The Annals of Statistics
1993, Vol. 21, No. 1, 378–390

ADMISSIBILITY RESULTS IN LOSS ESTIMATION

BY CHITRA LELE

Stanford University

In this paper we consider the problem of estimating the loss of point estimators and study admissibility of such estimators of loss. We adapt and extend the "unified admissibility proof" idea of Brown and Hwang to this problem. We first present the result in the Gaussian setup. We then generalize the procedure to general exponential family distributions and apply it to the Poisson distribution. The result for the gamma distribution is also stated. The role played by the "polydisc transform" (cf. Johnstone and MacGibbon) in making explicit the relationship between the Gaussian and Poisson cases is discussed.

1. Introduction. It is often of interest to assess the discrepancy between an unknown (vector) parameter $\theta \in \mathbb{R}^p$ and its point estimator $\delta(X)$. This discrepancy (or loss) being a function of both the parameter and the data is itself unknown, but a data-dependent measure $\gamma(X)$ of this loss can serve as a measure of the performance of $\delta(X)$. In this paper we shall concentrate on developing methods for proving admissibility of loss estimators in certain situations. It would be interesting to see whether, as in the point estimation setting, certain natural estimators are admissible for some dimensions and inadmissible for others. In Lele (1990), inadmissibility of certain unbiased estimators of loss is proved by construction of dominating estimators, and is illustrated with some computational results and data examples.

To begin with, note that an unbiased estimator of risk is also unbiased for estimating loss. Now let us formulate the problem in detail. Let $X \sim p_\theta(x)$, where $p_\theta(x)$ is the density of X relative to a dominating measure $\nu(dx)$ and $\theta \in \Theta \subset \mathbb{R}^p$. Let $\delta(X) \in \mathbb{R}^p$ be an estimator of θ and let the loss incurred be $L(\delta(X), \theta)$. We now need another distance measure to study how well $\gamma(X)$ behaves as an estimator of the loss $L(\delta(X), \theta)$; we use squared error for this measure. Note that L does not have to be a squared error loss function. The risk incurred by $\gamma(X)$ is $R(\gamma, \theta) = E_\theta[\gamma(X) - L(\delta(X), \theta)]^2$. Of course, this risk depends also on the decision rule $\delta(X)$, but δ is assumed fixed and given.

This approach is both conditional and frequentist, but still all the conventional definitions and methods can be used in studying the admissibility of loss estimators. We present some here for completeness. We say that $\gamma(X)$ is inadmissible if there exists another loss estimator $\tilde{\gamma}(X)$, such that $R(\tilde{\gamma}, \theta) \le R(\gamma, \theta)$ for all θ with strict inequality for some θ. The Bayes rule γ_G, corresponding to the prior $G(d\theta)$, is the minimizer of

$$r(\gamma, G) = \int R(\gamma, \theta) G(d\theta)$$

Received September 1990; revised October 1991.
AMS 1991 *subject classification*. Primary 62C15.
Key words and phrases. Admissibility, loss estimator, Blyth's method, exponential families, polydisc transform.

and

$$\min_{\gamma} r(\gamma, G) = r(\gamma_G, G) = r(G).$$

If $r(G) < \infty$, then the Bayes rule γ_G also minimizes the posterior expected distance

$$E_G\Big[\big(L(\delta(X), \theta) - \gamma\big)^2 | X = x\Big]$$

and then, $\gamma_G = E_G[L(\delta(X), \theta)|X]$. If $r(G) = \infty$, this γ_G is called a formal or a generalized Bayes rule. Note that $\delta(X)$ can be arbitrary.

If $g(\theta)$ is the density function corresponding to $G(d\theta)$, then we shall often use the notation $r(\gamma, g)$ in place of $r(\gamma, G)$.

This approach is similar to that of Johnstone (1988), who considered (in)admissibility of unbiased loss estimators in the Gaussian setting for the mle and the James–Stein point estimator.

The problem of estimated confidence has been studied rather extensively. There has also been recent interest in the area of loss estimation. Rukhin (1988) considered loss functions for the simultaneous estimation of θ and $L(\delta(X), \theta)$ and proved some admissibility results. Berger and Lu (1989) considered the problem of constructing improvements over $\gamma(X) = p$ (distinct from the unbiased estimator of loss) for estimating the loss associated with the posterior point estimator for the Strawderman prior and also for the James–Stein positive part estimator. More recently, Brown and Hwang (1989) have considered a variant of the present problem; they consider the unbiased estimator of the coverage function in the Gaussian context and prove its admissibility for $p \leq 4$, a result similar to that of Johnstone (1988). We also refer to a related result in DasGupta (1989) later on.

2. Unified admissibility proof. Our goal is to develop a single methodology for proving admissibility of a variety of Bayes loss estimators in a variety of distributions. We shall adapt the methods used by Brown and Hwang (1982) in the point estimation situation towards this end. We shall present a detailed application of the methodology in the Gaussian case and will state the results in the Poisson and the gamma distributions which can be derived from a generalization of the technique.

Johnstone [(1988), page 364] gives a version of Blyth's lemma that is useful in the loss estimation context. Let us recall that the difference in integrated risks, occurring in the statement of Blyth's lemma, can be written as

$$\begin{aligned} & r(\gamma, G_n) - r(G_n) \\ (2.1) \qquad & = \iint \Big[\big(\gamma(x) - L\big)^2 - \big(\gamma_{G_n}(x) - L\big)^2\Big] p_\theta(x) G_n(d\theta) \nu(dx) \\ & = \int \big[\gamma(x) - \gamma_{G_n}(x)\big]^2 m_n(x)\, d\nu(x). \end{aligned}$$

[Here, $L(\delta(x), \theta)$ is abbreviated by L and $m_n(x)$ is the marginal distribution of X.]

Consider an exponential family with density

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

relative to a σ-finite Borel measure ν on $\mathscr{X} \subset \mathbb{R}^p$, and $\Theta \subset \mathbb{R}^p$ is the natural parameter space,

$$\Theta = \left\{\theta \colon \int e^{\theta \cdot x} \nu(dx) < \infty\right\}.$$

Assume Θ is open in $\mathbb{R}^p$. The mean can be expressed as $E_\theta(X) = \nabla\psi(\theta)$.

First, using the exponential identities (derived via integration by parts) of the following lemma, express the difference between posterior loss estimators corresponding to different priors in terms of the derivatives of the priors. The lemma follows from Hudson (1978).

We shall let $\Theta = \mathbb{R}^p$, unless stated otherwise.

LEMMA 2.1. *The following relation holds true for a continuously differentiable function* $g\colon \mathbb{R}^p \to \mathbb{R}$, *for which all integrals exist*:

$$\begin{aligned} &\int \|X - \nabla\psi(\theta)\|^2 g(\theta) f_\theta(X)\, d\theta \\ &\quad = \int [\Delta g(\theta) + g(\theta)\Delta\psi(\theta)] f_\theta(X)\, d\theta. \end{aligned} \tag{2.2}$$

In particular, if $g(\theta) \equiv 1$, (2.2) simplifies to

$$\int \|X - \nabla\psi(\theta)\|^2 f_\theta(X)\, d\theta = \int \Delta\psi(\theta) f_\theta(X)\, d\theta.$$

For $L(\theta, X) = \|X - \nabla\psi(\theta)\|^2$, the posterior loss estimator corresponding to the prior g is therefore

$$\begin{aligned} \gamma_g(X) &= E_g(L(\theta, X)|X) \\ &= \frac{\int (\Delta g(\theta))(g(\theta))^{-1} f_\theta(X) g(\theta)\, d\theta}{\int f_\theta(X) g(\theta)\, d\theta} + \frac{\int \Delta\psi(\theta) f_\theta(X) g(\theta)\, d\theta}{\int f_\theta(X) g(\theta)\, d\theta}. \end{aligned}$$

Using (2.1),

$$\Delta_n = r(\gamma_g, g_n) - r(g_n) = \int (\gamma_g - \gamma_{g_n})^2 (I_x g_n) \nu(dx),$$

where $I_x h = \int h(\theta) f_\theta(x)\, d\theta$. Note that x will be used instead of X when it is a variable of integration.

To prove admissibility using Blyth's lemma, we need to prove that $\Delta_n \to 0$ as $n \to \infty$, for a sequence of priors $g_n \to g$.

Let us now consider the generalized Bayes point estimator [cf. Stein (1981) and Brown and Hwang (1982)]

$$\delta_g(X) = X + \frac{I_X \nabla g}{I_X g}.$$

Defining γ_{g_n} as

$$\gamma_{g_n} = E_{g_n}\left(\left\| X + \frac{I_X \nabla g_n}{I_X g_n} - \nabla\psi(\theta)\right\|^2 \middle| X\right),$$

the difference between γ_{g_n} and γ_g is

$$\begin{aligned}\gamma_{g_n} - \gamma_g = &\left[\frac{I_X \Delta g_n}{I_X g_n} + \frac{I_X(g_n \Delta\psi)}{I_X g_n} - \left\|\frac{I_X \nabla g_n}{I_X g_n}\right\|^2\right] \\ &- \left[\frac{I_X \Delta g}{I_X g} + \frac{I_X(g\Delta\psi)}{I_X g} - \left\|\frac{I_X \nabla g}{I_X g}\right\|^2\right].\end{aligned} \tag{2.3}$$

Now concentrate on the simplest case of the Gaussian distribution.

2.1. *Normal distribution.* In this case $\Delta\psi$ = constant, and the difference in (2.3) depends only on terms involving derivatives of the priors. So,

$$\Delta_n = \int \left(\frac{I_x \Delta g_n}{I_x g_n} - \frac{I_x \Delta g}{I_x g} - \left\|\frac{I_x \nabla g_n}{I_x g_n}\right\|^2 + \left\|\frac{I_x \nabla g}{I_x g}\right\|^2\right)^2 (I_x g_n)\nu(dx). \tag{2.4}$$

We assume that conditions of Lemma (2.1) also hold for g_n. The integrand in (2.4) converges to zero as n tends to infinity, and so we seek sufficient conditions for application of the dominated convergence theorem to conclude that the integral vanishes as $n \to \infty$.

We shall first state the theorem and its two corollaries and then present the proof.

THEOREM 2.1. *In case of the Gaussian distribution,* $\gamma_g = E_g(L|X)$ *is admissible as an estimator of the loss*

$$L = \left\| X + \frac{I_x \nabla g}{I_x g} - \theta\right\|^2$$

if the prior g satisfies the following two conditions:

$$\int \frac{(\Delta g)^2}{g}\, d\theta + \int \frac{\|\nabla g\|^4}{g^3}\, d\theta < \infty, \tag{2.5}$$

$$\int_{R^p - S} \frac{g(\theta)}{\|\theta\|^4 \ln^2(\|\theta\| \vee 2)}\, d\theta < \infty, \tag{2.6}$$

where $S = \{\theta\colon \|\theta\| \le 1\}$.

We now apply this theorem to two different priors $g(\theta)$ and derive admissibility results for the corresponding loss estimators. This will also illustrate that the conditions of the theorem are often easy to verify.

(i) With $g(\theta) \equiv 1$, $\delta_g(X) = X$ and for $L = \|X - \theta\|^2$, $\gamma_g(X) = p = \gamma_{\text{unb}}(X)$. The condition (2.5) holds trivially for this g, and condition (2.6) holds only for $p \leq 4$. Hence we have Johnstone's (1988) result:

COROLLARY 2.1. *$\gamma_{unb}(X) \equiv p$ is admissible as an estimator of $\|X - \theta\|^2$ for $p \leq 4$.*

Johnstone (1988) has also constructed an improved loss estimator for $p > 4$.

(ii) Next consider the Strawderman prior and the point and loss estimators that correspond to it. We examine admissibility of the posterior estimator of the loss associated with the generalized Bayes point estimator. A simplified version of the Strawderman prior [i.e., a special case of the general prior stated in, e.g., Berger and Lu (1989)] is

$$g_m(\theta) = \int_0^1 \{\det[D(\lambda)]\}^{1/2} \exp\left\{-\frac{\theta^T D(\lambda)\theta}{2}\right\} \lambda^{m-1-p/2}\, d\lambda,$$

where $D(\lambda) = \lambda(1-\lambda)^{-1} I$ and I is the identity matrix. The point estimator δ_m which is Bayes for this prior, is admissible for $m \geq (p-2)/2$ [cf. Berger (1985)]; in particular, when $m = (p-2)/2$ for $p \geq 3$.

Consider the simplest case with $m = (p-2)/2$. Then, letting $v = \|X\|^2$, the point estimator is

$$\delta(X) = X - \frac{r_{(p-2)/2}(v)}{v} X$$

and the prior with respect to which this δ is Bayes is

$$\begin{aligned} g(\theta) &= \int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left\{-\frac{\lambda}{1-\lambda}\frac{\theta^t\theta}{2}\right\} \frac{1}{\lambda^2}\, d\lambda \\ &= \Gamma\left(\frac{p}{2} - 1\right)\left(\frac{\theta^t\theta}{2}\right)^{-(p/2)+1} \end{aligned} \tag{2.7}$$

Define $u(v)$ so that $v \cdot u(v) = r_{(p-2)/2}(v)$. Then for estimating $L = \|\delta(X) - \theta\|^2$, the posterior loss estimator is $\gamma_{\text{post}} = 2 + vu(v) - vu^2(v)$. Both the conditions can be verified to hold for this prior and we have another corollary of the theorem:

COROLLARY 2.2. *γ_{post} is admissible as an estimator of L for all $p \geq 3$.*

The result also holds true for a general $m = (p-2)/2 + \beta$, $\beta > 0$.

REMARK. Note that the prior yielding an admissible point estimator gives an admissible loss estimator as well, that is, posterior mean and posterior expected loss are both admissible.

Let us now prove the theorem.

PROOF OF THEOREM 2.1. As indicated before, the task is of bounding the integral in (2.4). To do this, we initially bound the integral by four different terms, and then find conditions on the prior in order for these four terms to be bounded.

Consider the sequence of priors $\{g_n(\theta)\}$ with $g_n(\theta) = \xi_n(\theta)g(\theta)$ and $\xi_n(\theta) \le 1$ for all n. Substitute $\xi_n g$ for g_n and separate out the terms involving the derivatives of g alone using the inequality

$$(a_1 + \cdots + a_q)^2 \le q(a_1^2 + \cdots + a_q^2) \tag{2.8}$$

for $q = 2$ to give

$$\begin{aligned}\Delta_n &\le 2\int\left(\frac{I_x(g\Delta\xi_n) + 2I_x(\nabla\xi_n \cdot \nabla g)}{I_x g_n} - \left\|\frac{I_x\nabla g_n}{I_x g_n}\right\|^2\right)^2 (I_x g_n)\nu(dx)\\ &\quad + 2\int\left(\frac{I_x(\xi_n\Delta g)}{I_x g_n} - \frac{I_x\Delta g}{I_x g} + \left\|\frac{I_x\nabla g}{I_x g}\right\|^2\right)^2 (I_x g_n)\nu(dx)\\ &= 2A_n + 2B_n.\end{aligned} \tag{2.9}$$

Let us first deal with B_n. Using the Cauchy–Schwarz inequality and the fact that $g_n \le g$,

$$B_n \le \int I_x\left[g\left(\frac{\Delta g}{g} - \frac{I_x\Delta g}{I_x g} + \left\|\frac{I_x\nabla g}{I_x g}\right\|^2\right)^2\right]\nu(dx). \tag{2.10}$$

Expanding the square and simplifying the integrand in (2.10),

$$I_x\left[g\left(\frac{\Delta g}{g} - \frac{I_x\Delta g}{I_x g} + \left\|\frac{I_x\nabla g}{I_x g}\right\|^2\right)^2\right] \le I_x\frac{(\Delta g)^2}{g} + \frac{\|I_x\nabla g\|^4}{(I_x g)^3}.$$

Now, application of Hölder's inequality implies that B_n is bounded by

$$\int\frac{(\Delta g)^2}{g}\,d\theta + \int\frac{\|\nabla g\|^4}{g^3}\,d\theta.$$

Let us now consider A_n. Use of (2.8) with $q = 3$, along with repeated application of the Cauchy–Schwarz inequality, allows bounding A_n by three integrals as follows:

$$\begin{aligned}A_n &\le 3\int\frac{(\Delta\xi_n(\theta))^2}{\xi_n(\theta)}g(\theta)\,d\theta + 12\int\frac{\|\nabla\xi_n(\theta)\|^2}{\xi_n(\theta)}\frac{\|\nabla g(\theta)\|^2}{g(\theta)}\,d\theta\\ &\quad + 3\int\frac{\|\nabla(g(\theta)\xi_n(\theta))\|^4}{(g(\theta)\xi_n(\theta))^3}\,d\theta\\ &= 3C_n + 12D_n + 3E_n.\end{aligned} \tag{2.11}$$

To choose an appropriate sequence $\{\xi_n\}$, first define

$$\eta_n(\theta) = \begin{cases} 1, & \|\theta\| \le 1, \\ 1 - \dfrac{\ln\|\theta\|}{\ln n}, & 1 \le \|\theta\| \le n, \\ 0, & \|\theta\| \ge n, \end{cases} \tag{2.12}$$

for $n = 2, 3, \dots$.

We encounter the same difficulty as in Johnstone (1988) in using $\xi_n = \eta_n^2$. Hence, define

$$\xi(\eta) = \begin{cases} 2^{k-1}\eta^k, & 0 \le \eta \le 1/2, \\ 1 - 2^{k-1}(1-\eta)^k, & 1/2 \le \eta \le 1, \end{cases} \tag{2.13}$$

for some integer $k > 1$ and let $\xi_n(\theta) = \xi(\eta_n(\theta))$.

Compute the derivatives of ξ_n and substitute them in the integrands of C_n, D_n and E_n. It can then be seen that, provided $k \ge 4$, the growth condition (2.6) implies finiteness of C_n; the same condition is also sufficient for E_n to be finite and (2.6), in conjunction with the condition $\int(\|\nabla g\|^4/g^3)\,d\theta < \infty$ included in bounding B_n, also ensures the boundedness of D_n.

Thus the condition (2.5) takes care of finiteness of B_n, and together with condition (2.6), it also implies boundedness of C_n, D_n, E_n and hence of A_n. See Lele (1990) for further details of this proof. □

2.2. *Extension to other exponential family distributions.* We now return to the general exponential family setting, but restrict our attention to information-normalized loss functions, that is, weighted loss functions

$$L(\delta, \theta) = \sum w_i(\theta)(\delta_i - \nabla_i\psi(\theta))^2$$

with weights

$$w_i(\theta) \propto \left(\nabla_i^2\psi(\theta)\right)^{-1} = \left(\frac{\partial^2}{\partial\theta_i^2}\psi(\theta)\right)^{-1}.$$

Going over the procedure of bounding the difference in integrated risks appearing in Blyth's condition as in the Gaussian case, and choosing an appropriate sequence of priors $\{\xi_n\}$, we can write down sufficient conditions on g that ensure admissibility of $\gamma_g = E_g[L(\delta, \theta)|X]$.

In particular, let us see what the conditions are in the Poisson distribution case. Consider $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, \dots, p$, $X = (X_1, \dots, X_p)^T$, $\lambda = (\lambda_1, \dots, \lambda_p)^T$. Here the results can be conveniently stated in terms of λ, so we shall not refer to the canonical parameter θ. The information-normalized loss is $L_{-1}(\delta, \lambda) = \sum\lambda_i^{-1}(\delta_i(X) - \lambda_i)^2$. Let $\delta_\pi(X)$ denote the Bayes estimator corresponding to the prior $\pi(\lambda)$ and consider the sequence of priors $\{\pi_n\}$ converging to π. The sequence is of the form $\pi_n = \xi_n \pi$, wherein

$$\xi_n(\lambda) = \begin{cases} 2^{k-1}\eta_n^k, & 0 \le \eta_n \le 1/2, \\ 1 - 2^{k-1}(1-\eta_n)^k, & 1/2 \le \eta_n \le 1, \end{cases}$$

and η_n is defined as

$$\eta_n(\lambda) = \begin{cases} 1, & \Lambda \le 1, \\ 1 - \dfrac{\ln \Lambda}{\ln n}, & 1 \le \Lambda \le n, \\ 0, & \Lambda \ge n, \end{cases}$$

for $n = 2, 3, \dots,$ and $\Lambda = \Sigma_1^p \lambda_i$.

After performing the necessary calculations [cf. Lele (1990)] we get the following theorem:

THEOREM 2.2. *In the context of the Poisson distribution, $\gamma_\pi = E_\pi[L_{-1}|X]$ is admissible as an estimator of $L_{-1} = \Sigma\lambda_i^{-1}(\delta_{\pi,i}(X) - \lambda_i)^2$ if π satisfies the two conditions*

$$\int_{R_+^p} \frac{\left(\lambda \cdot \nabla^2\pi + \nabla \cdot \pi\right)^2}{\pi} d\lambda + \int_{R_+^p} \frac{\left(\lambda \cdot (\nabla\pi)^2\right)^2}{\pi^3} d\lambda < \infty \tag{2.14}$$

and

$$\int_{R_+^p - S} \frac{\pi(\lambda)}{\Lambda^2 \ln^2(\Lambda \vee 2)} d\lambda < \infty, \tag{2.15}$$

where $S = \{\lambda\colon \Lambda \le 1\}$ and

$$\nabla^2\pi = \left(\frac{\partial^2}{\partial\lambda_i^2}\pi\right)_{p\times 1}, \qquad \nabla \cdot \pi = \sum_i \frac{\partial}{\partial\lambda_i}\pi(\lambda), \quad (\nabla\pi)^2 = \left(\left(\frac{\partial}{\partial\lambda_i}\pi(\lambda)\right)^2\right)_{p\times 1}$$

Once again, we apply this theorem to two different point estimators to yield two corollaries.

(i) For $\pi(\lambda) \equiv 1$, $\delta(X) = X$ and $E[\Sigma\lambda_i^{-1}(X_i - \lambda_i)^2|X] = p$. The integrands in condition (2.14) are identically equal to zero and condition (2.15) holds only for $p \le 2$. Hence we have the result:

COROLLARY 2.3. *$\gamma_{unb}(X) \equiv p$ is admissible as an estimator of $\Sigma\lambda_i^{-1}(X_i - \lambda_i)^2$ for $p \le 2$.*

An improvement for $p > 2$ is constructed in Lele (1990).

(ii) Now let us consider the Clevenson–Zidek point estimator, estimating λ under L_{-1} loss. The estimator is

$$\delta_{\mathrm{CZ}} = \delta_{\mathrm{CZ}}(X) = \left(1 - \frac{p - 1 + \beta}{\sum X_i + p - 1 + \beta}\right) X.$$

This estimator is proved to be admissible for $p > 1$ and $\beta \geq 0$ [cf. Clevenson and Zidek (1975)]. Consider the simplest case of $\beta = 0$. Then the prior corresponding to which the Clevenson–Zidek estimator is Bayes, is

$$\pi_{CZ}(\lambda) = \frac{\Gamma(p-1)}{\left(\sum \lambda_i\right)^{p-1}}. \tag{2.16}$$

Both the conditions can easily be seen to hold for this prior.

COROLLARY 2.4. *The posterior loss estimator*

$$p - \frac{(p-1)^2}{\sum X_i + p - 1}$$

of the loss $\Sigma \lambda_i^{-1}(\delta_{CZ,i} - \lambda_i)^2$ *is admissible for all* $p > 1$.

(Note that the C–Z point estimator reduces to the mle when $p = 1$.)

REMARK. As in the Gaussian case, here too, the prior that produces an admissible point estimator also yields an admissible loss estimator.

There also exists an extension of the admissibility theorem for the situation where $\Theta = \mathbb{R}^p_-$. Once again we restrict to information-normalized loss functions, and the sequence $\{\xi_n\}$ is the same as defined above in the Poisson case, but in the definition of η_n, Λ is defined by the relation $\Lambda^2 = \Sigma \ln^2|\theta_i|$.

The following theorem can be proved.

THEOREM 2.3. *For an exponential family distribution with* $\Theta + \mathbb{R}^p_-$, *the posterior estimator with respect to the prior* g, *of the loss* $L[\delta(X), \theta]$ *is admissible if* g *satisfies the two conditions*

$$\int_{R^p_-} \frac{\left[\sum (w_i g)''_i\right]^2}{g}\, d\theta + \int_{R^p_-} \frac{1}{g^3}\left(\sum \frac{\left[(w_i g)'_i\right]^2}{w_i}\right)^2 d\theta < \infty \tag{2.17}$$

and

$$\int_{R^p_- - S} \sum \frac{w_i^2}{\theta_i^4} \frac{g(\theta)}{\Lambda^2 \ln^2(\Lambda \vee 2)}\, d\theta < \infty, \tag{2.18}$$

where $S = \{\theta\colon \Lambda \leq 1\}$ *and*

$$(w_i g)'_i = \frac{\partial}{\partial \theta_i}(w_i g), \qquad (w_i g)''_i = \frac{\partial^2}{\partial \theta_i^2}(w_i g).$$

Let us first note that the above theorem can be further extended to include other conjugate priors and priors associated with conjugate priors. Consider prior densities of the form $g^*(\theta) = g(\theta)\exp\{\alpha \cdot \theta - \eta\psi(\theta)\}$, for $\alpha \in \mathbb{R}^p$, $\eta > -1$

and consider the corresponding generalized Bayes point and loss estimators. Conditions on $g(\theta)$ which are sufficient for the admissibility of the posterior Bayes loss estimator corresponding to the prior $g^*(\theta)$ are minor modifications of conditions (2.17) and (2.18): The integrands in (2.17) and (2.18) get multiplied by the factor $\exp\{\alpha \cdot \theta - \eta\psi(\theta)\}$.

Using this modification, we can establish a result for the Gamma distribution with the scale parameter unknown. Consider the following setup:

$$X_i \sim_{\text{ind}} \frac{\exp\{-X_i/\alpha_i\} X_i^{(\beta-1)}}{\alpha_i^{\beta}\Gamma(\beta)},$$

where $\beta > 1$ is known and $\alpha_i > 0$, $X_i > 0$, $X = (X_1, \dots, X_p)^T$ and $\alpha = (\alpha_1, \dots, \alpha_p)^T$. Here, $\theta_i = \alpha_i^{-1}$ and the information-normalized loss function is

$$L(\delta, \theta) = \sum \theta_i^2 \left(\delta_i - \frac{\beta}{|\theta_i|} \right)^2.$$

The best invariant point estimator $\delta(X) = \beta X/(\beta + 1)$ is generalized Bayes for the prior $g^*(\theta) \propto \prod_{i=1}^{p} |\theta_i|^{-1}$. The following corollary is now easily verifiable.

COROLLARY 2.5. *In the above gamma setting, the posterior loss estimator* $p\beta^2/(\beta + 1)$ *is admissible for* $p \leq 2$.

DasGupta (1989) considers the squared error loss function for the gamma distribution and has an inadmissibility result for $p \geq 3$ for the best unbiased loss estimator.

It is also possible to derive a result for a loss function which is not information-normalized. The admissibility theorem for the Poisson distribution with squared error loss is stated below, along with a corollary. Refer to Lele (1990) for the proof which involves some manipulation in order to transform the problem to a familiar form, after which we can resort to the methodology of Theorem 2.1.

We consider the same Poisson setting as before, but look at the loss

$$L(\delta(X), \lambda) = \sum_i (\delta_i(X) - \lambda_i)^2.$$

The maximum likelihood point estimator $\delta(X) = X$ is formal Bayes with respect to the prior $\pi(\lambda) \propto (\prod \lambda_j)^{-1}$ and the corresponding posterior estimator of L is $\gamma_{\text{post}}(X) = \sum X_i$. On transformation to the canonical parameter θ such that $\theta_i = \ln(\lambda_i)$, $i = 1, \dots, p$, and defining $\varphi(\theta) = \pi(\lambda)$ we have the theorem stated below.

THEOREM 2.4. *Let $\varphi(\theta)$ be a prior satisfying the regularity conditions of Hudson* (1978). *Then the posterior expected loss $\gamma_\varphi(X)$ is admissible for* *estimating the squared error loss L if $\varphi(\theta)$ satisfies the following conditions*:

$$\int_{R^p} \frac{(\Delta\varphi(\theta))^2}{\varphi(\theta)}\,d\theta + \int \frac{\|\nabla\varphi(\theta)\|^4}{\varphi^3(\theta)}\,d\theta < \infty, \tag{2.19}$$

$$\int_{R^p} \frac{(\nabla\cdot\varphi(\theta))^2}{\varphi(\theta)}\,d\theta < \infty, \tag{2.20}$$

$$\int_{R^p - S} \frac{\varphi(\theta)}{\|\theta\|^2 \ln^2(\|\theta\| \vee 2)}\,d\theta < \infty. \tag{2.21}$$

Here $S = \{\theta\colon \|\theta\| \le 1\}$.

The unbiased loss estimator ΣX_i corresponds to the prior $\varphi(\theta) \equiv 1$. So, the first two conditions above hold trivially, and the last condition holds only for $p \le 2$. Thus follows the result:

COROLLARY 2.6. *In the p-dimensional Poisson situation*, ΣX_i *is admissible as an estimator of* $L = \Sigma(X_i - \lambda_i)^2$ *when* $p \le 2$.

The problem of finding improvements for $p > 2$ is still unresolved.

3. Role of the polydisc transform. We will now employ the polydisc transform to explore the relation between the Gaussian and the Poisson cases in the context of admissibility. The transform is described in detail, along with its properties, in Johnstone and MacGibbon (1992). The definition and some fundamental aspects are reproduced here. Define a many-to-one mapping $\tau\colon \mathbb{R}^{2p} \to \mathbb{R}^p_+$ as

$$\tau\colon (\omega_1, \dots, \omega_{2p-1}, \omega_{2p}) \to \left(\omega_1^2 + \omega_2^2, \dots, \omega_{2p-1}^2 + \omega_{2p}^2\right). \tag{3.1}$$

The set $\Omega = \tau^{-1}(T)$ is called the polydisc transform of T. A function $v(\tau)$ defined on T induces a function $u(\omega) = v(\tau(\omega))$ on Ω. We apply the transform to the parameter space, and in our context, τ will be replaced by λ. Define $\beta(\omega) = \pi(\lambda(\omega))$. The following identities can be verified to hold:

$$\|\nabla\beta(\omega)\|^2 = 4\sum \lambda_i(\pi_i'(\lambda))^2 = 4\lambda\cdot(\nabla\pi)^2, \tag{3.2}$$

$$\Delta\beta(\omega) = 4\left[\Delta\cdot\pi(\lambda) + \lambda\cdot\nabla^2\pi\right], \tag{3.3}$$

where $(\nabla\pi)^2 = (((\partial/\partial\lambda_i)\pi(\lambda))^2)_{p\times 1}$ and $\nabla^2\pi(\lambda) = ((\partial^2/\partial\lambda_i^2)\pi(\lambda))_{p\times 1}$.

We can now see, in the context of loss estimation, how a Poisson problem in p dimensions can be transformed to a corresponding Gaussian problem in $2p$ dimensions with this transformation. The transformation is applied in the point estimation context in Lele (1990).

Consider the conditions for admissibility of the posterior loss estimator in the Gaussian setting (as derived in Section 2.1) and the corresponding ones for the Poisson with L_{-1} loss (as in Section 2.2). So we have to compare conditions (2.5) and (2.6) with conditions (2.14) and (2.15). It is not hard to verify [see Lele (1990)] that after application of the polydisc transformation, the left-hand side expressions of (2.14) and (2.15) transform to the $2p$ dimensional versions of the left-hand side expressions of (2.5) and (2.6) respectively.

Moreover, the Clevenson–Zidek prior (2.16) is equivalent to the Strawderman prior (2.7) after the transformation, and the same holds true even if we do not restrict to the special values of the parameters m and β in the priors.

Brown (1979) first suggested that admissibility considerations for point estimators in the Poisson distribution in p dimensions correspond to those in the Gaussian distribution in $2p$ dimensions. The polydisc transform provides some insight into how this phenomenon occurs, and we find that the equivalence also carries over to the loss estimation setting.

4. Concluding remarks. A necessary condition for admissibility of loss estimators, namely, "if a loss estimator is admissible, then it is generalized Bayes" can be proved to hold in the case of exponential families [cf. Lele (1990)]. The point estimation analog of this result is extremely familiar [cf. Brown (1986)]. The proof in the point estimation framework extends naturally to yield this result for loss estimators.

Acknowledgments. I profoundly and sincerely thank Professor Iain Johnstone for his guidance and support throughout this research. I would also like to express my gratitude to the Associate Editor and the referees who made constructive suggestions regarding presentation of this paper.

REFERENCES

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

BERGER, J. O. and LU, K. L. (1989). Estimation of normal means: Frequentist estimation of loss. *Ann. Statist.* **17** 890–906.

BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960–994.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Decision Theory*. IMS, Hayward, CA.

BROWN, L. D. and HWANG, J. T. (1982). A unified admissibility proof. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. Berger, eds.) 205–230. Springer, New York.

BROWN, L. D. and HWANG, J. T. (1989). Admissibility of confidence estimators. Unpublished manuscript.

CLEVENSON, M. and ZIDEK, J. (1975). Simultaneous estimation of the mean of independent Poisson laws. *J. Amer. Statist. Assoc.* **70** 698–705.

DASGUPTA, A. (1989). Inadmissibility result for squared error loss: With applications. *Ann. Statist.* **17** 1360–1374.

HUDSON, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

JOHNSTONE, I. M. (1988). On inadmissibility of unbiased estimates of loss. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) 361–379. Springer, New York.

JOHNSTONE, I. M. and MACGIBBON, K. B. (1992). Minimax estimation of a constrained Poisson vector. *Ann. Statist.* **20** 807–831.

LELE, C. (1990). Admissibility results in loss estimation. Technical Report 352, Dept. Statistics, Stanford Univ.

RUKHIN, A. L. (1988). Estimated loss and admissible loss estimators. In *Statistical Decision Theory and Related Topics IV* (S. S. Gupta and J. O. Berger, eds.) 409–418. Springer, New York.

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

DIVISION OF EPIDEMIOLOGY
HRP MODULAR 2
STANFORD UNIVERSITY SCHOOL
OF MEDICINE
STANFORD, CALIFORNIA 94305