A UNIFIED ADMISSIBILITY PROOF

Lawrence D. Brown[1] and Jiunn Tzon Hwang[2]

Department of Mathematics
Cornell University
Ithaca, New York, U.S.A.

I. INTRODUCTION

This paper contains an admissibility proof for a variety of generalized Bayes estimators. The context is the problem of estimating the natural mean vector of an exponential family under a quadratic-form loss.

The ambition of the paper is two-fold. One is to establish the admissibility of certain natural procedures whose admissibility was previously in question, and to enable the proposal of new admissible procedures in certain situations. The second is to provide a simpler, more transparent proof even in previously established cases - a proof which also displays the similarity of all problems within the context of the theorem.

The theorem here includes Karlin's theorem on the admissibility of linear estimators in one-parameter exponential families. See Karlin [19]; also Cheng [10], Zidek [25] and Ghosh and Meeden [12].

But, the result here is not limited to one parameter exponential families; and covers a wide variety of generalized Bayes estimators, not merely linear estimators. In this same sense the

[1]Research supported in part by National Science Foundation grant MCS 7824175.

[2]Research supported in part by National Science Foundation grant MCS 8003568.

ISBN 0-12-307501-7

proof here may be thought of as a double extension of Cheng's [10] proof of Karlin's theorem.

There is another proof in the literature with considerable superficial similarity to the proof here. Zidek [25] gives a proof of admissibility in a very general one parameter context. See also Portnoy [21]. Zidek's theorem also contains Karlin's theorem. It may even be that Zidek's theorem includes much of ours in the one parameter situation, however Zidek's main regularity condition is relatively obscure in the general case. The precise connection is therefore not easy to establish.

The principal elements of our proof are Blyth's lemma, Green's Theorem (integration by parts), and the Cauchy-Schwartz inequality. Zidek's proof involves exactly these same elements, but they are applied differently and hence lead to a different result.

II. SETTING

Introduce the standard elements of an estimation problem: the sample space, $\mathcal{X}$, the parameter space, Θ, and decision space, $\mathcal{G}$, each contained in $\mathbb{R}^p$. Assume the unknown distribution on $\mathcal{X}$ is from an exponential family with densities

$$f_\theta(x) = e^{\theta\cdot x-\psi(\theta)}$$

relative to a σ-finite Borel measure, ν, on $\mathcal{X}$. Take Θ to be the natural parameter space,

$$\Theta = \{\theta: \int e^{\theta\cdot x}\,\nu(dx) < \infty\}.$$

Assume Θ is open in $\mathbb{R}^p$. Then

$$E_\theta(X) = \nabla\psi(\theta).$$

See, for example, Barndorff-Nielson [1]. Assume for now that the loss function is

(2.1) $L(\theta,a) = ||a - \nabla\psi(\theta)||^2 = ||a - E_\theta(X)||^2.$

Other quadratic type losses will be considered later. The risk of a non-randomized estimator δ: $\mathcal{X} \to \mathcal{G}$ is

$$R(\theta,\delta) = E_\theta(L(\theta,\delta(X)).$$

For convenience, introduce the notation

$$I_x h = \int h(\theta)e^{\theta\cdot x-\psi(\theta)}d\theta.$$

Let G be a non-negative measure on Θ with differentiable density g. Assume $G(K) < \infty$ for every compact $K \subset \Theta$. Suppose

(2.2) $I_x(||\nabla g||) < \infty$ for all $x \in \mathcal{X}$.

Then, *define*

$$\delta_g(x) = x + \frac{I_x(\nabla g)}{I_x g}$$

with the obvious convention that $\frac{I_x(\nabla g)}{\infty} = 0$. As motivation for this definition, note that if $\Theta = \mathbb{R}^p$ and if $I_x g < \infty$ then

(2.3) $$\delta_g(x) = x + \frac{I_x(\nabla g)}{I_x g} = \frac{I_x(g\ \nabla\psi)}{I_x g}$$

$$= \frac{\int_\Theta \nabla\psi(\theta)f_\theta(x)g(\theta)d\theta}{\int_\Theta f_\theta(x)g(\theta)d\theta}$$

by Green's theorem and (2.2). Thus in this case δ_g is the generalized Bayes estimator corresponding to g. The expression (2.3) is frequently valid also when $\Theta \neq \mathbb{R}^p$, and in these cases δ_g is again the appropriate generalized Bayes estimator. See the remark following (5.1).

III. BASIC RESULT FOR $\Theta = \mathbb{R}^p$

Impose two conditions on the generalized density, g, in addition to the mild conditions implicit in Section 2:

The *growth condition*,

$$(3.1) \qquad \int_{\mathbb{R}^p - S} \frac{g(\theta)}{||\theta||^2 \ell n^2(||\theta|| \vee 2)} d\theta < \infty$$

where $S = \{\theta: ||\theta|| \leq 1\}$ and $a \vee b = \max\{a,b\}$.

And, the *asymptotic flatness condition*,

$$(3.2) \qquad \int I_x\{g|| \frac{\nabla g}{g} - \frac{I_x \nabla g}{I_x g} ||^2\} \nu(dx) < \infty.$$

This form of the condition is not easy to verify. A more transparent but slightly less general condition is

$$(3.3) \qquad \int_{\mathbb{R}^p} \frac{||\nabla g(\theta)||^2}{g(\theta)} d\theta < \infty.$$

LEMMA 3.1. *Equation* (3.3) *implies Equation* (3.2).

Proof. $I_x(g||\frac{\nabla g}{g} - \frac{I_x \nabla g}{I_x g}||^2)$

$$= I_x(\frac{||\nabla g||^2}{g}) - \frac{||I_x \nabla g||^2}{I_x g}$$

$$\leq I_x(\frac{||\nabla g||^2}{g})$$

and

$$\int_{\mathcal{X}} I_x(\frac{||\nabla g||^2}{g}) \nu(dx) = \int \frac{||\nabla g(\theta)||^2}{g(\theta)} d\theta.$$

One final minor technical assumption is

$$(3.4) \qquad \sup\{R(\theta, \delta_g): \theta \in K\} < \infty \text{ for all compact sets } K \subset \Theta.$$

Here is the basic theorem for $\Theta = \mathbb{R}^p$ and ordinary quadratic loss (2.1).

THEOREM 3.1. *Assume* (3.1), (3.2) *and* (3.4). *Then* δ_g *is admissible.*

Before proving Theorem 3.1, we first introduce some notations. Let

$$B(g,\delta) = \int R(\theta,\delta)g(\theta)d\theta.$$

The sequence $h_n: \Theta \to [0,1]$ of absolutely continuous functions will be explicitly defined later so that, for every n, $h_n(\theta) = 1$ if $\theta \in S$ for some set S with $\int_S g(\theta)d\theta > 0$, and

$h_n(\theta) = 0$ for $||\theta|| > n$. Let $g_n(\theta) = h_n^2(\theta)g(\theta)$ and let

$$\Delta_n = B(g_n,\delta_g) - B(g_n,\delta_{g_n}).$$

In common with many admissibility proofs, the proof of Theorem 3.1 is based on Blyth's method. See Blyth [6], Stein [22] as well as page 386 of Berger's book [4]. The proof is easily supplied below.

Blyth's method: δ_g *is admissible if* $\exists\{h_n\} \ni \Delta_n \to 0$.

Proof. Suppose δ_g not admissible and let $R(\theta,\delta') \leq R(\theta,\delta_g)$ with $\delta' \neq \delta_g$(a.e.(ν)). Let $\delta'' = (\delta' + \delta_g)/2$. Then $R(\theta,\delta'') < R(\theta,\delta_g)$ $\forall\theta$ by Jensen's inequality. Then

$$\begin{aligned}\Delta_n &= \sup\{B(g_n,\delta_g) - B(g_n,\delta): \delta\}\\ &\geqq B(g_n,\delta_g) - B(g_n,\delta'')\\ &= \int (R(\theta,\delta_g) - R(\theta,\delta''))g_n(\theta)d\theta\\ &\geqq \int_S (R(\theta,\delta_g) - R(\theta,\delta''))g(\theta)d\theta.\end{aligned}$$

This implies

$$\Delta_n \not\to 0,$$

which is a contradiction.

Proof of Theorem 3.1. A familiar algebraic manipulation yields

(3.5) $$\Delta_n = \int ||\delta_g(x) - \delta_{g_n}(x)||^2 (I_x g_n)\nu(dx)$$

because

$$\Delta_n = \iint (||\delta_g(x) - \nabla\psi(\theta)||^2 - ||\delta_{g_n}(x) - \nabla\psi(\theta)||^2) f_\theta(x)\nu(dx) g_n(\theta)d\theta$$

$$= \int \{(\delta_g(x) - \delta_{g_n}(x)) \cdot (\delta_g(x) + \delta_{g_n}(x) - 2\frac{I_x(g_n\nabla\psi)}{I_x g_n})(I_x g_n)\nu(dx)$$

$$= \int ||\delta_g(x) - \delta_{g_n}(x)||^2 (I_x g_n)\nu(dx).$$

Hence, by the differentiability of g and h_n,

$$\Delta_n = \int || \frac{I_x(\nabla g)}{I_x g} - \frac{I_x(\nabla g_n)}{I_x g_n} ||^2 (I_x g_n)\nu(dx)$$

$$= \int || \frac{I_x(\nabla g)}{I_x g} - \frac{I_x(h_n^2 \nabla g)}{I_x g_n} - \frac{I_x(g\nabla h_n^2)}{I_x g_n} ||^2 (I_x g_n)\nu(dx).$$

Continuing from the above,

$$\Delta_n \leq 2\int || \frac{I_x(\nabla g)}{I_x g} - \frac{I_x(h_n^2 \nabla g)}{I_x g_n} ||^2 (I_x g_n)\nu(dx)$$

$$+ 2\int || \frac{I_x(g\nabla h_n^2)}{I_x g_n} ||^2 (I_x g_n)\nu(dx)$$

$$= 2(B_n + A_n). \quad \text{(say)}$$

Showing $A_n \to 0$:

$$A_n = 4\int \left|\left| \frac{I_x(gh_n \nabla h_n)}{I_x(gh_n^2)} \right|\right|^2 (I_x g_n)\nu(dx)$$

$$\leq 4\int I_x(g||\nabla h_n||^2)\nu(dx) \quad \text{(by Cauchy-Schwarz inequality)}$$

$$= 4\int ||\nabla h_n(\theta)||^2 g(\theta)d\theta.$$

Let

$$(3.6) \qquad h_n(\theta) = \begin{cases} 1 & ||\theta|| \leq 1 \\ 1 - \dfrac{\ell n(||\theta||)}{\ell n(n)} & 1 \leq ||\theta|| \leq n \\ 0 & ||\theta|| \geq n, \end{cases} \qquad n = 2,3,\ldots .$$

Clearly

$$(3.7) \qquad ||\nabla h_n(\theta)||^2 = \frac{1}{||\theta||^2 \ell n^2(n)} \chi_{1\leq||\theta||\leq n}(\theta)$$

$$\leq \frac{1}{||\theta||^2 \ell n^2\ (||\theta||\vee 2)} \chi_{||\theta||\geq 1}(\theta).$$

Note that $||\nabla h_n(\theta)||^2 \to 0$ for each $\theta \in \Theta$. Condition (3.1) (the growth condition) and the bound in (3.7) yield that

$$\int \sup_n ||\nabla h_n(\theta)||^2 g(\theta)d\theta < \infty.$$

Hence $A_n \to 0$ by the dominated converence theorem.

Showing $B_n \to 0$:

The integrand of B_n is

$$||I_x(g_n \frac{I_x(\nabla g)}{I_x g} - h_n^2 \nabla g)||^2 / I_x g_n$$

$$= ||I_x(g_n(\frac{I_x \nabla g}{I_x g} - \frac{\nabla g}{g}))||^2 / I_x g_n$$

$$\leq I_x(g_n||\frac{I_x(\nabla g)}{I_x g} - \frac{\nabla g}{g}||^2) \quad \text{(Cauchy-Schwarz)}$$

$$\leq I_x(g||\frac{I_x(\nabla g)}{I_x g} - \frac{\nabla g}{g}||^2) \quad \text{since} \quad g_n = h_n^2 g \leq g.$$

Note the integrand of B_n approaches zero for each x. Apply the flatness condition (3.2) with the above bound to get $B_n \to 0$ by the dominated convergence theorem.

Hence

$$\Delta_n \leq A_n + B_n \to 0.$$

So δ_g is admissible by Blyth's method.

IV. APPLICATIONS OF THEOREM 3.1

The following is an interesting general Corollary to Theorem 3.1.

COROLLARY 4.1. *If* $\Theta = \mathbb{R}^p$ *and* p = 1 *or* p = 2 *then the estimator* $\delta(x) = x$ *is admissible.*

Proof. Let g = 1. Then $\delta_g(x) = x$ since $\nabla g \equiv 0$. In this case, the regularity conditions of Theorem 3.1 are trivial to verify.

When p = 1 then this corollary is a special case of Karlin's theorem. The result for p = 2 is new, although special cases have been previously established as noted in Examples 4.1 and 4.2, below.

Example 4.1; *Normal distributions.* Suppose $X \sim N(\theta, I)$. By Corollary 4.1, *if* $p = 1$ *or* $p = 2$ *then* $\delta(x) = x$ *is admissible.* (For $p = 1$ there are proofs of this result which predate Karlin's paper. See Blyth [6], and Hodges and Lehmann [13]. For $p = 2$, the first proof of this result is in Stein [23]).

In general:

(i) If $g(\theta) \leq ||\theta||^{2-p-\varepsilon}$ for some $\varepsilon > 0$ and

$$\left|\left|\frac{\nabla g(\theta)}{g(\theta)}\right|\right| = O\left(\frac{1}{||\theta||}\right)$$

then (3.1) and (3.3) are easy to check. Hence δ_g is admissible.

(ii) If $g(\theta) \leq ||\theta||^{2-p}$ and

$$\left|\left|\frac{\nabla g(\theta)}{g(\theta)}\right|\right| = O\left(\frac{1}{||\theta||}\right) \text{ and } \left|\frac{\partial^2 g(\theta)}{\partial\theta_i \partial\theta_j}\right| = O\left(\frac{1}{||\theta||^2}\right)$$

then (3.2) can be verified with some difficulty. (Extend Lemma 3.4.1 of Brown (1971).) Condition (3.1) is still easy to check. Hence δ_g is admissible. Note that if $g(\theta) = ||\theta||^{+r}$ is smooth, as above, then $\delta_g(x) - x \sim rx/||x||^2$ as $||x|| \to \infty$. See Brown [7] and [8] or Berger and Srinivasan [5].

The generalized Bayes estimators arising out of the prior $g(\theta)$ satisfying (i) or (ii) are the commonly proposed admissible estimators. See, for example, Strawderman [24] and Berger [3]. The above results are also valid for the case $X \sim N(\theta, \Sigma)$ with Σ a known, non-singular covariance matrix. (Only the asymptotic formula for $\delta_g(x)$ need be modified.)

Example 4.2; *Poisson distributions.* Suppose X_i are independent Poisson variables with mean λ_i, $i = 1, \ldots, p$. Then the natural parameter is $\theta = (\theta_1, \ldots, \theta_p)$ with $\theta_i = \ln \lambda_i$; and $\Theta = \mathbb{R}^p$. Again by Corollary 4.1, if $p = 1$ or $p = 2$ then $\delta(x) = x$ is admissible. This result for $p = 2$ was conjectured in Brown [8] and proved in Peng [20].

In general if g satisfies (i) of Example 4.1 then δ_g is admissible. If g satisfies (ii) of Example 4.1 then δ_g is *probably* also admissible, but we have not yet checked (3.2).

If $g(\theta) = ||\theta||^{-r}$ then

$$(\delta_g(x)-x)_i = \frac{r \ln x_i}{\sum_{j=1}^{p} \ln^2(x_j+1)}$$

as $x_i \to \infty$. These admissible estimators are thus similar to the (probably inadmissible) estimators in Peng [20] and in Hwang [15] improving on $\delta(x) = x$. Problems involving the weighted quadratic loss function first used by Clevenson and Zidek [11] may be treated using Theorem 2, to follow.

V. EXTENSIONS OF THE BASIC THEOREM

The basic theorem can be extended in several significant directions with only minor modifications of the proof. This section outlines three such directions and concludes with a unified statement of the resultant extended theorem. One combined effect of these extensions is to drop the previous assumption that $\Theta = \mathbb{R}^p$.

Other choices of $\{h_n\}$: In order for Blyth's method to be valid, one needs only to choose h_n so that $h_n(\theta) \geq 1$ (or even any positive number not necessary 1, as in Corollary 5.1) for all θ in a set S with $\int_S g(\theta) > 0$ and

(5.0) $$\int h_n^2(\theta)g(\theta)d\theta < \infty.$$

This latter condition will automatically be satisfied if for every n

(5.1) Closure $\{\theta: h_n(\theta) > 0\}$ is a compact subset of Θ.

The condition (5.1) is also convenient for establishing the validity of (2.3) for the prior density $h_n^2(\theta)g(\theta)$, as exploited in the algebraic manipulations in the proof of Theorems 3.1 or 5.1.

In order to show that $A_n \to 0$ as in the proof of Theorem 3.1 it is desired to choose $\{h_n\}$ so that

$$(5.2) \qquad \int ||\nabla h_n(\theta)||^2 g(\theta)d\theta \to 0.$$

As in that proof this is usually accomplished by choosing $\{h_n\}$ so that $\nabla h_n(\theta) \to 0$ for each $\theta \in \Theta$ as $n \to \infty$ and so that

$$(5.3) \qquad \int \sup_n ||\nabla h_n(\theta)||^2 g(\theta)d\theta < \infty.$$

This flexibility to choose $\{h_n\}$ other than as in the proof of Theorem 3.1 will only very rarely be useful if $\Theta = \mathbb{R}^p$, however it is usually essential when $\Theta \neq \mathbb{R}^p$.

One common class of examples has $\Theta = \mathop{X}\limits_{i=1}^{p} (-\infty,0)$. A natural choice of $\{h_n\}$ is then

$$(5.4) \qquad h_n(\theta) = \begin{cases} 1 & \Lambda \leq 1 \\ 1 - \dfrac{\ln \Lambda}{\ln n} & 1 \leq \Lambda \leq n \\ 0 & \Lambda \geq n \qquad n = 2,3,4\ldots \end{cases}$$

with $\Lambda^2 = \Lambda^2(\theta) = \sum_{i=1}^{p} \ln^2|\theta_i|$.

The growth condition (5.3), enabling the proof that $A_n \to 0$, is then

$$(5.5) \qquad \int_{\Theta-S'} \frac{g(\theta)}{\theta_i^2 \Lambda(\theta)\ln^2(\Lambda(\theta)\vee 2)}\, d\theta < \infty$$

with $S' = \{\theta\colon \Lambda(\theta) < 1\}$.

Weighted quadratic loss functions: The methods of Section 3 can readily be extended to quadratic type losses more general than (2.1). Thus (2.1) will now be replaced by a loss of the form

$$L(\theta,a) = \sum_{i=1}^{p} v_i(\theta)(a_i - \beta_i \nabla_i \psi(\theta))^2 \tag{5.6}$$

$$= \sum_{i=1}^{p} v_i(\theta)(a_i - \beta_i E_\theta(X_i))^2$$

with $\beta_i > 0$ being fixed constants, and

$$\nabla_i \psi(\theta) = (\nabla\psi(\theta))_i = \frac{\partial}{\partial\theta_i}\psi(\theta).$$

It seems clear to us that the following considerations could be further readily extended to a loss of the form

$$L(\theta,a) = (a-B(\theta)\nabla\psi(\theta))'V(\theta)(a-B(\theta)\nabla\psi(\theta))$$

with $V(\theta)$ positive definite and $B(\theta)$ non-singular. To preserve algebraic and conceptual simplicity in the statement of Theorem 5.1, below, we have not pursued this possibility.

Other (nearly conjugate) priors: Note that the prior density $g(\theta) = c$ is a conjugate prior for the exponential family. If g satisfies the flatness condition (3.2), then the prior $g(\theta)$ can be thought of as nearly equal to this conjugate prior, since (3.2) forces g to behave asymptotically much like a constant. (More precisely, $\nabla \ln g \sim \nabla \ln c = 0$.) Other conjugate priors are of interest, as well as priors affiliated with them in the above sense. Such priors do not satisfy the flatness condition (3.2). Fortunately the theory can be easily modified to accommodate them.

Fix $\eta > -1$ and $\alpha \in \mathbb{R}^p$ and consider a prior density of the form

(5.7) $$g(\theta)e^{-\eta\psi(\theta)+\alpha\cdot\theta}$$

Let

$$I_x^* h = \int_\Theta h(\theta)e^{\theta\cdot(x+\alpha)-(\eta+1)\psi(\theta)}d\theta.$$

Replace the assumption (2.2) with

(5.8) $$I_x^*(|\nabla_i(v_i g)|) < \infty \quad \forall i,\ \forall\, x \in \mathcal{X}.$$

Let $\delta(x) = \delta_g^*(x)$ have coordinates

(5.9) $$\delta_i(x) = \frac{x_i+\alpha_i}{\eta+1} + \frac{I_x^*(\nabla_i(v_i g))}{(\eta+1)I_x^*(v_i g)}.$$

Note that under (5.8) and some further minor assumptions $\delta(x)$ is the generalized Bayes estimator corresponding to the density (5.7).

Growth condition: With the loss as in (5.6) and the prior density as in (5.7) the relevant growth condition now takes one of the following forms. If $\Theta = \mathbb{R}^p$:

(5.10) $$\sum_{i=1}^{p} \int_{\Theta-S} \frac{v_i(\theta)g(\theta)e^{-\eta\psi(\theta)+\alpha\cdot\theta}}{||\theta||^2 \ln^2(||\theta||\vee 2)}\, d\theta < \infty.$$

(Compare this to (3.1).)

If $\Theta = \mathop{X}_{i=1}^{p} (-\infty,0)$:

(5.11) $$\sum_{i=1}^{p} \int_{\Theta-S'} \frac{v_i(\theta)g(\theta)e^{-\eta\psi(\theta)+\alpha\cdot\theta}}{\theta_i^2\Lambda^2(\theta)\ln^2(\Lambda(\theta)\vee 2)}\, d\theta < \infty.$$

(Compare this to (5.5).)

More generally: there exists $\{h_n\}$ satisfying (5.1) and (5.0) such that

(5.12) $$\sum_{i=1}^{p}\int (\nabla_i h_n(\theta))^2 v_i(\theta)g(\theta)e^{-\eta\psi(\theta)+\alpha\cdot\theta}d\theta \to 0.$$

(Compare this to (5.2). Note also that there is a natural extension of (5.3).)

Asymptotic flatness condition: This condition is now

(5.13) $$\sum_{i=1}^{p}\int I_x^*\{v_i g(\frac{\nabla_i(v_i g)}{v_i g} - \frac{I_x^*(\nabla_i(v_i g))}{I_x^*(v_i g)})^2\}\nu(dx) < \infty.$$

As in Lemma 3.1 this condition is implied by

(5.14) $$\sum_{i=1}^{p}\int_{\Theta} \frac{(\nabla_i[v_i(\theta)g(\theta)])^2}{v_i(\theta)g(\theta)} e^{-\eta\psi(\theta)+\alpha\cdot\theta}d\theta < \infty.$$

Here is a formal statement of the theorem.

THEOREM 5.1. *Let the loss be given by* (5.6). *Fix* $\eta > -1$ *and* g *satisfying* (5.8). *Assume that the growth condition* ((5.10) *or* (5.11) *or* (5.12)) *is satisfied and that the asymptotic flatness condition* ((5.14) *or* (5.13)) *is satisfied. Let* δ *be defined by* (5.9) *and assume the mild boundedness condition,* (3.4), *is satisfied. Then* δ *is admissible.*

Proof. The proof follows exactly the proof of Theorem 3.1. It is only necessary to substitute I_x^* for I_x and make some consequent changes in the relevant algebraic expressions, and to use the form of $\{h_n\}$ appropriate to the assumed growth condition. (Use (3.6) when $\theta = \mathbb{R}^p$ and (5.4) when $\Theta = \times_{i=1}^{p} (-\infty,0)$.)

Karlin's theorem is a corollary of Theorem 5.1. This is given below.

COROLLARY 5.1. *Suppose* $L(\theta,a) = (a-\psi'(\theta))^2$, $p = 1$ *and* $\Theta = (a,b)$ *with* $-\infty \le a < b \le \infty$. *Suppose*

$$\int_a e^{\eta\psi(\theta)-\alpha\theta}d\theta = \infty$$

and

$$\int^{b} e^{\eta\psi(\theta)-\alpha\theta}\, d\theta = \infty.$$

Then the estimator $\delta(x) = (x+\alpha)/(\eta+1)$ *is admissible.*

Proof. Choose η, α as in this corollary, and $g \equiv 1$. Then δ_g^* is as given in the corollary. The flatness condition, (5.14), is trivially satisfied since $\nabla g \equiv 0$. The growth condition (5.12) is

$$(5.15) \qquad \int (h_n'(\theta))^2 e^{-\eta\psi(\theta)+\theta}\, d\theta \to 0.$$

This condition is satisfied by the choice

$$h_n(\theta) = \begin{cases} \int_\theta^{b_n} e^{\eta\psi(t)-\alpha t}dt \Big/ \int_{(a+b)/2}^{b_n} e^{\eta\psi(t)-\alpha t}dt & \text{if } \frac{a+b}{2} \le \theta < b_n \\ \int_{a_n}^{\theta} e^{\eta\psi(t)-\alpha t}dt \Big/ \int_{a_n}^{(a+b)/2} e^{\eta\psi(t)-\alpha t}dt & \text{if } a_n \le \theta \le \frac{a+b}{2} \\ 0 & \text{otherwise} \end{cases}$$

where $a_n \searrow a$ and $b_n \nearrow b$. (This choice of h_n minimizes the left side of (5.15) subject to the constraints $h_n(a_n) = h_n(b_n) = 0$, $h_n((a+b)/2) = 1$).

Remarks. Example 6.3 involves a minor generalization of Theorem 5.1. Probably this type of generalization would be relevant to other problems involving discrete sample spaces. (It did not, however, seem to be of use in Example 6.2.).

It is of interest to inequire whether there is a valid converse to Theorem 5.1. That is, suppose the flatness condition (5.13) is satisfied but the general growth condition (5.12) is not, and the revised growth condition, outlined in Example 6.3 is also not satisfied. Does this imply that δ_g is inadmissible? Certain heuristic considerations indicate such a converse may be valid. In particular, in the normal setting of Example 4.1, this converse is valid (with a slightly modified flatness condition which may however, be implied by (5.13). Consult Brown [7].) A special case of this conjecture would be to determine whether the converse of Karlin's theorem is valid. This converse has been verified in special cases by Joshi [18] and by Johnstone [16], but completely general results have not yet been obtained.

VI. APPLICATIONS OF THEOREM 5.1

Example 6.1; *Gamma distributions*. Let X_i be independent gamma variables, $i = 1,\ldots,p$ with scale σ_i (unknown) and expectation $k\sigma_i$ (k is a known parameter). This forms an exponential family with natural parameter θ having coordinates $\theta_i = -\sigma_i^{-1}$ and with $\Theta = \overset{p}{\underset{i=1}{X}} (-\infty,0)$ and $\psi(\theta) = \sum_{i=1}^{p} -k \ \ell n|\theta_i|$. Note that $E_\theta(X_i) = \nabla_i\psi = k/|\theta_i|$. Consider a loss function of the form

$$L(\theta,a) = \Sigma|\theta_i|^{m+2}(a_i-k/|\theta_i|)^2 = \Sigma|\theta_i|^{m}(a_i|\theta_i|-k)^2.$$

The case $m = 0$ corresponds to the standard invariant quadratic loss.

The best invariant estimator is $\delta(x) = kx/(1+k)$, corresponding to the prior $\Pi|\theta_i|^{-m-1} = (\Pi|\theta_i|^{-2-m})e^{-\eta\psi(\theta)}$ with $\eta = 1/k$. Theorem 2 yields:

This estimator is admissible if $p = 1$, *or* $p = 2$ *and* $m = 0$ ($\Rightarrow$ standard invariant loss).

In other cases Theorem 5.1 fails for good reason, since δ is inadmissible by Berger [2].

For an admissible estimator use $\eta = 1$ and

$$g(\theta) = \frac{e(\theta)}{\Pi|\theta_j|^{m+2}(\Sigma|\theta_j|^{-m})^r} \quad \text{with} \quad r = p-1.$$

and with $e(\theta) = f(\Lambda(\theta))$ where f is an asymptotically flat function satisfying a suitable order condition, below. Condition (5.8) is satisfied if $m < k$. If $f(\lambda) = O((1+\lambda)^{2-p})$ then (5.11) is satisfied since substitution and some direct bounds yield (for $p \geq 2$)

$$\sum \frac{v_i(\theta)g(\theta)e^{-\eta\psi(\theta)}}{\theta_i^2} = O\left(\frac{e(\theta)}{\prod_j |\theta_j|}\right)$$

for the given choice of $r = p-1$. If $f(\lambda) = O((1+\lambda)^{-p-\varepsilon})$, some $\varepsilon > 0$, then (5.14) is satisfied since substitution and some direct bounds yield

$$\sum (\nabla_i \ln v_i g)^2 v_i g\ e^{-\eta\psi(\theta)} = O\left(\frac{e(\theta)}{\prod |\theta_j|}\right).$$

(We expect that (5.13) will be satisfied under a somewhat less stringent growth condition on f. Whether or not it is satisfied under the more aesthetic condition $f = O((1+\lambda)^{2-p})$ we cannot foresee.)

We have found the form of the resultant generalized Bayes estimators hard to conveniently describe with precision. However, *when* $m = -1$ a very crude, heuristic approximation using (1) yields

$$\delta_i(x) = \frac{x_i}{1+\eta} + \frac{-rm\ q(x,m,k)}{(1+\eta)\Sigma(1/x_j)}$$

with $0 < \inf q < \sup q < \infty$. This compares well with Berger's [2] estimator

$$\delta_i(x) = \frac{x_i}{1+\eta} + \frac{-rm}{(1+\eta)^2\Sigma(1/x_j)}.$$

Presumably one can derive similar comparisons for other values of m. (Note also that for $m = +1$ we need $k > 1$, as does Berger.)

It is interesting to observe that only the value $r = p-1$ yields a prior satisfying the growth condition (5.11). One may reasonably presume that the other values of r lead to inadmissible estimators. This apparently corresponds to the fact that (for $m \leq 0$) Berger's estimator for his constant $c = p-1$ dominates his estimator for all other values of c. (See also Brown, L. [9], p. 10-12.)

Example 6.2; *Geometric distribution*. Let X_i be independent geometric (π_i) variables. Thus $P\{X_i = x\} = \pi_i^x(1-\pi_i)$, $x = 0,1,\ldots$ This forms an exponential family with natural parameter having coordinates $\theta_i = \ln \pi_i$ and with $\theta = \overset{p}{\underset{i=1}{X}} (-\infty,0)$ and $\psi(\theta) = -\sum_{i=1}^{p} \ln(1-e^{\theta_i})$. Note that $\nabla_i\psi(\theta) = e^{\theta_i}/(1-e^{\theta_i}) = \pi_i/(1-\pi_i)$ and $\partial^2/\partial\theta_i^2\, \psi(\theta) = \mathrm{Var}_\theta(X_i) = e^{\theta_i}/(1-e^{\theta_i})^2 = \pi_i/(1-\pi_i)^2$.

Consider the ordinary quadratic loss (2.1), and the estimator $\delta(x) = x/2$. This estimator is generalized Bayes for the conjugate prior $1\cdot e^{-\psi(\theta)}$. *When* $p = 1$ Theorem 5.1 (or Corollary 5.1) shows this estimator to be admissible. But *when* $p \geq 2$ the growth conditions (5.11) or (5.12) fails; and $\delta(x) = x/2$ is probably inadmissible.

Now, suppose

$$L(\theta,a) = \sum_{i=1}^{p} \frac{(1-\pi_i)^2}{\pi_i} \left(a_i - \frac{\pi_i}{1-\pi_i}\right)^2,$$

which is reasonable since $\mathrm{Var}\, X_i = \pi_i/(1-\pi_i)^2$. Then $\delta(x) = x/2$ is generalized Bayes for the prior $\left(\prod_{i=1}^{p} [e^{\theta_i}/(1-e^{\theta_i})^2]\right)e^{-\psi(\theta)}$.

When $p = 2$ condition (5.11) is now satisfied, so that $\delta(x) = x/2$ is admissible. The same result holds for

$$L(\theta,a) = \sum_{i=1}^{p} \frac{(1-\pi_i)^2}{\pi_i(1+\pi_i)} \cdot \left(a_i - \frac{\pi_i}{1-\pi_i}\right)^2$$

for which $\delta(x) = x/2$ is constant risk minimax.

It should be possible to use Theorem 5.1 to describe generalized priors which lead to admissible, and hopefully reasonable, estimators for the cases where $\delta(x) = x/2$ is inadmissible.

The above results generalize easily to situations where the X_i are negative binomial variables.

Example 6.3; *Poisson distributions*. Suppose X_i are independent Poisson variables with mean λ_i as in Example 4.2. Recall that $\theta_i = \ln \lambda_i$. Now suppose

$$L(\theta,a) = \sum_{i=1}^{p} \lambda_i^{-1}(a_i-\lambda_i)^2.$$

This replaces the ordinary quadratic loss discussed in Example 4.2. Clevenson and Zidek [11] first studied this problem and recommended the estimator

$$\delta_{CZ}(X) = [1 - \frac{\beta+p-1}{\beta+p-1+\Sigma X_j}]X, \ p \geq 2, \ \beta \geq 0$$

which is generalized Bayes relative to the prior (in λ)

$$(6.1) \qquad f(\lambda) = (\Sigma\, \lambda_j)^{-(p-1)} \int_0^\infty u^{\beta+p-2}(u+\Sigma\, \lambda_j)^{-\beta} e^{-u} du.$$

This estimator satisfies $R(\theta,\delta_{Cz}) < p \equiv R(\theta,X)$. In terms of θ, the prior is of course $g(\theta) = \Pi e^{\theta_i} f(e^{\theta_1},\ldots,e^{\theta_p})$. We show this estimator is admissible if $\beta \geq 0$.

The proof requires a small but significant revision of Theorem 5.1. Note that if δ dominates δ_{Cz} then

$$(6.2) \qquad \delta_i(X) = 0 \text{ whenever } X_i = 0,$$

for otherwise $\lim_{\theta\to -\infty} R(\theta,\delta) = \infty$ whereas $R(\theta,\delta_{Cz}) \leq p$. Thus one may consider only the class of procedures satisfying (6.2) in the proof of Theorem 5.1. The revised growth condition (5.10) then becomes

$$(6.3) \qquad \sum \int_{R^p} (\nabla_i h_n(\theta))^2 v_i(\theta) g(\theta) j_i(\theta) d\theta < \infty$$

where $j_i(\theta) = P_\theta(X_i > 0) = 1-e^{-e^{\theta_i}}$. Similarly the asymptotic flatness conditions (5.13) and (5.14) should be modified as (6.4) and (6.5). In (6.4), $\mathcal{X}_i$ represents the set of all x for which x_i is a positive integer and x_j's are non-negative integers, and v is the discrete measure that put mass $(\Pi\, x_i!)^{-1}$ on $(x_1,\ldots,x_p)$.

$$(6.4) \qquad \sum_{i=1}^{p} \int_{\mathcal{X}_i} I_x \{v_i g (\frac{\nabla_i v_i g}{v_i g} - \frac{I_x \nabla_i v_i g}{I_x v_i g})^2\} \, v(dx) < \infty.$$

$$(6.5) \qquad \sum_{i=1}^{p} \int_{\Theta} \frac{[\nabla_i v_i(\theta) g(\theta)]^2}{v_i(\theta) g(\theta)} j_i(\theta) d\theta < \infty.$$

Let

$$\begin{aligned} h_n(\theta) &= 1 && \text{if } 0 \le \Lambda = \textstyle\sum \lambda_i < 1 \\ &= 1 - \frac{\ell n\ \Lambda}{\ell n\ n} && 1 < \Lambda < n \\ &= 0 && \text{otherwise.} \end{aligned}$$

We first prove that such h_n satisfies condition (5.0). Note $R(\theta,\delta_{CZ})$ is bounded by p and

$$(6.6) \qquad \int h_n^2(\theta) g(\theta) d\theta \le \int_{\Lambda \le n} f(\lambda) d\lambda.$$

Let $f_0(\sum \lambda_i)$ denote the integral term in the definition, (6.1), of $f(\lambda)$. Note that $f_0(t) = O((1+t)^{-\beta})$, and in particular $f_0 \equiv$ constant for $\beta = 0$. Hence from (6.6)

$$(6.7) \qquad \int h_n^2(\theta) g(\theta) d\theta \le \int_{\Lambda \le n} \Lambda^{-p+1} O((1+\Lambda)^{-\beta}) \Lambda^{p-1} d\Lambda \, dr_1,\ldots,dr_{p-1}$$

where $\lambda_1,\ldots,\lambda_p$ is transformed to Λ, $r_1,\ldots,r_{p-1}$ by the relation

$\Lambda = \sum\lambda_i$ and $\gamma_i = \lambda_i/\Lambda$. The upper bound in (6.7) is clearly finite. To check the growth condition (6.3), we note that

$$\nabla_i h_n(\theta) = -\lambda_i/[\Lambda \ \ell n\ n] \qquad 1 \leq \Lambda \leq n$$
$$= 0 \qquad \text{otherwise.}$$

Hence the ith term in (6.3) is bounded by

$$\frac{1}{\ell n^2 n} \int_{1 < \Lambda < n} (1-e^{-\lambda_i}) f(\lambda)\lambda_i/\Lambda^2 d\lambda.$$

Omitting $(1-e^{-\lambda i})$ and using the same transformation as in deriving (6.7), the last expression is smaller than

$$\text{(6.8)} \qquad \frac{1}{\ell n^2 n} \int_1^n f_0(\Lambda)/\Lambda \, d\Lambda$$

since $f_0(\Lambda) = 0((1+\Lambda)^{-\beta})$, so (6.8) approaches zero for $\beta \geq 0$.

The expression in (6.7) therefore goes to zero and the growth condition (6.3) is satisfied.

For the asymptotic flatness condition, we consider first (6.5). Note that

$$|\nabla_i v_i g/v_i g| \leq (p+\beta-1)\lambda_i/\Lambda.$$

Hence the ith term of (6.5) is bounded by

$$\text{(6.9)} \qquad \int (p+\beta-1)^2 (1-e^{-\lambda_i}) f(\lambda)/\lambda_i d\lambda.$$

Again the transformation used in obtaining (6.7) can be applied to (6.9) which gives

$$\text{(6.10)} \qquad \int (p+\beta-1)^2 (1-e^{-\Lambda r_i}) f_0(\Lambda)/[\Lambda\ r_i] d\Lambda\ dr_1 \ldots dr_{p-1}$$

$$\leq (p+\beta-1)^2 \int_0^\infty \int_0^\Lambda (1-e^{-u} 0((\Lambda+1)^{-\beta})/[u\Lambda] du d\Lambda.$$

Note that

$$\int_0^2 \int_0^\Lambda (1-e^{-u})O((\Lambda+1)^{-\beta})/[u\Lambda]\,du\,d\Lambda$$

is bounded by $kc \int_0^2 \int_0^\Lambda \Lambda^{-1} du\, d\Lambda = 2ck$ where c denotes

$\max_{0\le u\le 2} (1-e^{-u})u^{-1} < \infty$, and $k = \max O[(\Lambda+1)^{-\beta}]$ for $0 < \Lambda < 2$.

Therefore the upper bound in (6.10) is, in turn, smaller than

$$(p+\beta-1)^2(2ck + \int_2^\infty \int_0^\Lambda O((1+\Lambda)^{-\beta})(1-e^{-u})/[u\Lambda]\,du\,d\Lambda.$$

Let Γ represent the double integral in the above expression. Again using the fact $(1-e^{-u})u^{-1} \le c$ for $0 \le u \le 1$, Γ is clearly less than

$$c \int_2^\infty \int_0^1 O((1+\Lambda)^{-\beta})\Lambda^{-1}\,du\,d\Lambda$$

$$+ \int_2^\infty \int_1^\Lambda O((1+\Lambda)^{-\beta})\Lambda^{-1}u^{-1}\,du\,d\Lambda.$$

Since $\int_2^\infty O((1+\Lambda)^{-\beta})\Lambda^{-1}d\Lambda$ is finite, it is sufficient to show

$\int_2^\infty \int_1^\Lambda O((1+\Lambda)^{-\beta})\Lambda^{-1}u^{-1}\,du\,d\Lambda < \infty$, to establish the asymptotic flatness condition. This is obvious for $\beta > 0$, since for sufficiently large Λ, $(1+\Lambda)^{-\beta} < (\ln \Lambda)^{-3}$.

To verify the asymptotic flatness condition for $\beta = 0$, we consider (6.4), which clearly equals

$$(6.11) \qquad \sum_{i=1}^{p} \int_{\mathcal{X}_i} \{I_x \frac{[\nabla_i v_i g]^2}{v_i g} - \frac{(I_x \nabla_i v_i g)^2}{I_x v_i g}\}\, \nu(dx).$$

Direct calculation gives the formula

(6.12) $$I_x \Lambda^{k_0} \prod_{i=1}^{p} \lambda_i^{k_i}$$

$$\overset{\text{defn.}}{=} \int \Lambda^{k_0} \prod_{i=1}^{p} \lambda_i^{k_i} e^{\theta x - \psi(\theta)} d\theta$$

$$= (z+k_0-1+\sum k_i)! \prod_{i=1}^{p} (k_i+x_i-1)!/[z-1+\sum k_i]$$

where $z = \sum x_i$ and k_0 and k_i's are arbitrary integers such that the factorials make sense. From (6.1), $g(\theta) = (p-2)! (\Pi\lambda_i)(\sum\lambda_i)^{-p+1}$. By using (6.12), (6.11) can be shown to equal

(6.13) $$(p-1)^2(p-2)! \sum_{i=1}^{p} \int_{\mathcal{X}_i} \frac{(z-1)!}{(z+p-1)!}[\frac{x_i+1}{z+p} - \frac{x_i}{z+p-1}](\pi x_i!)\nu(dx)$$

$$= (p-1)^2(p-2)! \sum_{i=1}^{p} \sum_{x\varepsilon\mathcal{X}_i} \frac{(z-1)!}{(z+p-1)!} [\frac{x_i+1}{z+p} - \frac{x_i}{z+p-1}].$$

By adding all the terms corresponding to x such that $\sum x_i = z$, (6.13) equals the multiplication of $(p-1)^2(p-2)!$ and

(6.14) $$\sum_{z=1}^{\infty} \frac{(z-1)!}{(z+p-1)!} [1- \frac{z}{z+p-1}] \frac{(z+p-1)!}{z!(p-1)!},$$

since there are $(z+p-1)!/[z!(p-1)!]$ many points whose coordinates are non-negative integers summing up to z. The expression in (6.14) is now

$$\sum_{z=1}^{\infty} \frac{(p-1)}{(z+p-1)z(p-1)!}$$

which is clearly finite. The asymptotic flatness condition (6.4) is now satisfied for $\beta = 0$. Theorem 5.1 implies that δ_{CZ} is admissible if $\beta \geq 0$ and $p \geq 2$. This proves the conjecture of Brown (1979). For $\beta < 0$, δ_{CZ} is inadmissible as conjectured in Brown [8] and proved in Hwang [15]. The technique developed here therefore proves the best admissibility results that one can hope for δ_{CZ}.

After this paper was presented, Iain Johnstone [17] proved the admissibility of a wide class of specific estimators for this Poisson problem by using different methods. His results include our Example 6.3 as a special case.

REFERENCES

[1] Barndorff-Neilsen, O. (1977). *Information and Exponential Families in Statistical Theory*. John Wiley, New York.

[2] Berger, J. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of Gamma scale parameters. *Ann. Statist.* *8*, 545-571.

[3] Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Ann. Statist.* *8*, 716-761.

[4] Berger, J. (1980). *Statistical Decision Theory: Foundations, Concepts, and Methods*. Springer-Verlag, New York.

[5] Berger, J. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. *Ann. Statist.* *6*, 783-801.

[6] Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* *22*, 22-42.

[7] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* *42*, 855-904.

[8] Brown, L. D. (1979). A heuristic method for determining admissibility of estimators-with applications. *Ann. Statist.* 7, 961-994.

[9] Brown, L. D. (1981). The differential inequality of a statistical estimation problem. Preprint, Cornell University, Ithaca.

[10] Cheng, P. (1964). Minimax estimates of parameters of distributions belonging to the exponential family. *Chinese Math. - Acta* 5, 277-299.

[11] Clevenson, M. and Zidek, J. (1977). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc. 70*, 698-705.

[12] Ghosh, M. and Meeden, G. (1977). Admissibility of linear estimators in the one parameter exponential family. *Ann. Statist.* 5, 772-778.

[13] Hodges, J. L., Jr., and Lehmann, E. L. (1951). Some applications of the Cramer-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Probab. 1*, Univ. of California Press, Berkeley.

[14] Hwang, J. T. (1979). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. Submitted to Ann. Statist.

[15] Hwang, J. T. (1980). Semi Tail Upper Bounds on the class of a admissible estimators in discrete exponential families with applications to Poisson and Negative binomial families. Submitted to Ann. Statist.

[16] Johnstone, Iain (1981). Lecture in Purdue 3rd Symposium.

[17] Johnstone, Iain (1981). Admissible Estimators of Poisson Means, Birth-Death Processes and Discrete Dirichlet Problems. Ph.D. thesis, Cornell University, Ithaca.

[18] Joshi, V. M. (1969). On a theorem of Karlin regarding admissible estimates for exponential populations. *Ann. Math. Statist. 40*, 216-223.

[19] Karlin, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist. 29*, 411-415.

[20] Peng, J.C.M. (1975). Simultaneous estimation of the parameters of independent Poisson distributions. Technical Report 78, Dept. Statist., Stanford Univ.

[21] Portnoy, S. (1971). Formal Bayes estimation with application to a random effects model. *Ann. Math. Statist 42*, 1379-1402.

[22] Stein, C. (1955). A necessary and sufficient condition for admissibility. *Ann. Math. Statist. 26*, 518-522.

[23] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probability* 1. University of California Press, Berkeley.

[24] Strawderman, W. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* 42, 385-388.

[25] Zidek, James V. (1970). Sufficient conditions for admissibility under squared errors loss of formal Bayes estimators. *Ann. Math. Statist.* 41, 446-456.