{h}\cdot t)$$

where $\nabla\cdot t = \sum_{i=1}^{k}\frac{\partial}{\partial x_i} t_i$, provided that

$$(4)\qquad E_\theta(|\frac{\partial t_i}{\partial x_i}|) < \infty$$

and

$$E_\theta(|\frac{\nabla h}{h}\cdot t|) < \infty, \qquad i=1,\ldots,k.$$

(In expressions (2), (3), (4) and similar expressions below define $\frac{h_i'}{h} = 0$ if $h = 0$.)

Proof. For (2) note that $\frac{\partial}{\partial x_i}(th) = (t_i' + \frac{h_i'}{h} t)h$ and apply Theorem 4.7. For (3) apply (2) with $i=1,\ldots,k$ and sum. ||

Remarks. Expression (2) immediately yields

$$(5)\qquad \theta E_\theta(t) = -E_\theta(\nabla t + t\,\frac{\nabla h}{h})$$

provided the expectations exist. (3) can also be derived directly from Green's theorem which implies (under suitable conditions) that

$$(6)\qquad \int s(x)(\nabla e^{\theta\cdot x})dx = -\int(\nabla\cdot s(x))e^{\theta\cdot x}\,dx \quad .$$

It can also be worthwhile to apply Theorem 4.7 repeatedly, as in the next proposition which is needed for Theorem 4.10.

4.9 Proposition

Let p_θ be as in Corollary 4.8. Assume that h_i' is also absolutely continous, and that

(1) $$E_\theta\left(\left|\frac{h_i'}{h}\right|\right) < \infty \quad ,$$

and

(2) $$E_\theta\left(\left|\frac{h_{ii}''}{h}\right|\right) < \infty \, , \qquad i=1,\ldots,k$$

(where $h_{ii}'' = \frac{\partial^2}{\partial x_i^2} h$). Then

(3) $$||\theta||^2 = E_\theta\left(\frac{\nabla^2 h}{h}\right)$$

(where $\nabla^2 h = \sum_{i=1}^{k} h_{ii}''$).

Proof. Apply Theorem 4.6 twice for each $i=1,\ldots,k$ and sum over i. ||

Combining the preceding results yields the following unbiased estimator of risk for squared error loss.

4.10 Theorem

Let $\{p_\theta\}$ be an exponential family whose densities are of the form 4.8(1) with h satisfying 4.9(1), (2). Let $\delta\colon R^k \to R^k$ be any absolutely continuous estimator of θ. Suppose

(1) $$E_\theta(||\delta||^2) < \infty$$

and

(2) $$E_\theta\left(\left|\delta_i' + \frac{h_i'}{h}\,\delta\right|\right) < \infty \, , \qquad i=1,\ldots,k \quad .$$

Then

(3) $$E_\theta(||\delta - \theta||^2) = E_\theta\left(||\delta||^2 - 2\left(\nabla \cdot \delta + \frac{\nabla h}{h} \cdot \delta\right) + \frac{\nabla^2 h}{h}\right)$$

Proof. Note that

$$E_\theta(||\delta - \theta||^2) = E_\theta(||\delta||^2 - 2\theta \cdot \delta + ||\theta||^2) \quad .$$

Now use 4.8(3) and 4.9(3) to arrive at (3). ||

Remarks. The left side of (3) is the risk function for squared error loss. As previously, we frequently use the notation $R(\theta, \delta)$ for a risk function when the loss function (here $||\delta - \theta||^2$) is clear from the context. The integrand of the right side of (3) is free of θ; hence this integrand is an unbiased estimate of $R(\theta, \delta)$. For most applications of (3) one actually needs only an unbiased estimate of $R(\theta, \delta_1) - R(\theta, \delta_2)$ where δ_1 and δ_2 are two given estimators. In that case, the term $||\theta||^2$, leading to $\frac{\nabla^2 h}{h}$ in (3), cancels. Assumption 4.9(2) is therefore not needed to arrive at an unbiased estimate of the form

$$(4) \qquad R(\theta, \delta_1) - R(\theta, \delta_2) = E_\theta(||\delta_1||^2 - ||\delta_2||^2 + 2(\nabla \cdot (\delta_1 - \delta_2) + \frac{\nabla h}{h} \cdot (\delta_1 - \delta_2))) \quad .$$

4.11 Application (James-Stein estimator)

The neatest application of Theorem 4.10 is to prove the minimaxity of the James-Stein estimator for a multivariate normal mean. (The original result in James and Stein (1961) uses a different method of proof.) Let X be k-variate normal, $k \geq 3$, with mean $\xi(\theta) = \theta$ and covariance I. Consider the problem of estimating ξ under squared error loss. The usual estimator $\delta_0(x) = x$ is minimax. However, when $k \geq 3$ it is not admissible. Let

$$(1) \qquad \delta(x) = (1 - \frac{r(||x||)}{||x||^2})x$$

where r is absolutely continuous, non-decreasing, and

$$(2) \qquad 0 \leq r(\cdot) \leq 2(k - 2) \quad .$$

Then

$$R(\theta, \delta) \leq R(\theta, \delta_0) = k \qquad (3)$$

Strict inequality holds in (3) except when $r \equiv 0$ or when $r \equiv 2(k - 2)$, as can be seen from (5) below.

The normal density is of the form 4.8(1) and $\frac{\nabla h}{h} = -x$. With δ as in (1)

$$||\delta_0||^2 - ||\delta||^2 + 2\frac{\nabla h}{h} \cdot (\delta_0 - \delta) = -\frac{r^2(||x||)}{||x||^2}$$

so that 4.10(4) yields

$$R(\theta, \delta_0) - R(\theta, \delta) = E_\theta\left(2\nabla \cdot \frac{r(||X||)X}{||X||^2} - \frac{r^2(||X||)}{||X||^2}\right) \qquad (4)$$

(It remains to check the regularity conditions needed for 4.10(4), and these will be discussed below.)

Observe that $\nabla \cdot \frac{x}{||x||^2} = \frac{k-2}{||x||^2}$. Hence (4) yields

$$R(\theta, \delta_0) - R(\theta, \delta) = E_\theta\left(\frac{r(||X||)}{||X||^2}(2(k-2) - r(||X||)) + 2\frac{r'(||X||)}{||X||}\right) \qquad (5)$$

The unbiased estimator of the risk which appears on the right of (5) is non-negative because of (2); hence (3) follows. The first estimator of James and Stein was of the form (1) with $r \equiv k - 2$, which is the best possible constant value of r. However, a better estimator (as also noted by James and Stein) is

$$\delta^+(x) = \left(1 - \frac{k-2}{||x||^2}\right)^+ x \qquad (6)$$

which corresponds to the choice

$$r(t) = \min(t^2, k-2) \quad .$$

See Exercise 4.11.1. See also Exercises 4.11.5, 4.17.5, and 4.17.6 for generalizations.

(It is also of interest to note that in general if

$\delta_i = \delta_{0i} + \gamma_i$, $i=1,\ldots,k$, then 4.10(4) yields

(7) $$R(\theta, \delta_0) - R(\theta, \delta) = E_\theta[\sum_{i=1}^{k} 2 \frac{\partial}{\partial x_i} \gamma_i - \gamma_i^2] \quad .$$

The integrand is formally the same as the Cramer-Rao lower bound (in which $b(\cdot)$ replaces $\gamma(\cdot)$). See 4.5(7) (with $\lambda = 0$) and Exercise 4.5.6. Hence the fact that the inequality

(8) $$\sum_{i=1}^{k} 2 \frac{\partial}{\partial x_i} \gamma_i - \gamma_i^2 \geq 0$$

has a non-trivial solution if and only if $k \geq 3$ leads to the proof of the fact that $\delta_0(x) = x$ is inadmissible if and only if $k \geq 3$.)

The regularity conditions stated in Theorem 4.10 are not always satisfied by an estimator of the form (1). (If, for example, $r(x) \equiv k-2$ then δ is not continuous at $||x|| = 0$.) Justification of (4) therefore requires a supplementary argument: suppose δ is an estimator of the form (1) with a specified $r(\cdot)$. Let δ_ε be the estimator with $r(\cdot)$ replaced by

(9) $$r_\varepsilon(||x||) = \min(||x||^2/\varepsilon \ , \ r(||x||)) \quad .$$

Then δ_ε satisfieds the conditions of Theorem 4.10 so that (4) holds for δ_ε. Passing to the limit as $\varepsilon \downarrow 0$ yields that (4) also holds for δ.

There is a very extensive literature concerning the problem of estimating a multivariate normal mean. For an introduction and some references consult Lehmann (1983, Chapter 4).

4.12 Remark

For discrete exponential families there is an analog of the unbiased estimates in 4.8 and 4.10 which involves difference operators instead of partial derivatives. These results are based on the deceptively simple equality

(1) $$\sum_{x=0}^{\infty} \lambda h(x)\lambda^x = \sum_{x=1}^{\infty} h(x-1)\lambda^x$$

They have been particularly useful for certain problems involving Poisson or negative binomial variables. See Hudson (1978), Hwang (1982), and Ghosh, Hwang, and Tsui (1983) for some theory and applications.

GENERALIZED BAYES ESTIMATORS OF CANONICAL PARAMETERS

We first define the concept of a generalized Bayes estimator in the current context and state some foundational results. Then we discuss estimation of the canonical parameter of an exponential family. Later in this chapter we discuss estimation of the expectation parameter, including the topic of conjugate priors for exponential families.

4.13 Definition

Let $\{p_\theta: \theta \in \Theta\}$ be an exponential family of densities. Let $\zeta: \Theta \to R^\ell$ be measurable. Let G be a non-negative (σ-finite) measure on Θ, locally finite at every $\theta \in \Theta$. G is called a *prior measure* on Θ. Let $S \subset R^k$. Then $\delta: S \to R$ is *generalized Bayes* on S (for estimating ζ under squared error loss) if

$$(1) \qquad \delta(x) = \frac{\int \zeta(\theta) p_\theta(x) G(d\theta)}{\int p_\theta(x) G(d\theta)} , \qquad x \in S ,$$

where both numerator and denominator exist for all $x \in S$. We say δ is *generalized Bayes* if it is generalized Bayes on S where $\nu(S^c) = 0$. We will use the symbol δ_G to denote the generalized Bayes procedure for G, when this exists.

If the loss is squared error loss --

$$(2) \qquad L(\theta, a) = ||a - \zeta(\theta)||^2 \qquad --$$

for estimating $\zeta(\theta)$ and if the Bayes risk,

$$(3) \qquad B(G) = \inf_{\delta} B(G, \delta') = \inf_{\delta'} \int R(\theta, \delta')G(d\theta)$$

$$= \inf_{\delta'} \int E_\theta(L(\theta, \delta'(X))G(d\theta),$$

satisfies $B(G) < \infty$. Then by Fubini's theorem any *Bayes estimator* for G (i.e. one which minimizes $B(G, \delta)$) must also be generalized Bayes for G. One of the topics in which we shall be interested below is that of characterizing complete classes of procedures under squared error loss (2). Since L is strictly convex the nonrandomized procedures are a complete class. The following theorem is our main tool for proving complete class theorems. (In the current context a *complete class* is a set of procedures which contains all admissible procedures.)

4.14 Theorem

With $\{p_\theta\}$ and L as above every admissible procedure must be a limit of Bayes estimators for priors with finite support. More precisely, to every admissible procedure corresponds a sequence G_i of prior distributions supported on a finite set (and hence having finite Bayes risk) such that

$$(1) \qquad \delta_{G_i}(x) \to \delta(x) \qquad \text{a.e.}(\nu) \quad ,$$

where (as above) δ_{G_i} denotes the Bayes estimator for G_i.

Proof. This theorem is apparently "well known". Its proof is outside the intended scope of our manuscript. However, I do not know any adequate published reference for it, so a proof is given in the appendix to the monograph. See Theorem A12. Theorems 3.18 and 3.19 of Wald (1950) come close to the above theorem as do some comments in Sacks (1963) and in Le Cam (1955). ||

We now concentrate on estimation of the canonical parameter. In this case generalized Bayes estimators have a particularly convenient form, as described in the next theorem.

4.15 Theorem

Let $\{p_\theta\}$ be a canonical exponential family and let G be a prior measure on Θ for which the generalized Bayes procedure, δ_G for estimating θ

exists. Define the measure H by

$$H(d\theta) = e^{-\psi(\theta)}\, G(d\theta) \tag{1}$$

and (as usual) let $\lambda_H(x) = \int e^{\theta\cdot x} H(d\theta)$ denote its Laplace transform. Then δ_G satisfies

$$\delta_G(x) = \nabla \ln \lambda_H(x) = \nabla\psi_H(x) , \qquad x \in K^\circ . \tag{2}$$

(If $\nu(\partial K) = 0$ then, of course, (2) completely defines δ_G since $\nu((K^\circ)^{\text{comp}}) = \nu(\partial K) = 0$.)

Proof. By definition the generalized Bayes procedure is

$$\delta_G(x) = \frac{\int \theta\, e^{\theta\cdot x} H(d\theta)}{\int e^{\theta\cdot x} H(d\theta)} \qquad \text{a.e. } (\nu) \quad . \tag{3}$$

By assumption the integrals on the right of (3) exist a.e.(ν); hence $N_H \supset K^\circ$. The denominator exists on N_H, by definition, and by Theorem 2.2, the numerator exists on N_H° and is given by $\nabla\lambda_H(x)$. This proves (2). ||

If δ is only generalized Bayes on $S \subset K$ relative to G one clearly has an analogous representation of δ on S°, namely

$$\delta(x) = \nabla\psi_H(x) , \qquad x \in S^\circ \quad . \tag{4}$$

An interesting special consequence of the above is that if $k = 1$, and $|\delta(x) - x|$ is bounded, and $\lambda\delta(x)$ is generalized Bayes on K° for $0 < \lambda \leq 1$ then $\delta(x) = x + b$. See Meeden (1976).

The foundation for the following major theorem has been laid above and in Section 2.17. The first theorem of this type was proved by J. Sacks (1963) for dimension $k = 1$. Indeed Sacks claimed, but did not prove, validity of the result for arbitrary dimension. Brown (1971) proved the result for arbitrary dimensions when $\{p_\theta\}$ is a normal location family; and that proof was extended to arbitrary exponential families by Berger and Srinivasan

(1978). The proof below follows Brown and Berger-Srinivasan. The proof of Theorem 4.24 is somewhat more like Sacks' original proof.

4.16 Theorem

Let $\{p_\theta\}$ be a canonical k parameter exponential family. Then δ is admissible under squared error loss for estimating θ only if there is a measure H on $\bar{\Theta} \subset \bar{N}$ such that

$$(1) \qquad \delta(x) = \frac{\int \theta e^{\theta\cdot x} H(d\theta)}{\int e^{\theta\cdot x} H(d\theta)} = \nabla\psi_H(x) , \quad \text{for} \quad x \in K^\circ \quad \text{a.e.}(\nu) \quad .$$

Remarks. The expression (1) implicitly includes the condition $N_H \supset K^\circ$, so that both numerator and denominator in (1) are well defined for all $x \in K^\circ$.

If $H(\bar{\Theta} - \Theta) = 0$ so that $\Theta = \bar{\Theta} \subset N$, then one may define

$$(2) \qquad G(d\theta) = e^{\psi(\theta)} H(d\theta)$$

and rewrite (1) as

$$(3) \qquad \delta(x) = \frac{\int \theta p_\theta(x) G(d\theta)}{\int p_\theta(x) G(d\theta)} , \qquad x \in K^\circ \quad .$$

Thus δ is generalized Bayes on K° relative to G. This observation leads to Corollary 4.17 and to further remarks which appear after the corollary.

Proof. Let δ be admissible. By Theorem 4.14 there is a sequence of prior measures G_i, having finite support, such that $\delta_{G_i}(x) \to \delta_G(x)$ a.e.(ν). Let $x_0 \in K^\circ$ such that $\delta_{G_i}(x_0) \to \delta(x_0)$. Since G_i has finite support $\int e^{\theta\cdot x_0 - \psi(\theta)} G_i(d\theta) < \infty$. Let

$$(2) \qquad \tilde{H}_i(d\theta) = e^{-\psi(\theta)} G_i(d\theta) / \int e^{\zeta\cdot x_0 - \psi(\zeta)} G_i(d\zeta) \quad .$$

This is a normalized version of 4.15(1), so, letting $\psi_i = \psi_{\tilde{H}_i}$,

(3) $$\delta_{G_i}(x) = \nabla\psi_i(x) \quad .$$

Since $\int e^{x_0\cdot\theta}\tilde{H}_i(d\theta) = 1$ we assume without loss of generality the existence of a limiting measure H, for which $\tilde{H}_i \to H$ weak*. (Apply 2.16(iv) to the measure $e^{x_0\cdot\theta}\tilde{H}_i$ to get $e^{x_0\cdot\theta}\,\tilde{H}_i \to H^*$, say, and let $H = e^{-x_0\cdot\theta}\,H^*$.) Let $x' \in K°$ such that 4.14(1) holds at x'. Then there is a finite set $S \subset K°$ such that 4.14(1) holds on S and such that B = conhull S satisfies $x_0 \in B°$, $x' \in B°$. Let $x \in S$. Then

(4) $$\psi_i(x) - \psi_i(x_0) = \int_0^1 (x - x_0)\cdot\nabla\psi_i(x_0 + \rho(x - x_0))d\rho$$

$$\leq (x - x_0)\cdot\nabla\psi_i(x) \leq ||x - x_0||\ ||\delta_i(x)||$$

by Corollary 2.5. (Note that $\psi_i(x_0) \equiv 0$.) It follows that

(5) $$\limsup_{i\to\infty}\ \sup_{x\in S}\psi_i(x) = \sup_{x\in S}||\delta(x)||\ ||x - x_0|| < \infty \quad .$$

This is the principle assumption of Theorem 2.17, which now implies the existence of a subsequence H_i, and a limiting measure, which must be H, such that $\psi_i(x) \to \psi_H(x)$, $x \in B°$, and also $\nabla\psi_i(x) \to \nabla\psi_H(x)$, $x \in B°$, by 2.17(5). Since $\nabla\psi_i(x') = \delta_i(x') \to \delta(x')$ we have

(4) $$\delta(x') = \nabla\psi_H(x') \quad .$$

This proves (1) since x' is an arbitrary point of $K°$ satisfying 4.14(1), and since 4.14(1) is satisfied a.e.(ν). ||

<u>4.17 Corollary</u>

Suppose Θ is closed in R^k and

(1) $$\nu(\partial K) = 0 \quad .$$

Then the generalized Bayes procedures form a complete class.

Proof. As noted the admissible procedures are a (minimal) complete class. If δ is admissible then for some prior measure H on $\Theta = \bar{\Theta}$

(2) $$\delta(x) = \frac{\int \theta e^{\theta \cdot x} H(d\theta)}{\int e^{\theta \cdot x} H(d\theta)} \quad \text{a.e.}(\nu)$$

by 4.16(1) and (1), above. Let $G(d\theta) = e^{\psi(\theta)} H(d\theta)$ as in 4.16(2) to get the desired representation,

(3) $$\delta(x) = \frac{\int \theta p_\theta(x) G(d\theta)}{\int p_\theta(x) G(d\theta)} \quad \text{a.e. } (\nu) \quad . \qquad ||$$

Remarks. If ν is dominated by Lebesgue measure then (1) holds since the Lebesgue measure of the boundary of any convex subset of R^k is zero. (To see this note that if C is bounded and convex with $0 \in \text{int}C$ then

$$\partial C = \bigcap_{i=1}^{\infty} [(1 + \tfrac{1}{i})C - (1 - \tfrac{1}{i})C] = \bigcap_{i=1}^{\infty} C_i \text{ , say, where (as usual)}$$

$aC = \{x: \exists y \in C, \ x = ay\}$. See e.g. Rockafeller (1970). Then $\int_{aC} dx = a\int_C dx$ so that $\int_{\partial C} dx = \lim \int_{C_i} dx = \lim(\frac{1}{2i})\int_C dx = 0$. If C is unbounded apply the result for bounded C to $C \cap \{x: ||x|| < b\}$ and let $b \to \infty$.)

If $\nu(\partial K) \neq 0$ then there are, in general, admissible procedures which are not generalized Bayes. See Exercise 4.17.1. Similarly, if Θ is not closed in R^k there will again be admissible procedures which are not generalized Bayes, even when $\nu(\partial K) = 0$. See Exercise 4.17.2. When $\Theta = N$ and the exponential family is regular then Θ is closed if and only if $N = R^k$. Hence when $\Theta \neq R^k$ one cannot assert that all admissible procedures are generalized Bayes. However, the representation 4.16(1) remains valid. This representation is qualitatively similar to a generalized Bayes representation and is generally as useful as one.

Not all estimators which can be represented in the form 4.17(3) or 4.16(1) are admissible. In fact, many are not. Nevertheless, representations of this form are valuable stepping-off points for general admissibility proofs. See Brown (1971, 1979).

The most conspicuous example of an inadmissible generalized Bayes estimator occurs in the problem of estimating a multivariate normal mean already discussed in 4.11. The usual estimator $\delta(x) = x$ is generalized Bayes, but when $k \geq 3$ it is not admissible. When $k \geq 3$ the positive part James-Stein estimator, defined in 4.11(6), dominates $\delta(x) = x$. However, the positive part James-Stein estimator cannot be generalized Bayes (see Example 2.9); hence is itself inadmissible. So far as I know the problem of finding an (admissible) estimator which dominates 4.11(6) remains open. However, theoretical and numerical evidence indicates that such an estimator cannot have a much smaller risk at any parameter point; hence 4.11(6) remains one of the many reasonable alternatives to $\delta(x) = x$ when $k \geq 3$. (See e.g. Berger (1982).)

GENERALIZED BAYES ESTIMATORS OF EXPECTATION PARAMETERS CONJUGATE PRIORS

The statistical problem of estimating the expectation parameter $\xi(\theta)$, is more often of interest than that considered previously, of estimating the natural parameter. (Of course for normal location families the two problems are identical.) In this case, too, there is a representation theorem for generalized Bayes procedures and a complete class theorem based on a representation similar to that of generalized Bayes. (In some (not fully developed) sense the generalized Bayes representation available here is dual to that in the preceding section -- the differentiation operator is with respect to θ and appears inside the integral sign instead of being with respect to x and appearing outside it.) Both these main results are somewhat more limited than those for estimating θ; but are nevertheless useful.

A new feature of considerable statistical interest appears here. The linear estimators are (generalized) Bayes for the conjugate (generalized) priors. This result is presented first; the conjugate priors are defined in

4.18 and the existence and linearity of their (generalized) Bayes procedures is proved in Theorem 4.19.

4.18 Definition

Prior measures having densities relative to Lebesgue measure of the form

$$(1) \qquad g(\theta) = C\, e^{\theta\cdot\gamma-\lambda\psi(\theta)} \qquad \gamma \in R^k\,, \quad \lambda \geq 0 \qquad ,$$

are called conjugate prior measures. Note that if the prior is of the form (1) then the posterior distribution, calculating formally, has the same general form, with new parameters $\gamma + x$ and $\lambda + 1$. For a sample of size n the parameters become $\gamma + s_n = \gamma + \sum_{i=1}^{n} x_i$ and $\lambda + n$. (Note in (1) that $g = 0$ if $\theta \notin N$ since then $\psi(\theta) = \infty$.)

Arguments resembling those in the following proof show that the conjugate prior measure is finite, and hence can be normalized to be a prior probability distribution if and only if

$$(2) \qquad \lambda > 0 \quad \text{and} \quad \gamma/\lambda \in K^{\circ} \quad .$$

See Exercise 4.18.1.

For estimating $\xi(\theta) = E_\theta(X)$, under squared error loss, the Bayes procedures for conjugate priors are linear in x. This fact (often under extraneous regularity conditions) has been known for decades. See, for example, De Groot (1970, Chapter 9) and Raiffa and Schlaiffer (1961). The following precise statement and its converse first appeared in Diaconis and Ylvisaker (1979). (See Exercise 4.19.1 for a statement of the converse.)

4.19 Theorem

Let $\{p_\theta\}$ be a regular canonical exponential family and let $g(\theta)$ be a conjugate prior density as defined by 4.18(1). Then the generalized Bayes procedure for estimating $\xi(\theta)$ exists on the set

(1) $$S = \{x : \delta(x) = \frac{x + \gamma}{\lambda + 1} \in K^\circ\}$$

and has the linear form

(2) $$\delta(x) = \frac{x}{\lambda + 1} + \frac{\gamma}{\lambda + 1} = \alpha x + \beta \quad , \qquad x \in S \quad .$$

Remarks. If $\nu(S^c) = 0$ then δ is generalized Bayes. If $0 \in K$ this always occurs for $\gamma = 0$, $\lambda > 0$. It occurs for $\gamma = 0$, $\lambda = 0$ if (and only if) $\nu(\partial K) = 0$. It can occur for other values of γ, λ as well.

If $x \notin S$ then the generalized Bayes procedure does not exist at x since $\int e^{\theta \cdot x - \psi(\theta)} g(\theta) d\theta = \infty$. See Exercise 4.19.1.

For the relation between the condition that $\nu(S^c) = 0$, so that δ is generalized Bayes, and Karlin's condition, 4.5(2), see Exercise 4.19.2.

Proof. Let $x \in S$. The generalized Bayes procedure at x, if it exists, has the form

(3) $$\delta(x) = \frac{\int (\nabla\psi(\theta)) \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta)) d\theta}{\int \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta)) d\theta}$$

because of the form of g and of p_θ, and because $\xi(\theta) = \nabla\psi(\theta)$ on N and $g(\theta) = 0$ for $\theta \notin N$.

If the integrals in the numerator and denominator of (3) exist then Green's theorem in the form of 4.7(3) yields

(4) $$(x + \gamma) \int \exp((x + \gamma) \cdot \theta - (\lambda + 1)\psi(\theta)) d\theta$$

$$= (\lambda + 1) \int (\nabla\psi(\theta)) \exp((x + \gamma) \cdot \theta - (\lambda + 1)\psi(\theta)) d\theta \quad .$$

Rearranging terms in (4) yields (2). It remains only to verify that the numerator and denominator of (3) exist.

Let $z = \frac{x+\gamma}{\lambda+1}$. $z \in K^\circ$ since $x \in S$.

Hence

$$(5) \qquad \liminf_{||\theta||\to\infty} \frac{\psi(\theta) - \theta \cdot z}{||\theta||} > 0$$

by 3.5.2(1) (or by 3.6(3) and translation of the origin). It follows that for some $\varepsilon > 0$

$$(6) \qquad \exp\left((x + \gamma) \cdot \theta - (\lambda + 1)\psi(\theta)\right) = O(e^{-\varepsilon||\theta||}) \quad .$$

This proves existence of the integral in the denominator of (3).

Now consider $\xi_1 = \frac{\partial\psi}{\partial\theta_1}$ on N. For simplicity of notation below, let $\xi_1(\theta) = 0$ if $\theta \notin N$. Fix $\theta_2,\ldots,\theta_k$. $\xi_1(\theta_1,\theta_2,\ldots,\theta_k)$ is monotone in θ_1 for $\theta \in N$. Thus for some $q = q(\theta_2,\ldots,\theta_k) \in R$, $\xi_1(\theta_1,\theta_2,\ldots,\theta_k) \le 0$ for $\theta_1 < q$ and $\xi_1(\theta_1,\theta_2,\ldots,\theta_k) \ge 0$ for $\theta_1 > q$. Hence

$$(7) \qquad \int |\xi_1(\theta_1,\theta_2,\ldots,\theta_k)| \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))d\theta_1$$

$$= \lim_{B\to\infty} \int_{-B}^{q} - \xi_1(\theta_1,\theta_2,\ldots,\theta_k) \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))d\theta_1$$

$$+ \lim_{B\to\infty} \int_{q}^{B} \xi_1(\theta_1,\theta_2,\ldots,\theta_k) \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))d\theta_1 \ .$$

The function $\exp(-(\lambda+1)\psi(\theta_1,\theta_2,\ldots,\theta_k))$ is absolutely continuous in θ_1 since $\{p_\theta\}$ is regular. (If $\{p_\theta\}$ were not regular there could be a discontinuity at the boundary of N.) Let $\theta_q = (q(\theta_2,\ldots,\theta_k), \theta_2,\ldots,\theta_k)$. Ordinary integration by parts yields

$$(8) \qquad \lim_{B\to\infty} \int_{-B}^{q} - \xi_1(\theta_1,\theta_2,\ldots,\theta_k) \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))d\theta_1$$

$$= \lim_{B\to\infty} \Big\{-(x_1+\gamma_1) \int_{-B}^{q} \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))d\theta_1$$

$$+ \left[\exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))\right]_{\theta_1=-B}^{q} \Big\}$$

$$= -(x_1+\gamma_1) \int_{-\infty}^{q} \exp((x+\gamma) \cdot \theta - (\lambda+1)\psi(\theta))d\theta_1 + \exp((x+\gamma)\cdot\theta_q - (\lambda+1)\psi(\theta_q))$$

by (6). Note that (again by (6))

(9) $$\exp((x+\gamma)\cdot\theta_q - (\lambda+1)\psi(\theta_q)) = O(\exp(-\varepsilon \sum_{j=2}^{k} \theta_j^2)) \quad .$$

Reasoning similarly for the second integral on the right of (7), integrating both integrals over $\theta_2,\ldots,\theta_k$, and using (9) yields

(10) $$\int_{R^k} |\xi_1(\theta)| \exp((x+\gamma)\cdot\theta - (\lambda+1)\psi(\theta))d\theta < \infty \quad .$$

Finally, the identical reasoning on ξ_i, $i=1,2,\ldots,k$, shows that

$$\int (||\nabla\psi(\theta)||) \exp((x+\gamma)\cdot\theta - (\lambda+1)\psi(\theta))d\theta < \infty \quad ,$$

which verifies that the numerator of (3) exists. As noted previously, this completes the proof. ||

4.20 Application

For a given k-parameter exponential family $\{p_\theta\}$ the conjugate prior distributions, $\{g_{\gamma,\lambda}\}$, say, form a (k+1)-parameter exponential family with canonical statistics $\theta_1,\ldots,\theta_k$, $-\psi(\theta)$. This (k+1)-parameter family is minimal except when $\psi(\theta)$ is a linear function of θ. This linearity occurs when p_θ is the $\Gamma(\alpha, \sigma)$ family with known σ, and in certain multivariate generalizations of this univariate example.

Many familiar exponential families are the conjugate families of prior distributions for other familiar exponential families of distributions. (Conjugate prior measures which are not finite then appear as limits of these distributions.) For example, the $N(\gamma, \lambda^{-1}I)$ distributions are conjugate to the $N(\mu, I)$ family. The proper conjugate prior distributions for the $\Gamma(\alpha, \frac{1}{(-\theta)})$ family (α known, $\theta < 0$) are those of $-\Theta$ where $\Theta \sim \Gamma(\lambda\alpha, -\gamma)$, $\gamma < 0$, $\lambda > 0$. The proper conjugate priors for the $P(e^\theta)$ family have density

(1) $$g_{\gamma,\lambda}(\theta) = e^{\gamma\theta-\lambda e^\theta}, \qquad \gamma < 0, \qquad \lambda \geq 0$$

with respect to Lebesgue measure on $(-\infty, \infty)$. Thus the density of $\xi = e^\theta$ is

$\Gamma(-\gamma, 1/\lambda)$. See also Exercise 5.6.3.

The basic representation theorem for generalized Bayes procedures is a simple consequence of Green's Theorem 4.7(3), and is an obvious extension of 4.19(4) in the proof of Theorem 4.19. The regularity conditions in the following statement may be modified as noted in the remark following the theorem.

4.21 Theorem

Let $\{p_\theta\}$ be a regular canonical exponential family and let G be a prior measure on Θ. Suppose G has a density, g, with respect to Lebesgue measure. Suppose $g(\theta)e^{-\psi(\theta)}$ is absolutely continuous on R^k. Assume for $x \in S$

(1) $$\int e^{\theta\cdot x-\psi(\theta)}\, g(\theta)d\theta < \infty ,$$

(2) $$\int ||\nabla g(\theta)||\, e^{\theta\cdot x-\psi(\theta)}\, d\theta < \infty ,$$

and

(3) $$\int ||\nabla\psi(\theta)||\, g(\theta)e^{\theta\cdot x-\psi(\theta)}\, d\theta < \infty .$$

Then the generalized Bayes procedure, δ, for estimating $\xi(\theta)$ under squared error loss, exists on S and is given by the formula

$$\delta(x) = x + \frac{\int(\nabla g(\theta))e^{\theta\cdot x-\psi(\theta)}\, d\theta}{\int g(\theta)e^{\theta\cdot x-\psi(\theta)}\, d\theta} .$$

Remarks. If $\nu(S^c) = 0$ then, of course, the unrestricted generalized Bayes procedure exists and is given by (4).

Conditions (1) and (2) are of course necessary for the representation (4) to make sense. Condition (3) is necessary in order that the generalized Bayes estimator be well defined. However it can often be deduced as a consequence of (2) and so then need not be checked directly. Suppose

(5) $$g(\theta)e^{\theta\cdot x-\psi(\theta)} \le h(||\theta||)$$

for some function $h(\theta)$ satisfying

(5')
$$\int_0^\infty t^{k-1}\, h(t)dt \ < \ \infty \ .$$

Then (1) is satisfied, and condition (2) implies condition (3). See Exercise 4.21.1.

The representation (4) is exploited in Brown and Hwang (1982) as the starting point for a proof of admissibility of generalized Bayes estimators under certain (important) extra regularity conditions.

Proof. Conditions (1), (2), and (3) justify use of the integration by parts formula 4.7(3), which yields

(6)
$$\int x(g(\theta)e^{-\psi(\theta)})e^{\theta\cdot x}\, d\theta \ = \ \int (-\nabla g(\theta) + g(\theta)\nabla\psi(\theta))e^{\theta\cdot x-\psi(\theta)}\, d\theta \quad .$$

Rearranging terms (each of which exists by (1), (2), (3)) yields (4). ||

We now turn to the complete class theorem comparable to Theorem 4.16. The result proved below applies only to one parameter exponential families. It appears to us that there exists a satisfactory multiparameter analog of this result which, however, is somewhat more complex to state (and to prove). We hope to present this multiparameter extension in a future manuscript.

As with Theorem 4.16 the representation of admissible procedures involves a ratio of integral expressions similar to the formula for a generalized Bayes estimator. Again, under certain additional conditions, this representation reduces to precisely that of a generalized Bayes procedure. A new complication appears in the integral representation below. It applies only on an interval I_δ whose definition involves $\delta(\cdot)$ itself. (See 4.24(1).) However, as explained in the remarks following the theorem, the values of $\delta(x)$ for $x \notin \bar{I}$ are uniquely specified by monotonicity considerations. Hence the theorem actually describes exactly the values of $\delta(x)$ except for at most two points -- the endpoints of I_δ . In this sense the complication presented by the presence of I_δ is just a minor nuisance.

We begin with a technical lemma.

4.22 Lemma

Let ν_n be a sequence of probability measures on R^1. Suppose for some $\zeta > 0$

(1) $$\liminf_{n\to\infty} \nu_n(\{x > K\}) > \zeta > 0$$

for all $K < \infty$. Let $\varepsilon > 0$. Suppose $\lambda_{\nu_n}(\varepsilon) < \infty$, $n=1,\ldots$. Then

(2) $$\lim_{n\to\infty} \frac{\int_K^\infty e^{\varepsilon x}\, \nu_n(dx)}{\lambda_{\nu_n}(\varepsilon)} = 1$$

for all $K < \infty$.

Remarks. The negation of (1) is the condition

(3) $$\lim_{K\to\infty} \liminf_{n\to\infty} \nu_n(\{|x| > K\}) = 0 \quad .$$

This is the usual necessary and sufficient condition for there to exist a subsequence n' and a non-zero limiting measure ν such that $\nu_{n'} \to \nu$.

The conclusion (2) can be paraphrased by saying that the sequence of probability measures $e^{\varepsilon x}\, \nu_n(dx)/\lambda_{\nu_n}(\varepsilon)$ sends all its mass out to $+\infty$.

Proof. Let $K < \infty$, $1 < m < \infty$. Then

(4) $$\frac{\int_K^\infty e^{\varepsilon x}\, \nu_n(dx)}{\int_{-\infty}^K e^{\varepsilon x}\, \nu_n(dx)} \geq \frac{\int_{mK}^\infty e^{\varepsilon x}\, \nu_n(dx)}{\int_{-\infty}^K e^{\varepsilon x}\, \nu_n(dx)}$$

$$\geq e^{\varepsilon(m-1)K}\, \nu_n(\{x > mK\}) \quad .$$

Now let $n \to \infty$ and $m \to \infty$ to find

$$\text{(5)} \qquad \liminf_{n\to\infty} \frac{\int_K^\infty e^{\varepsilon x}\, \nu_n(dx)}{\int_{-\infty}^K e^{\varepsilon x}\, \nu_n(dx)} = \infty \quad ,$$

which proves (2). ||

4.23 Theorem

Let $\{p_\theta : \theta \in \Theta\}$ be a regular exponential family on R^1. Consider the problem of estimating the expectation parameter, $\xi(\theta)$, under squared error loss. Let δ be an admissible estimator. Then, $\delta(\cdot)$ must be a non-decreasing function. Let

$$\text{(1)} \qquad I_\delta = \{x: \nu(\{y: y > x,\ \delta(y) \in K^\circ\}) > 0 \text{ and } \nu(\{y: y < x,\ \delta(y) \in K^\circ\}) > 0\}.$$

Then there exists a finite measure V on $\bar{\Theta}$ such that for all $x \in I$

$$\text{(2)} \qquad \delta(x) = \frac{\int \frac{\xi(\theta)}{1 + |\xi(\theta)|} e^{\theta x}\, V(d\theta)}{\int \frac{1}{1 + |\xi(\theta)|} e^{\theta x}\, V(d\theta)} \quad .$$

Remarks. In (2) the functions $\frac{\xi(\theta)}{1 + |\xi(\theta)|}$ and $\frac{1}{1 + |\xi(\theta)|}$ have the obvious interpretation on the boundary of N. (In other words, if $N = (a, b)$ then $\frac{\xi(b)}{1 + |\xi(b)|} = 1$, $\frac{\xi(a)}{1 + |\xi(a)|} = -1$, etc., since $\lim_{\theta\uparrow b} \xi(\theta) = \infty$, $\lim_{\theta\downarrow a} \xi(\theta) = -\infty$.)

By monotonicity of δ, I must be an open interval. Say $I = (\underline{i}, \bar{i})$, $-\infty \le \underline{i} < \bar{i} \le \infty$. Suppose $K^\circ = (\underline{k}, \bar{k})$, $-\infty \le \underline{k} < \bar{k} \le \infty$. Then $\underline{k} \le \underline{i}$ ($\bar{i} \le \bar{k}$, respectively) and, by monotonicity and the definition of I, $\delta(x) = \underline{k}$ for $\underline{k} \le x < \underline{i}$ ($\delta(x) = \bar{k}$ for $\bar{i} < x \le \bar{k}$). For $\underline{i} < x < \bar{i}$, $\delta(x)$ is defined by (2). Thus, the theorem fails to define $\delta(x)$ only for $x = \underline{i}$ if $-\infty < \underline{k} < \underline{i}$ or for $x = \underline{k}$ if $-\infty < \underline{k} = \underline{i}$, and, if $\bar{k} < \infty$, for $x = \bar{i}$ or $\bar{k}$ depending on whether $\bar{i} < \bar{k}$ or $\bar{i} = \bar{k}$. If ν, the dominating measure for $\{p_\theta\}$, is continuous then these two points have measure 0 and the theorem completely describes δ. Similarly, if $K = (-\infty, \infty)$ then irrespective of ν the theorem completely describes δ

If $V(\bar{N} - N) = V(\{a, b\}) = 0$ then (2) can be rewritten as

$$(3) \qquad \delta(x) = \frac{\int \xi(\theta)\, p_\theta(x) G(d\theta)}{\int p_\theta(x) G(d\theta)} , \qquad x \in I^\circ ,$$

where

$$G(d\theta) = \frac{e^{\psi(\theta)}}{1 + |\xi(\theta)|} V(d\theta) .$$

Thus, δ is then generalized Bayes on I in the ordinary sense. (This must, of course, occur if $N = R^1$.) When $N \neq R^1$ there may exist admissible procedures having representation (2) but not (3). See Exercise 4.24

Finally, note as with Theorem 4.16 that there are many inadmissible procedures satisfying (2). See for example Exercise 4.5.4.

Proof. If G is a prior density then the Bayes procedure (assuming it is well defined for $x \in K$) is given by the formula

$$(4) \qquad \delta_G(x) = \frac{\int \xi(\theta) e^{\theta x - \psi(\theta)} G(d\theta)}{\int e^{\theta x - \psi(\theta)} G(d\theta)} = \frac{\int \xi(\theta) e^{\theta x} H(d\theta)}{\int e^{\theta x} H(d\theta)}$$

where $H(d\theta) = ce^{-\psi(\theta)} G(d\theta)$. $\xi(\theta)$ is monotone on N. The family of densities $e^{\theta x}/\int e^{\theta x} H(d\theta)$ is an exponential family (with parameter x) relative to the dominating measure H. In particular, it has monotone likelihood ratio. Hence, δ_G is monotone non-decreasing by Corollary 2.22. (δ_G is actually strictly increasing unless G is concentrated on a single point.) All admissible procedures are (a.e.(ν)) limits of Bayes procedures by Theorem 4.14, and limits of monotone functions are monotone. Hence all admissible procedures must be monotone non-decreasing. (A different proof of a better result is contained in Brown, Cohen and Strawderman (1976).)

Let δ be admissible and let δ_{G_i} be the sequence promised in Theorem 4.14 having $\delta_{G_n} \to \delta$ a.e.(ν). Since all δ_{G_n} are monotone non-decreasing there is no loss of generality in assuming $\delta_{G_n}(x) \to \delta(x)$ for all $x \in K°$, and we do so below.

Assume $0 \in I \subset K°$. Define the probability measures,

$$(5) \qquad V_n(d\theta) = \frac{(1 + |\xi(\theta)|)\, e^{-\psi(\theta)}\, G_n(d\theta)}{\int (1 + |\xi(\theta)|)\, e^{-\psi(\theta)}\, G_n(d\theta)} .$$

Let $\varepsilon > 0$ such that $\varepsilon \in K°$. Then

$$(6) \qquad \delta_{G_n}(\varepsilon) = \frac{\int \frac{\xi(\theta)}{1 + |\xi(\theta)|} e^{\varepsilon\theta} V_n(d\theta)}{\int \frac{1}{1 + |\xi(\theta)|} e^{\varepsilon\theta}\, V_n(d\theta)} .$$

Suppose for some $\zeta > 0$

$$(7) \qquad \liminf_{n\to\infty} V_n(\{\theta > K\}) > \zeta > 0 \quad \text{for all} \quad K < \infty .$$

Let θ_0 be the unique value such that $\xi(\theta_0) = 0$, and let $K > \theta_0$. The function $\frac{\xi(\theta)}{1 + |\xi(\theta)|}$ is increasing for $\theta > \theta_0$. Apply Lemma 4.23 to get

$$(8) \qquad \frac{\int \frac{\xi(\theta)}{1 + |\xi(\theta)|} e^{\varepsilon\,\theta} V_n(d\theta)}{\int e^{\varepsilon\theta}\, V_n(d\theta)} \geq \frac{-\int_{-\infty}^{K} e^{\varepsilon\theta}\, V_n(d\theta) + \frac{\xi(K)}{1 + \xi(K)} \int_K^{\infty} e^{\varepsilon\theta} V_n(d\theta)}{\int e^{\varepsilon\theta}\, V_n(d\theta)}$$

$$\to \frac{\xi(K)}{1 + \xi(K)} .$$

Similarly, $\frac{1}{1 + |\xi(\theta)|}$ is decreasing for $\theta > \theta_0$ so that

$$(9) \qquad \frac{\int \frac{1}{1 + |\xi(\theta)|} e^{\varepsilon\theta}\, V_n(d\theta)}{\int e^{\varepsilon\theta}\, V_n(d\theta)} \leq \frac{\int_{-\infty}^{K} e^{\varepsilon\theta}\, V_n(d\theta) + \frac{1}{1 + \xi(K)} \int_K^{\infty} e^{\varepsilon\theta}\, V_n(d\theta)}{\int e^{\varepsilon\theta}\, V_n(d\theta)}$$

$$\to \frac{1}{1 + \xi(K)} .$$

Substitute (8) and (9) into the formula, (6), for $\delta_{G_n}(\varepsilon)$ and let $K \to \infty$ to find

$$\delta(\varepsilon) \;=\; \lim_{n\to\infty} \delta_{G_n}(\varepsilon) \;\geq\; \lim_{K\to\infty} \xi(K) \;. \tag{10}$$

This holds for all $\varepsilon > 0$ with $\varepsilon \in K°$. It follows from (10) that $0 \notin I$, contrary to assumption. Hence (7) must be false. A symmetric argument shows that $\liminf_{n\to\infty} V_n(\{\theta < -K\}) > \zeta > 0$ is also impossible. Hence

$$\lim_{K\to\infty} \liminf_{n\to\infty} V_n(\{|\theta| > K\}) \;=\; 0 \;. \tag{11}$$

By translating the origin the same argument can be applied at any $x \in I \subset K°$. The conclusion is that $x \in I$ implies

$$\lim_{K\to\infty} \liminf_{n\to\infty} \frac{\int_{|\theta|>K} e^{\theta x}\, V_n(d\theta)}{\int e^{\theta x}\, V_n(d\theta)} \;=\; 0 \;. \tag{12}$$

This is slightly more than is needed to apply the corollary of Theorem 2.17 stated in Exercise 2.17.2. ((12) implies 2.17.2(2) with the roles of θ and x interchanged so that $P_{n,x}(d\theta) = e^{\theta x}V_n(d\theta)/\ e^{\theta x}V_n(d\theta)$.) The conclusion of this exercise is that there exists a subsequence $\{n'\}$ and a limiting measure V on $\bar{\Theta}$ such that

$$e^{\theta x}V_{n'}(d\theta) \;\to\; e^{\theta x}V(d\theta) \quad \text{and} \quad \lambda_{V_{n'}}(x) \;\to\; \lambda_V(x)\;, \quad x \in I° \;. \tag{13}$$

Note that $V(R^k) = \lambda_V(0) = \lim \lambda_{V_{n'}}(0) = 1$. Since both $\frac{\xi(\theta)}{1 + |\xi(\theta)|}$ and $\frac{1}{1 + |\xi(\theta)|}$ are bounded continuous functions on $\bar{\Theta}$, (13) and (6) yield directly that for $x \in I$

$$\delta(x) \;=\; \lim_{n'\to\infty} \delta_{G_{n'}}(x) \;=\; \frac{\int \frac{\xi(\theta)}{1 + |\xi(\theta)|}\, e^{x\theta}\, V(d\theta)}{\int \frac{1}{1 + |\xi(\theta)|}\, e^{x\theta}\, V(d\theta)} \;.$$

This verifies (2), and completes the proof. ||

EXERCISES

4.1.1

(i) Prove the Cauchy-Schwarz inequality (4.1(1)) with B_{22}^{-} in place of B_{22}^{-1} when B_{22} is singular. (ii) Show the inequality remains valid when T_1, T_2 are respectively ($\ell\times s$) and ($m\times s$) matrix valued random variables. [(i) Reproduce the proof of Theorem 4.1 with B_{22}^{-} in place of B_{22}^{-1}; or rotate coordinates so that B_{22} is diagonal with diagonal entries $d_{ii} > 0$, $1 \le i \le r$, and $d_{ii} = 0$, $r+1 \le i \le m$, and apply 4.1(1) for the first r coordinates of T_2.]

4.2.1

Let ν be a measure on R^k and let T be a real valued statistic. Suppose $0 \in N^\circ$ and $E(T^2) < \infty$. Show for every $\varepsilon > 0$ there is a polynomial $p(x)$ such that $E((T - p)^2) < \varepsilon$. (In other words, the monomials $x_1,\ldots,x_k,x_1^2$, $x_1 x_2,\ldots$ form a complete basis for $L_2(\nu)$.) [For $k = 1$ (for simplicity) let $f_0 = 1$, $f_1 = a_{11}x + a_{10}$, $f_2 = a_{22}x^2 + a_{21}x + a_{20},\ldots$ be orthonormal functions in $L_2(\nu)$. Let $\alpha_i = E(Tf_i)$. Then $\Sigma\alpha_i^2 \le ET^2$ so that $g = \sum_{i=0}^{\infty} \alpha_i f_i \in L_2(\nu)$. $T - g \in L_2(\nu)$, $0 \in N^\circ(T - g)$ and $\lambda_{T-g}^{(j)}(0) = 0$, $j=0,1,\ldots$.]

4.3.1

Verify formulae 4.3(3) and 4.3(5).

4.3.2

Let p_θ be a full canonical exponential family and let $\xi = \xi(\theta)$ denote the expectation parameter. Show that relative to this parameter the information matrix is $J(\xi) = \Sigma^{-1}(\theta(\xi))$.

4.4.1

Let M be a fixed $\ell\times\ell$ positive semi-definite symmetric matrix. Write the information inequality for $E_{\theta_0}((T - \mu)' M(T - \mu))$ where T is an ℓ-dimensional statistic with mean μ and finite covariance at θ_0. [This is immediate from Theorem 4.4 and $(T - \mu)' M(T - \mu) = \mathrm{Tr}(M(T - \mu)(T - \mu)')$.]

4.4.2

Show that the information inequality 4.4(1) is an equality if and only if for some matrix A and vector b

(a) $$T(x) = A(\nabla\theta(\rho_0))X + b \quad .$$

[Show the Cauchy-Schwarz inequality is an equality if and only if T_2 is an affine transformation of T_1.]

4.4.3

Let $\{p_\theta: \theta \in \Theta\}$ be a differentiable subfamily and T an ℓ-dimensional statistic. Suppose $\Sigma_{\theta_0}(T)$ exists for some $\theta_0 \in \Theta$. Then the information inequality is an equality for all $\theta \in \Theta$ if and only if Θ is an affine subspace of N and T is an affine function of the canonical minimal sufficient statistic for the exponential family $\{p_\theta: \theta \in \Theta\}$. (That such a characterization holds under mild regularity conditions for a general family $\{p_\theta\}$ was proved in Wijsman (1973) and Joshi (1976).) [Use Exercise 4.4.2.]

4.4.4

Suppose $\{p_\theta\}$ is a canonical one-parameter exponential family. Show that when the information inequality is not an equality it can be improved to an inequality of the form:

(1) $$\mathrm{Var}_{\theta_0} T \geq \varepsilon'(\theta_0)M(\theta_0)\varepsilon(\theta_0)$$

where $\varepsilon(\theta)$ is the $j\times 1$ vector with

(2) $$\varepsilon(\theta)_i = \frac{\partial^i}{\partial\theta^i} e(\theta) \qquad i=1,\dots,j$$

and $M(\theta)$ is an appropriate $j\times j$ symmetric matrix, not depending on T. In fact, $M^{-1}(\theta)$ is the covariance matrix at θ of the vector with coordinates

(3) $$p_\theta^{-1} \frac{\partial^i}{\partial\theta_i} p_\theta \qquad i=1,\dots,j \quad .$$

(The inequality (1) with M^{-1} as in (3) is called a *Bhattacharya* inequality.

Such inequalities are valid also for full k parameter exponential families and for ℓ-dimensional statistics, as well as for differentiable subfamilies (ρ replaces θ in (1) - (3)). See e.g. Lehmann (1983, p.129). [A direct proof is possible which also yields the formula (3). An alternate proof assumes $\theta_0 = 0$, $\psi(\theta_0) = 1$ (w.l.o.g.) and uses Exercise 4.2.1 to write

$$\int (T(x) - \alpha_0)^2 \, \nu(dx) \geq \sum_{i=1}^{j} \alpha_i^2 = \sum_{i=1}^{j} \int T(x) f_i(x) \nu(dx). \quad]$$

<u>4.4.5</u>

Suppose $X_1, \ldots$ are i.i.d. observations from a differentiable exponential subfamily. Let N be a stopping time with $P_{\theta_0}(N < \infty) = 1$ and

(1) $$E_{\theta_0}(\exp(\varepsilon N)) < \infty \quad \text{for some} \quad \varepsilon > 0 \quad .$$

Let $S_n = \sum_{i=1}^{n} X_i$ and let $T(S_N, N)$ be a statistic for which $\mathcal{V}_{\theta_0}(T) < \infty$. Then

(2) $$\mathcal{V}_{\theta_0}(T) \geq (E_{\theta_0}(N))^{-1} (\nabla e(\rho_0))' \, J^{-1}(\rho_0)(\nabla e(\rho_0))$$

where $e(\rho) = E_{\theta(\rho)}(T(S_N, N))$. [Prove directly or use Exercise 3.12.2 (iii) and Theorem 4.4. The regularity condition (1) can be considerably relaxed or modified, but some condition on N is needed in general. See Simons (1980).]

<u>4.4.6</u>

(i) When $\{p_\theta\}$ is a full canonical exponential family and $E_{\theta_0}(T^2) < \infty$, the Bhattacharya inequalities 4.4.4(1) tend to equality in the limit as $j \to \infty$. (ii) If $\{p_\theta\}$ is an m-dimensional differentiable subfamily with $m < k$ then there are statistics T for which the appropriate Bhattacharya inequalities do not tend to equality as $j \to \infty$. [(i) Use Exercise 4.2.1 and proceed from the proof sketched in the hint in Exercise 4.4.4. (ii) Consider a curved exponential family in the canonical version 3.11(1), and let $T(x) = x_2 - x_1^2$.]

<u>4.5.1</u>

Prove the assertion in 4.5 when $\beta \neq 0$. [Let $Y = X - \gamma$. Apply 4.5

to yield αY as an admissible estimator of $\xi(\theta) - \gamma$. Hence $\alpha Y + \gamma$ is admissible for $\xi(\theta)$.]

4.5.2

Show the condition 4.5(2) implies $\delta_{\alpha,\beta}(x) = (\alpha x + \beta) \in K$ a.e.(ν). [The theorem would be false otherwise! But a direct proof not involving the theorem is also of interest. Use Lemma 3.5.]

4.5.3

Suppose (λ, γ) satisfies condition 4.5(2), $\lambda' < \lambda$, and either $\gamma \in K^\circ$ or ν is a discrete measure. Then (λ', γ) satisfies condition 4.5(2). If $\gamma \in \partial K = K - K^\circ$, and (λ_1, γ), (λ_2, γ) both satisfy 4.5(2), and $\lambda_1 < \lambda < \lambda_2$ then (λ, γ) satisfies 4.5(2).

4.5.4

Let $X \sim \Gamma(a, \sigma)$, a known, and consider the problem of estimating $\sigma = E(X)$ under squared error loss. (i) Using Karlin's theorem verify that $\delta_{\alpha,\beta}(x) = \alpha x + \beta$ is admissible if $\alpha = \frac{1}{a+1}$, $\beta = 0$ or if $\alpha \le \frac{1}{a+1}$, $\beta > 0$. (ii) Show that if α,β do not satisfy these conditions then $\delta_{\alpha,\beta}$ is inadmissible since there is an admissible linear estimator which is better.

4.5.5

Consider the one-parameter exponential family defined by 3.4(1) with $\theta_2 = -1$ and $\theta = \theta_1 \in (-\infty, 0)$. Consider the problem of estimating $\xi(\theta)$ under squared error loss. Let $\delta_{\alpha,\beta}$ be a linear estimator as in 4.5(1). Observe that condition 4.5(2) of Karlin's theorem is not satisfied at $\bar\theta = 0$. Show that $\delta_{\alpha,\beta}$ is inadmissible. [For the case $\alpha = 1$, $\beta = 0$ let

$$(1) \qquad \delta_c'(x) = \begin{matrix} x & x \le c \\ c + (x-c)/2 & x > c \end{matrix} \; .$$

Then $R(\theta, \delta_c') \le R(\theta, \delta_{1,0})$ for $\xi(\theta) \le c$ and, for $\xi(\theta) \ge c$, a crude bound yields

(2) $$R(\theta, \delta') \le (\tfrac{1}{4}) \operatorname{Var}_\theta(X) + (\tfrac{1}{4})(\xi(\theta) - c)^2 + \xi^2(\theta)$$

$$= \xi^3(\theta)/8 + (\xi(\theta) - c)^2/4 + \xi^2(\theta) \quad .$$

Hence for c sufficiently large $R(\theta, \delta') < R(\theta, \delta_{0,1}) = \xi^3(\theta)/2$ also when $\xi(\theta) \ge c$.]

4.5.6

Let $\{p_\theta\}$ be as in 4.5. Suppose it is desired to estimate $g(\theta) = \xi(\theta) + W'(\theta)$ under squared error loss. Show the estimator $\delta_{\alpha,\beta}$ is admissible if

(1) $$\int \exp(\lambda\psi(\theta) + (1 + \lambda)W(\theta) - \gamma\lambda(\theta)d\theta$$

diverges at both $\underline{\theta}$ and $\bar{\theta}$. [Define $b(\cdot)$ as in 4.5. 4.5(7) becomes

(2) $$2b'(\theta) - 2(\lambda\xi(\theta) + (1 + \lambda)W'(\theta))b(\theta) + (1 + \lambda)b^2(\theta) \le 0 \quad .]$$

(See Ghosh and Meeden (1977). Although an estimator $\delta_{\alpha,\beta}$ may be admissible here, it is not clear that it is desirable, whereas for the case $W \equiv 0$ of 4.5 these estimators are very natural.)

4.5.7

Let $\{p_\theta\}$ be a canonical two dimensional exponential family with $N = R^2$. Consider the problem of estimating $\xi(\theta)$ with squared error loss (so that $R(\theta, \delta) = E_\theta(||\delta(X) - \xi(\theta)||^2)$). Show that the estimator $\delta(x) = x$ is admissible. Apply this result when (X_1, X_2) are independent normal, independent Poisson, independent binomial, or the sample means from Von-Mises variables. [Using the bivariate information inequality leads to replacement of 4.5(7) by

(1) $$2\nabla \cdot b(\theta) + ||b(\theta)||^2 \le 0$$

where $\nabla \cdot b(\theta) = \sum_{i=1}^{2} \frac{\partial b_i(\theta)}{\partial \theta_i}$. If b satisfies (1) so does

(2) $$\bar{b}(\theta) = (2\pi)^{-1} \int_0^{2\pi} Q_\phi^{-1}\, b(Q_\phi \theta) d\phi$$

where

$$Q_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} .$$

$\bar{b}$ is spherically symmetric; hence can be written as $\bar{b}(\theta) = \beta(||\theta||)\theta$. Let $t = ||\theta||$. (1) becomes

(3) $$2k\beta(t) + 2t\beta'(t) + t^2\beta^2(t) \leq 0 \quad .$$

Now let $K(t) = t^2\beta(t)$ to get

(4) $$2K'(t) + K^2(t)/t \leq 0$$

in place of 4.5(8). (Note how the argument fails if $k > 2$!)] (Stein (1956), Brown and Hwang (1982, Corollary 4.1).)

<u>4.5.8</u>

Let $X \sim \Gamma(\alpha, \sigma)$, $\alpha > 0$ a specified constant. Consider the problem of estimating $\sigma = -\frac{1}{\theta}$ under the loss function

(1) $$L(\sigma, a) = \frac{a}{\sigma} - \ln(\frac{a}{\sigma}) - 1 \quad .$$

(See Chapter 5 for a natural interpretation of this loss. See also Exercises 4.11.3 and 4.11.4.) Let $\delta_0(x) = \frac{x}{\alpha}$ and let $\delta(x) = (1 + \phi(x))\delta_0(x)$ be any estimator. Let

(2) $$e(\theta) = E_\theta(\phi) \quad \text{and} \quad W(t) = t - \ln(1 + t) , \quad t > -1 \quad .$$

(i) Show that

(3) $$R(\theta, \delta) - R(\theta, \delta_0) \geq -\frac{\theta e'(\theta)}{\alpha} + W(e(\theta))$$

(ii) Use (3) to show that δ_0 is admissible among all estimators having $e(\theta) \leq B$ for all $\theta \in (-\infty, 0)$. (δ_0 is actually admissible with no restriction on δ. See Brown (1966).)

[(i) $R(\theta, \delta) - R(\theta, \delta_0) = E_\theta(-\theta X\phi(X) - \ln(1 + \phi(X)))$ $= -\theta e'(\theta)/\alpha + E_\theta(\phi(X) - \ln(1 + \phi(X)))$. For (ii) follow the pattern of the proof of Theorem 4.5. (It is also possible to use (3) to prove δ_0 is admissible with no restriction on δ.)]

4.6.1

Prove 4.6(2). [Use the information inequality to write

$$\int h(\rho)R(\rho, \delta)\, d\rho \geq m + \int\{2h(\rho)\, \mathrm{Tr}(\nabla b(\rho)) + h(\rho)\mathrm{Tr}(J(\rho)b(\rho)b'(\rho))\}d\rho \ .$$

Integrate by parts the first term in the integrand in order to get an integral whose integrand is a quadratic in $b(\rho)$ for each fixed ρ. Minimize this integrand for each ρ to get 4.5(2).] See Exercise 5.8.1 for a statistical application of 4.5(2).

4.6.2

In preparation for the proof of 4.6(6) prove the following facts: (i) For each K, $\nabla e_{(K)}(\rho)$ exists for all but at most a countable number of points, ρ.

Fix ρ_0, K for which $\nabla e_{(K)}(\rho_0)$ exists. Let $\delta^*(x) = \delta_{(K)}(x;\, \rho_0)$, $e^*(\rho) = E_{\theta(\rho)}(\delta^*(X))$, and $D = (d_{ij}) = \nabla e^*(\rho_0) - \nabla e_{(K)}(\rho_0)$. Show

(ii) $d_{ij} = 0$, $i \neq j$, and

(iii) $|d_{ii}| \leq P_{\theta(\rho_0)}(|X_i - \theta_i| \geq K)$.

Let $|D| = (|d_{ij}|)$ with d_{ij} as above and let $J = J(\rho_0)$ be symmetric positive definite with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m > 0$. Show

$$\text{(iv)} \quad \mathrm{Tr}(JDJ^{-1}D) \leq \frac{\lambda_1}{\lambda_m} \mathrm{Tr}|D| \leq \frac{\lambda_1 R_K(\rho_0,\delta)}{\lambda_m K^2} \ .$$

[Since $|d_{ii}| \leq 1$ the eigenvalues -- and hence diagonal elements -- of JDJ^{-1} all have magnitude at most $\frac{\lambda_1}{\lambda_m}$. Then $R_K = \mathrm{Tr}(J^{-1}E) \geq \frac{1}{\lambda_1}\mathrm{Tr}\, E$ where $E = E_{\rho_0}((\delta - \rho_0)(\delta - \rho_0)')$.]

4.6.3

Also in preparation for the proof of 4.6(6) prove the following matrix inequalities

(i) $\mathrm{Tr}(JA'J^{-1}A) \geq 0$ for any $(k\times k)$ positive definite symmetric J and any $(k\times k)$ matrix A.

(ii) $\mathrm{Tr}(J(A' + B')J^{-1}(A + B)) \geq \alpha\, \mathrm{Tr}(JA'J^{-1}A) - \frac{\alpha}{1-\alpha}\, \mathrm{Tr}(JB'J^{-1}B)$.

[(i) Diagonalize J (and J^{-1}) and then write out $\mathrm{Tr}(\cdot)$ as a sum of individual terms. (ii) follows from (i).]

4.6.4

Now prove 4.6(6). [Write the information inequality for δ^*. Substitute $\nabla e^*(\rho_0) = \nabla e_{(K)}(\rho_0) + D$ and use 4.6.2(iv) and 4.6.3(ii). (Note that both these inequalities are nearly trivial when $k = 1$, so in that case the overall proof is much simpler to follow.)]

4.6.5

The inequality 4.6(6) is never sharp (except sometimes in the limit as $K \to \infty$). To examine how far from sharp the inequality is compare R_K and the best lower bound from 4.6(6) in the case where $k = 1$, L is ordinary squared error loss, $X \quad N(\theta, 1)$, $\rho = \theta$, and $\delta(x) = ax$ $(0 < a \leq 1)$. [For $a = 1$, $K = 1$, I get $R_K = .516 \geq .250$ = best lower bound. For $a = 1$, $K = 3$ I get $R_K = .991 \geq .5625$ and for $a = 1$, $K = 10$ $R_K = .999+ \geq .891$.]

4.6.6

Prove 4.6(7). [See 4.6.1.]

4.6.7

Investigate the sharpness of (7) by comparing the Bayes risk for L_K and the bound on the right of 4.6(7) when $k = 1$, L is ordinary squared error loss, $X \sim N(\theta, 1)$, $\rho = \theta$, and h is a normal $(0, \sigma^2)$ density. (Note: h does not have compact support, but it can be shown (Exercise !) that the tails of h decrease fast enough so that 4.6(7) is still valid.) [When $K = \infty$

so that 4.6(7) reduces to 4.6(2) the Bayes risk is $\sigma^2/(1 + \sigma^2)$ and the lower bound is $(\sigma^2 - 1)/\sigma^2$. Thus even when $K = \infty$ the bound is not sharp, although it is asymptotically sharp as $\sigma^2 \to \infty$ also.]

4.11.1

Let δ denote the James-Stein estimator 4.11(1) with $r \equiv k - 2$ and let δ^+ denote the corresponding "positive part" estimator 4.11(6). Show that $R(\theta, \delta^+) < R(\theta, \delta)$. [Write $R(\theta, \delta) - R(\theta, \delta^+) = E_\theta(g(||X||^2))$. Note $S^-(g) = 1$ and $IS^-(g) = -1$, and (trivially) $E_0^+(g(||x||^2)) > 0$. Use Exercise 2.21.1.]

4.11.2

Suppose $X \sim N(\mu, \sigma^2 I)$ $(X \in R^k)$ and, independently, $V/\sigma^2 \sim \chi^2_m$. It is desired to estimate μ with squared error loss -- σ^2 is unknown. Let $k \geq 3$. Let $\hat{\sigma}^2 = V/m$ and

$$\delta(x) = \left(1 - \frac{s(||x||^2, \hat{\sigma}^2)}{||x||^2/\hat{\sigma}^2}\right) x$$

where $0 \leq s(\cdot) \leq 2(k-2)m/(m + 2)$ and $s(\cdot, \hat{\sigma}^2)$ is differentiable and non-decreasing for each value of $\hat{\sigma}^2$. Show that $\delta(x)$ is better than $\delta_0(x) = x$. [Assume (w.l.o.g.) that $\sigma^2 = 1$. Condition on $\hat{\sigma}^2$; apply 4.11(5) with $r(\cdot) = \hat{\sigma}^2 s(\cdot, \hat{\sigma}^2)$; and take the expectation over $\hat{\sigma}^2$. (A frequently recommended choice for s is $s(||x||^2, \hat{\sigma}^2) = \min(||x||^2/\hat{\sigma}^2, (k-2)m/(m + 2))$ corresponding to 4.11(6).]

4.11.3

Let $_i$ be independent $\Gamma(\alpha_i, \sigma_i)$ variables with α_i known, $i=1,\ldots,k$. Consider the problem of estimating $\sigma = (\sigma_1,\ldots,\sigma_k)$ with loss function $L(\sigma, a) = \Sigma\sigma_i(1 - a_i/\sigma_i)^2$. The best linear estimator for this problem is δ_0 with $\delta_{0i}(x) = x_i/(\alpha_i + 1)$. (i) When $k = 1$ this estimator is admissible. [Use Theorem 4.5.] (ii) for $k \geq 2$ define δ by

$$\delta_i(x) = x_i/(\alpha_i + 1) + (k-1)\alpha_i + 1/\sum_{j=1}^{k}(\alpha_j + 1)^3/x_j \quad . \tag{1}$$

Show that $R(\sigma, \delta) < R(\theta, \delta_0)$. (This is the easiest of several interesting related results in Berger (1980b).) [Let $\phi_i(x) = (\alpha_i + 1)(\delta_i(x) - \delta_{0i}(x))/x_i$. Using Corollary 4.7 show

(2) $$R(\sigma, \delta_0) - R(\sigma, \delta) =$$

$$-E[\Sigma \left(\frac{2X_i^2}{(\alpha_i + 1)^2} \frac{\partial}{\partial x_i} \phi_i(X) \frac{X_i \phi_i^2}{\alpha_i + 1} + \frac{2X_i^2 \phi_i(X)}{(\alpha_i + 1)^2} \frac{\partial}{\partial x_i} \phi_i(X) \right)]$$

since $\sigma_i(1 - a/\sigma_i)^2 = (a\sqrt{-\theta_i} - 1/\sqrt{-\theta_i})^2$. Then show the expectand on the right of (2) is negative. (Use the fact that $\frac{\partial}{\partial x_i} \phi_i(x) < 0$ to eliminate the terms involving $\phi_i \frac{\partial}{\partial x_i} \phi_i$.)]

<u>4.11.4</u>

Let $X_i \sim \Gamma(\alpha_i, \sigma_i)$, $\alpha_i > 0$ specified constants, $i=1,\ldots,k$, as in Exercise 4.11.3. Consider the loss function

(1) $$L(\sigma, a) = \sum_{i=1}^{k} (a_i/\sigma_i - \ln(a_i/\sigma_i) - 1) .$$

Define δ_0 by $\delta_{0i}(x) = x/\alpha_i$. (See Exercise 4.5.7.) Let $k \geq 3$ and define δ by $\delta_i(x) = (1 + \phi_i(x))\, \delta_{0i}(x)$ where

(2) $$\phi_i(x) = -\frac{c\alpha_i \ln x_i}{1 + \Sigma(\alpha_i \ln x_i)^2}$$

with $0 < c \leq 1$. Show that $R(\sigma, \delta) < R(\sigma, \delta_0)$, $\sigma > 0$. [The unbiased estimator of $R(\sigma, \delta) - R(\sigma, \delta_0)$ is

(3) $$\Sigma[(\frac{x_i}{\alpha_i}) \frac{\partial \phi_i}{\partial x_i} + \phi_i - \ln (1 + \phi_i)] .$$

(The following algebra can be simplified by changing variables in (3) to $y_i = \alpha \ln x_i$, $i=1,\ldots,k$.) Then show this is always positive, using the facts that $|\phi_i| \leq c/2$ and $t - \ln (1 + t) \leq 2t^2/3$ for $|t| \leq ½$. (You will see that values of c somewhat larger than 1 can also be used in (2).)] See Dey, Ghosh, and Srinivasan (1983).

(Change variables in Exercise 4.5.7(3) to $\sigma = -\frac{1}{\theta} = \xi(\theta)$, and compare with the i -th term in brackets in (3), above. This identity of expressions is analogous to that which occurs in the estimation of normal means with squared error loss. See 4.11(8).)

4.11.5

Let $X \sim N(\theta, I)$. Consider the problem of estimating $\theta \in R^k$ under squared error loss. Suppose for some $C < \infty$, $\varepsilon > 0$

(1) $$(\delta_1(x) - x) \cdot x \geq 2 - k + \varepsilon \quad \text{for} \quad ||x|| > C .$$

Then $\delta_1(x)$ is inadmissible.

[Let $\delta_2(x) = \delta_1(x) - \varepsilon[(||x|| - C)^+ \wedge 1] \dfrac{x}{||x||^2}$

and use 4.10(4).] (Note that this generalizes Example 4.11 since $\delta_1(x) = x$ satisfies (1) when $k \geq 3$.)

4.15.1

(i) Show that for estimating the natural parameter the correspondence between prior measures and their generalized Bayes procedures is one-one if Supp ν has a non-empty interior (i.e. show $\delta_G = \delta_H$ a.e.(ν) implies $G = H$). [Use Theorem 4.15 and Corollary 2.13.] (ii) Give an example to show that this unicity may fail if $(\text{Supp } \nu)^\circ = \phi$.

4.15.2

Show that every admissible estimator of θ under squared error loss satisfies the monotonicity condition

(1) $$(x_2 - x_1) \cdot (\delta(x_2) - \delta(x_1)) \geq 0 \quad \text{a.e.}(\nu \times \nu) .$$

[Use 4.14, 4.15, and 2.5. (Do not use 4.16(1) for this would not yield (1) for $x_i \in \partial K$.)]

4.16.1

Let $X \sim P(\lambda)$. Let $c_0 \le 0$. Show that the estimator $\delta(0) = c_0$, $\delta(x) = \ln x$, $x=1,2,\ldots$, is not an admissible estimator of the natural parameter $\theta = \ln \lambda$ under squared error loss. (δ is the "maximum likelihood estimate" of θ; see Chapter 5. Also, the squared error loss function $L(\theta, a) = (a - \theta)^2$ can be justified in its own right, or one can transform to $\lambda = e^\theta$ and let $b = e^a$. The loss then takes the form $(\ln b - \ln \lambda)^2$ $= (\ln (b/\lambda))^2 = L^*(\lambda, b)$. The inadmissibility result, above, then says also that $\delta^*(x) = x$ is an inadmissible estimator of λ under loss L^*. Losses of the form L^* appear naturally in scale invariant problems; see Brown (1968).)

[Use Theorem 4.16. If δ is of the form 4.16(1) then, by monotonicity,

$$\text{(1)} \qquad \ln [x] \le \frac{\lambda_H'(x)}{\lambda_H(x)} \le \ln ([x] + 1), \qquad x \ge 1 .$$

Hence $\lambda_H(x) \to \infty$ as $x \to \infty$ but $\lambda_H(x) = o(e^{\varepsilon x})$ as $x \to \infty$, $\forall\, \varepsilon > 0$. This is impossible by Lemma 3.5 and Exercise 3.5.1.]

4.17.1

Let $X \sim \text{Bin}(n, p)$, $n \ge 3$, and consider the problem of estimating the natural parameter $\theta = \ln (p/(1 - p))$ under squared error loss. Show that the procedure

$$\delta(x) = \begin{matrix} -1 & x = 0 \\ 0 & 1 \le x \le n-1 \\ 1 & x = n \end{matrix}$$

is admissible. But, δ is not generalized Bayes. (Note that Corollary 4.17 is not valid here because 4.17(1) is not satisfied. Of course, Theorem 4.16 is satisfied with H giving unit mass to the point $\theta = 0$.)

[Let δ' be another estimator. Suppose $\delta'(0) = -1 + \alpha$, $\alpha > 0$. Then

$$\lim_{\theta \to -\infty} |\theta|^{-1}(R(\theta, \delta') - R(\theta, \delta)) = \alpha > 0 .$$

Hence $R(\theta, \delta') \le R(\theta, \delta)$, $\forall\ \theta$, implies (i) $\delta'(0) \le -1$. Similarly (ii) $\delta'(n) \ge 1$. Among all procedures satisfying (i), (ii) δ uniquely minimizes $R(0, \delta)$. Hence δ is admissible. If δ were generalized Bayes the prior G would have to have support {0} by 2.5; but this would imply $\delta(0) = 0 = \delta(n)$.]

4.17.2

Let $Z \sim \Gamma(\alpha,\sigma)$ as in 4.17.3, below. (i) Show that the estimator $\delta_0(x) \equiv 0$ cannot be represented as a generalized Bayes estimator of $\theta = 1/\sigma$. (ii) For $\alpha < 2$ show δ_0 is admissible. (iii) For $\alpha > 2$ show δ_0 is inadmissible. [(ii) If $\delta \neq \delta_0$ then, for some $\epsilon > 0$, $R(\theta,\delta) > \epsilon\theta^{\alpha}$ as $\theta \to 0$. (iii) Let $\delta(x) = (\alpha-2)/x$.]

4.17.3

Let $Z \sim \Gamma(\alpha, \sigma)$, α known. Then the distributions of $X = -Z$ form an exponential family with natural parameter $\theta = 1/\sigma$. Consider the problem of estimating θ with squared error loss. Show that

$$\delta(x) = be^{x} \quad (= be^{-z}) \tag{1}$$

can be represented in the form 4.16(1). [Let H be a Poisson distribution! It can further be shown that δ is admissible when $\alpha \le 2$ since it uniquely minimizes

$$H(\{0\}) \limsup_{\theta\to 0} R(\theta, \delta)e^{\psi(\theta)} + \sum_{i=1}^{\infty} H(\{i\})R(i, \delta)e^{\psi(i)} \quad .] \tag{2}$$

4.17.4

Let $\{p_\theta\}$ be any exponential family with K compact (Binomial, Multinomial, Fisher, Von Mises, etc.). Show that $\delta(x) = x$ is an admissible estimator of θ under squared error loss. [Show that δ is Bayes for the prior distribution, G, with density $c \exp(\psi(\theta) - ||\theta||^2/2)$ and that $B(G) < \infty$. Admissibility then follows from basic decision theoretic results. See, e.g. Lehmann (1983, Theorem 3.1). Use Exercise 3.4.1 to verify that B(G) is finite (and also that G is finite).] Caution! $\delta(x) = x$ is not a very natural estimator of θ, in spite of its admissibility. Hence its use in this problem is not necessarily recommended (unless the prior G is indeed as above). If Supp ν is finite then $\delta(x) = x$ is a natural estimator of $\xi(\theta)$, and is admissible under squared error loss for estimating $\xi(\theta)$. See Exercise 4.5.5 and also Brown (1981b).

4.17.5

Let $X \sim P(\lambda)$. Consider the problem of estimating λ under loss function

$$(1) \qquad L(\lambda, a) = (\ln(a/\lambda))^2 \quad .$$

Show that estimator $\delta_1(x) = e^x$ is generalized Bayes, but not admissible. [The question is equivalent to asking whether the estimator $\delta(x) = x$ is generalized Bayes, or admissible for estimating the canonical parameter, θ, under squared error loss. Reason as in Exercise 4.17.4 to show $\delta(x) = x$ is generalized Bayes. However, for estimating θ, direct calculation shows that $\delta'(x) = bx$, $e^{-1} \leq b < 1$ is better than $\delta(x)$. This inadmissibility result shows that the general result of Exercise 4.17.4 does not extend to problems with K not compact, even when $k = 1$. (All estimators of the form $\delta(x) = bx$, $0 < b \leq 1$, are generalized Bayes for estimating θ. We conjecture that none of them are admissible.)]

4.17.6

Let $X \sim N(\theta, I)$. Consider the problem of estimating $\theta \in R^k$ under squared error loss. (i) Let G be a generalized prior density. Show that the generalized Bayes estimator (if it exists) can be written in the form

$$(1) \qquad \delta_G(x) = x + \frac{\nabla g^*(x)}{g^*(x)}$$

where

$$(2) \qquad g^*(x) = \int p_\theta(x)\, G(d\theta) \quad .$$

(ii) Consider the linear partial differential inequality

(3) $$\nabla \cdot (g^*(x) \nabla u(x)) \leq 0 \qquad ||x|| > 1$$

subject to the condition that u is continuous on $||x|| \geq 1$, and

(4) $$u(x) = 1 \quad \text{for} \quad ||x|| = 1, \qquad u(x) \leq 1 \quad \text{for} \quad ||x|| > 1 \ .$$

Show that if (3), (4) have a non-constant solution which also satisfies

(5) $$E_\theta(||\frac{\nabla u}{u}||^2) < \infty , \qquad \theta \in R^k \ ,$$

then δ_G is inadmissible.

[(ii) Let $\delta(x) = \delta_G(x) + \frac{\nabla u}{u}$. Use (5) and Green's theorem to justify an expression like 4.10(4) for $R(\theta, \delta_G) - R(\theta, \delta)$ but with an extra term involving a surface integral over $\{x: ||x|| = 1\}$. This extra term is non-negative because of (4), and the remainder of the expression is non-negative because of (3). (Note that Exercise 4.11.5 is a special case of the above. Brown (1971) proves that solubility of (3),(4) implies inadmissibility of δ_G (condition (5) is not required), and conversely if $\frac{\nabla g^*(x)}{g^*(x)}$ is bounded -- and somewhat more generally -- then insolubility of (3), (4) implies admissibility of δ_G. See also Srinivasan (1981).]

4.17.7 (Berger and Srinivasan (1978).)

(i) Again let $X \sim N(\theta, I)$ and consider the problem of estimating $\theta \in R^k$ under squared error loss. Suppose

(1) $$\delta(x) = x + \frac{Bx}{x'Mx} + O(\frac{1}{||x||^2})$$

for two constant k×k matrices B and M. Show that δ is inadmissible unless $B = cM$ for some $c \in R$.

[Theorem 4.17 and the representation 4.17.5(1) imply

$$\nabla(\ln g^*(x)) = \frac{Bx}{x'Mx} + O(\frac{1}{||x||^2}) \ .$$

By considering line integrals over closed paths show this is impossible unless B = cM. The calculations are easier if B and M are simultaneously diagonalized (w.l.o.g.). Then when k = 2 the only paths that need be considered are those bounding sets of the form $\{x: x_1 \geq 0, x_2 \geq 0, r \leq ||x|| \leq r + \varepsilon\}$.]

(ii) Suppose, instead, that $X \sim N(\theta, \Sigma)$ with Σ known (positive definite); and δ is given by (1). Now write a necessary condition on B and M for admissiblity of δ. Does the condition involve Σ? What if the loss function is $L(\theta, a) = (a - \theta)' D(a - \theta)$ for some (known positive definite matrix D?

4.18.1

Verify the assertion in 4.18(2). [Use Lemma 3.5 and Exercise 3.5.1(2).]

4.20.1

If $x \notin S$ as defined in 4.20(1) then $\int p_\theta(x)g(\theta)d\theta = \infty$, so that the generalized Bayes procedure for the conjugate prior does not exist at x. [See Exercise 4.19.1.]

4.20.2

Show that Karlin's condition 4.5(2) implies that $S \supset K^\circ$. (Hence, if $\nu(\partial K) = 0$ it implies that the estimator 4.20(2) = 4.5(1) is generalized Bayes.) (ii) Give an example where 4.5(2) is satisfied but 4.20(2) = 4.5(1) is not generalized Bayes.

4.20.3

Let $\{p_\theta: \theta \in \Theta\}$ be a stratum of an exponential family, as defined in Exercise 3.12.1. Suppose it is desired to estimate $\eta(\theta) = \dfrac{\xi_{(1)}(\theta)}{\xi_{(2)}(\theta)}$ under squared error loss. (Note that in the sequential setting of 3.12.2(iii) and 3.12.3, $\eta(\theta) = E_\theta(Y)$ is a very natural quantity to estimate.) State general conditions to justify the formal manipulation --

$$(1) \qquad \frac{x_{(1)}}{x_{(2)}} = \frac{\int \eta(\theta)e^{\theta\cdot x}\, d\theta_{(1)}}{\int e^{\theta\cdot x}\, d\theta_{(1)}}, \qquad x_{(1)} \in (K_{(1)})^\circ \quad \text{--}$$

which says that $\delta(x) = \dfrac{x_{(1)}}{x_{(2)}}$ is generalized Bayes on $(K_{(1)})^\circ$ relative to the prior measure $d\theta_{(1)}$ on $\Theta^* = \{(\theta_{(1)}, \theta_{(2)}(\theta_{(1)})) \in \Theta : \theta_{(1)} \in (N_{(1)})^\circ\}$.
(The conclusion is justified in the situation of 3.12.2(ii) and in that of 3.12.2(iii) if $p_\theta(N \le N_0) = 1$, and somewhat more generally.)

4.20.4

Generalize 4.20.3(1) to obtain a **representation** for certain estimators of the form $\dfrac{x_{(1)} + a}{x_{(2)} + b}$.

4.21.1

Show that 4.21(5) and 4.21(2) imply 4.21(1) and 4.21(3).
[4.21(1) is trivial from 4.21(5'). For 4.21(3) reason as in the proof of Theorem 4.19. The key fact is that, with q as defined there,

$$\int_{-B}^{q} -\xi_1(\theta_1,\theta_2,\ldots,\theta_k)g(\theta)e^{\theta\cdot x-\psi(\theta)}\, d\theta_1 = -\int_{-B}^{q} \left(x_1 g(\theta) + \frac{\partial g(\theta)}{\partial\theta_1}\right)e^{\theta\cdot x-\psi(\theta)}$$

$$+ \left. g(\theta)e^{\theta\cdot x-\psi(\theta)}\right]_{-B}^{q}, \quad \text{etc.}$$

Now integrate over $\theta_2,\ldots,\theta_k$ and let $B \to \infty$. The first part of the expression is bounded because of 4.21(1), (2), and the second part because of 4.21(5').]

4.21.2 (Converse to Theorem 4.19.)

Let G be a prior measure whose Bayes procedure for estimating $\xi(\theta)$ exists on S and satisfies $\delta(x) = \alpha x + \beta$. Suppose $S^\circ \ne \phi$. Assume further that G possesses a density g satisfying 4.21(1), (2), (3). Then G is a conjugate prior measure, and its conjugate prior density, 4.18(1), has $\alpha = 1/(\lambda + 1)$ and $\beta = \gamma/(\lambda + 1)$.) Apply 4.7(3) to the last integral of the equality

(1) $(\lambda+1)\int(\nabla\psi(\theta))g(\theta)e^{\theta\cdot x-\psi(\theta)}d\theta \;=\; \gamma\int g(\theta)e^{\theta\cdot x-\psi(\theta)}d\theta + x\int g(\theta)e^{\theta\cdot x-\psi(\theta)}d\theta$,

rearrange terms and invoke completeness to find

(2) $$\nabla g(\theta) \;=\; (\gamma - \lambda\nabla\psi(\theta))g(\theta) \qquad .]$$

(Diaconis and Ylvisaker (1979) show that this statement is true without this "further" assumption that G possess a density.)

(A question of interest is whether this unicity result extends to non-linear generalized Bayes estimators. To be more precise suppose the generalized Bayes procedures for estimating $\xi(\theta)$ under priors G and H exist and are equal everywhere on S with $S° \neq \phi$. Does this imply G = H? In the case of the normal distributions or the Poisson distribution the answer is yes. See 4.15.1 for the normal distribution and Johnstone (1982) for the Poisson distribution.)

4.24.1

Suppose $\delta(\cdot)$ is admissible for estimating ξ under squared error loss. Then $\nu\{x : \delta(x) \notin K\} = 0$.

[Define $\delta'(x)$ as the projection of $\delta(x)$ on K. If $\nu\{x : \delta(x) \neq \delta'(x)\} \neq 0$ then $R(\theta, \delta') < R(\theta, \delta)$ whenever $R(\theta, \delta) < \infty$. (If δ is admissible there must exist some θ for which $R(\theta, \delta) < \infty$.)]

4.24.2

(i) Verify that the conclusion of Theorem 4.24 remains valid when $\{p_\theta\}$ is a steep exponential family and $\Theta \subset N°$. (ii) Even more generally, it is valid for any one-parameter exponential family if

(1) $$\Theta \;\subset\; \{\theta : E_\theta(X) = \xi(\theta) \in R\}$$

and if the definition 4.24(1) is modified to

(2) $$I'_\delta = \{x : \nu(\{y: y > x,\ \delta(y) \in \xi(N^\circ)^\circ\}) > 0$$

and $$\xi(\{y: y < x,\ \delta(y) \in \xi(N^\circ)^\circ\}) > 0\} \quad .$$

(iii) Extend Theorem 4.24 to the problem of estimating $\rho(\theta)$ under squared error loss where $\rho\colon N^\circ \to R$ is a non-decreasing function. [The formulation and proof are identical to (ii), above.]

4.24.3

Let $\nu = \nu_1 + \nu_2$ where ν_1 is Lebesgue measure on $(0, 3)$ and ν_2 gives mass 1 to each of the points $x = 1,2$. Consider the estimator δ of ξ (under squared error loss) given by

(1) $$\delta(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{2} & x = 1 \\ 1\frac{1}{2} & 1 < x < 2 \\ 2\frac{1}{2} & x = 2 \\ 3 & x > 2 \end{cases} \quad .$$

(i) Show that δ has the representation 4.24(2) on $I = (1,2)$, but (ii) this representation cannot be extended to the points $x = 1,2$ even though $\delta(x) \in K^\circ$ for these points. (iii) Show that δ is a pointwise limit of a sequence of Bayes procedures. (δ is also admissible. See Exercise 7.9.1.)

4.24.4

Let X have the geometric distribution with parameter p(Ge(p)), under which

(1) $$\Pr\{X = x\} = p(1 - p)^x \qquad x=0,1,\ldots \quad .$$

(i) Show that $\delta(x) = x/2$ is an admissible estimator of $E_p(X) = (1-p)/p$ under squared error loss. [Use Karlin's Theorem 4.5. Note also that the estimators $\delta(x) = cx$ with $c > \frac{1}{2}$ fail to satisfy 4.5(2) and are not admissible.] (ii) Suppose it is known in addition that $p \le \frac{1}{2}$, so that $E_p(X) \ge 1$. Using Theorem 4.24 show that the truncated version of δ-- namely $\delta'(x) = \max(\delta(x), 1)$ --

is inadmissible. (iii) Can you find an (admissible) estimator better than δ' ??)