ALL ESTIMATES WITH A GIVEN RISK, RICCATI DIFFERENTIAL EQUATIONS, AND A NEW PROOF OF A THEOREM OF BROWN

by

Anirban DasGupta*   and   William E. Strawderman*
Purdue University   Rutgers University

Technical Report #95-22

Department of Statistics
Purdue University

May 1995
Revised July 1996
Revised October 1996

ALL ESTIMATES WITH A GIVEN RISK,
RICCATI DIFFERENTIAL EQUATIONS,
AND A NEW PROOF OF A THEOREM OF BROWN

by

Anirban DasGupta   and   William E. Strawderman

Purdue University   Rutgers University

Abstract

For the canonical problem of estimating the mean of a multivariate normal distribution with a known covariance matrix using a squared error loss, we give a general method for finding estimates that have identical risk functions as that of a given inadmissible estimate. In the case of more than one dimension, the estimates considered are spherically symmetric, but in one dimension no such assumption is made. Generally speaking, we characterize all estimates which have the risk duplication property. It is proven that every James - Stein estimator except $(1 - \frac{p-2}{\|\underset{\sim}{X}\|^2})\underset{\sim}{X}$ can be duplicated in risk by infinitely many estimators. A general theorem is presented from which a principal inadmissibility result for spherically symmetric estimates in Brown (1971) follows. This and the other results all basically depend on solving Riccati differential equations of an appropriate kind. A curious result is that two smooth estimates whose graphs intersect cannot have identical risk. Several results for the entire Discrete Exponential family and the Binomial case demonstrate that the phenomena are fundamentally different in continuous and discrete cases. Our results indicate a new method for constructing explicit dominating estimates that may work in many problems.

1. Introduction

In many statistical problems involving an unknown parameter, it is a commonly known fact that two different estimators cannot be unbiased for a parameter, or in general, cannot have identical expectations under all values of the parameter. This is a consequence of the completeness of the statistical family under consideration, when completeness is present. It is a less commonly asked question if two different estimators can have the same risk under all values of the parameter: notice that completeness of the family does not preclude identical risks of two different estimators. The case of the unbiased estimator X and the James-Stein estimator $(1 - 2(p - 2)/|X|^2)X$ for the mean of a $N(\theta, I)$ distribution under squared error loss is a classic example. One can rephrase this phenomenon in a different although equivalent way: given a function $r(\theta)$ of the parameter θ, is it possible to characterize (may be with some constraints on the form) all estimates which have $r(\theta)$ as their risk functions? From a purely decision theoretic viewpoint, there is clearly nothing to choose from among these estimators. Still, it would be a novel feature of decision theory that estimators which are different from each other with probability 1 under all θ, can actually result in exactly the same risk for all θ! In addition as our results show in the particular problem of estimating a multivariate normal mean, the risk duplicability will imply a number of new interesting phenomena and provide better insight into existing important results.

The main thrust in this article is on the problem of estimating the mean θ of a univariate or multivariate normal distribution with identity covariance matrix under squared error loss. Multivariate will include the case $p = 2$. We will present a general theory and technique for characterizing all estimators of θ that have a given function $r(\theta)$ as their risk. In the multivariate case, the normal estimators considered will be restricted to be spherically symmetric. A typical result will look like the following: let $\delta_0(X)$ be a given estimator of a θ; then $\delta_1(X)$ has identical risk as $\delta_0(X)$ for all values of θ, if and only if $\delta_1(X)$ has a certain explicit form, the form depending obviously on the problem at hand. For example, every James-Stein estimator $(1 - c/|X|^2)X$ can be duplicated exactly in risk if $c \neq p - 2$, by other spherically symmetric estimators; the estimator with $c = p - 2$ cannot be. In the one dimensional normal case, we show new estimators which have constant risk under all values of θ. These are two particular types of many examples that follow from

the general theory.

Besides, any estimator $\delta_0(X)$ which is duplicable in risk is automatically inadmissible: one can simply take any other estimator $\delta_1(X)$ with the same risk and take a convex combination of $\delta_0(X)$ and $\delta_1(X)$. This is better than each of $\delta_0(X)$, $\delta_1(X)$. It turns out that the estimators which duplicate X in risk in the multivariate normal problem happen to have an extreme value property in the class of Efron-Morris (1976) estimators. Our results also unearth a hitherto unknown feature of minimax estimation of normal means: that contrary to common belief, an estimator $(1 - r(|X|^2)/|X|^2)X$ can be minimax even if $r(.)$ is monotone decreasing! Therefore, one consequence of our results in the normal case is that the class of minimax estimators for $p \geq 3$ is really much richer than it was previously thought to be.

The mathematics of this article uses the Stein unbiased estimate of risk as the fundamental step, and then makes rather heavy use of the theory of linear differential and Riccati equations. Therefore, the aspects of linear differential equations that are crucial for this article are adequately presented in a brief technical appendix. Section 2 treats the univariate normal case and presents a general result and two examples. James-Stein estimators are considered in Section 3 and a general inadmissibility result on spherically symmetric estimators is presented in Section 4. A more direct proof (using our methods) of an important inadmissibility result in Brown (1971) is given in Section 5. Section 6 presents some brief results for the discrete Exponential family.

The main contributions of our article are the following:

a. We exhibit new phenomena in the context of minimax estimation of normal means.

b. We exhibit a number of examples that we believe are quite intriguing.

c. We present a more direct proof of an inadmissibility result in Brown (1971).

d. We present a new automatic method of obtaining an explicit improved estimator by using differential equations rather than differential inequalities; this method should work in a broad variety of non-normal problems as well, spherically symmetric ones in particular.

e. We demonstrate a new role of ordinary differential equations in statistical inference.

f. We give some evidence that discrete problems are fundamentally different.

2. The Univariate Normal Case

2.1. Basic Assumption. Since the Stein unbiased estimate of risk is used as the starting point for all of our results, we will make the assumption that every estimator under consideration satisfies the following assumption A:

Assumption A: Let $X \sim N(\theta, 1)$; let $\delta(X)$ be an estimate of θ. Write $\delta(X) = X + g(X)$; assume $g(X)$ has finite second moment under all θ, is once differentiable and $g(x)\exp(-(\theta - x)^2/2) \to 0$ for all θ, as $|x| \to \infty$. In particular, the risk of $\delta(X)$ admits the expression

$$R(\theta, \delta) = 1 + E\{g^2(X) + 2g'(X)\} \tag{2.1}$$

2.2. A General Theorem On Nonduplicability. Perhaps the simplest example of two different estimators with identical risk in the univariate normal case are $X + 1$ and $X - 1$. Notice the interesting fact that these two estimators have graphs that never intersect; the following theorem says if the graphs of two smooth estimators intersect, then they cannot have identical risk for all θ.

Theorem 1. Let $\delta_0(X)$ and $\delta_1(X)$ be two estimators of θ and assume both are analytic functions of X. Suppose there exists $x = k$ (say) such that $\delta_0(x) = \delta_1(x)$. Then $R(\theta, \delta_0)$ cannot equal $R(\theta, \delta_1)$ for all θ.

Proof: Write $\delta_1(X) = \delta_0(X) + g(X)$. Then, by virtue of (2.1),

$$\begin{aligned}
0 &= R(\theta, \delta_1) - R(\theta, \delta_0) \\
&= E\{g^2(X) + 2g(X)\delta_0(X) - 2Xg(X) + 2g'(X)\} \tag{2.2}
\end{aligned}$$

$\Rightarrow g^2(x) + 2g(x)\delta_0(x) - 2xg(x) + 2g'(x) = 0$ for all x (since the normal family is complete and g' is continuous)

By hypothesis, $g(k) = 0$. (2.2) therefore implies $g'(k) = 0$. Since (2.2) is an identity for all x, and since $g(x)$ is infinitely differentiable as it is the difference of two analytic functions, a differentiation of (2.2) yields:

$$2g(x)g'(x) + 2g'(x)\delta_0(x) + 2g(x)\delta_0'(x) - 2g(x) - 2xg'(x) + 2g''(x) = 0, \quad \text{for all } x \tag{2.3}$$

Using $x = k$ in (2.3) gives $g''(k) = 0$.

An induction argument leads to $g^{(n)}(k) = 0$ for all n, i.e., g and all its derivatives vanish at k. Since analytic functions can be expanded as $g(x) = \Sigma(x - k)^n/n!\ g^{(n)}(k)$, it follows that $g(x) = 0$ for all x, giving the assertion of the theorem.

Remark. Since generalized Bayes (GB) estimators are analytic in this problem, Theorem 1 says it is hard to duplicate the risk of one GB estimate by using another GB estimate.

We now give two examples.

Example 1. Consider the estimator $X + k$. Then $\delta(X) = X + g(X)$ has the same risk as $X + k$ if and only if

$$g^2(x) + 2g'(x) - k^2 = 0 \quad \text{for all} \ x \tag{2.4}$$

Equation (2.4) is a Riccati differential equation of the form

$$g'(x) + r(x)g^2(x) + s(x) = 0, \text{with} \quad r(x) = 1/2, \text{and} \ s(x) = -k^2/2. \tag{2.5}$$

It follows from the general theorems of Riccati differential equations (see the appendix) that the general solution to (2.4) has the form

$$g(x) = k\{Aexp(kx/2) - Bexp(-kx/2)\}/\{Aexp(kx/2) + Bexp(-kx/2)\} \tag{2.6}$$

among which, the estimates $X + g(X)$ with $A, B \geq 0$, not both zero, give rise to estimates with finite risk. Other values of A, B solve (2.4) but do not give estimators with finite risk.

Therefore, there is a one parameter family of estimators having the constant risk $1 + k^2$. Of these, $B = 0$ results in the estimator $X + k$, and $A = B$ results in the estimator $X + ktanh(kX/2)$. The correction term $ktanh(kX/2)$ arises in unrelated work in Casella and Strawderman (1981).

We will now ask a question that appears to be playful at first sight, but the answer, in our view, is a bit surprising. There could be a deeper phenomenon here of which the following is a special case, but we do no know.

Proposition 2. Consider the class of all estimators (2.6) that duplicate the risk of the linear estimator $X + k$. Define the curvature of an estimator $d(X)$ as $(d''(X))^2$. Then,

for each estimator in (2.6), except the linear one $X + k$, the mean curvature $\int (d''(x))^2 \, dx$ exists, and is exactly equal to the same constant $2/15 \ k^5$.

Remark: Our actual intention was to find the most nonlinear estimate with the risk of $X + k$. The Proposition says that every nonlinear estimate of this kind is exactly equally nonlinear on the average. Note that the more common definition of curvature is $(d''(x))^2 / (1 + d'(x))^2)^{3/2}$. Approximation theorists sometimes use $(d''(x))^2$ as curvature. See, for instance, Wahba (1990, pp. viii).

It is simple common sense to think that the result of Proposition 2 is not a pure accident, but has something to do with their equal risks. It will therefore be enlightening to find a proof of the Proposition that uses the equality of risks, instead of a dull integration proof. This is what we now do.

Proof of Proposition 2:

Step 1. It is enough to prove the Proposition for $k = 1$, as the general k will follow on just a change of scale.

Step 2. For $k = 1$, write the estimate as

$$X + \frac{(exp(x/2) - yexp(-x/2))}{(exp(x/2) + yexp(-x/2))}$$
$$= X + g(X)(\text{say}).$$

Thus, $d''(x) = g''(x)$ for all x.

Step 3. By Stein's identity, g satisfies $g^2 + 2g' = 1$ Hence, each of the following follows easily:

$$g'' = -gg'$$
$$g'' = -g/2 + g^3/2$$
$$gg'' = -g'g^2$$

Step 4. Therefore,

$$\int (d''(x))^2 \, dx$$

$$= 1/3 \int g'(g^3)' dx$$

$$= -1/3 \int g''(2g'' + g) dx \quad \text{(by Step 3 and integration by parts)}$$

$$= -2/3 \ LHS - 1/3 \int gg'' dx$$

$$= -2/3 \ LHS + 1/9 \int (g^3)' dx \quad \text{(by Step 3)}$$

$$= -2/3 \ LHS + 2/9$$

$$\Longrightarrow 5/3 \ LHS = 2/9,$$

giving the Proposition.

Example 2. Estimators with quadratic risk. Consider the function $r(\theta) = a\theta^2 + b\theta + c$; clearly, linear estimators of the form $sX + t$ have quadratic risks. Indeed, the risk of $sX + t$ is evidently the quadratic

$$(s - 1)^2 \theta^2 + 2t(s - 1)\theta + t^2 + s^2. \tag{2.7}$$

Notice that in the notation of $r(.)$ defined above, these quadratics have the property $a \geq 0, c \geq 0$; it is completely straightforward to show that only such quadratics can be risk functions, and furthermore $c = 0$ corresponds only to the estimator identically equal to 0; also $a = 0$ implies $b = 0$, reducing the risk to a constant, which was the topic of the previous example. We shall therefore assume that $a > 0, c > 0$; furthermore, for linear estimators the constant c is $(1 + \sqrt{a})^2 + b^2/(4a)$. Thus we will consider quadratic risk functions $a\theta^2 + b\theta + c$, in which $a > 0, b$ is a real number, and c is as above. We will ask if nonlinear estimators can have the same risk as that of linear ones. The estimators we duplicate in risk are thus of the form $(1 + \sqrt{a})X + b/(2\sqrt{a})$ where b is any arbitrary real.

As usual, write a generic estimator as $X + g(X)$; it turns out that it is convenient to write $g(X)$ itself as $g(X) = \sqrt{a}X + h(X)$.

With this notation, the risk identity $R(\theta, X + g(X)) = a\theta^2 + b\theta + c$ is equivalent to

$$g^2(x) + 2g'(x) - (ax^2 + bx + c - 1 - a) = 0$$

$$\Leftrightarrow h' + 1/2 \ h^2 + \sqrt{a}xh + 1/2((\sqrt{a} + 1)^2 - c - bx) = 0 \tag{2.8}$$

The Riccati equation (2.8) reduces to the more convenient form

$$f' + rf^2 + s = 0 \tag{2.9}$$

where $f(x) = h(x).exp[\sqrt{a}/2 \ x^2], r(x) = 1/2 \ exp(-\sqrt{a}/2 \ x^2)$, and $s(x) = -1/2 \ (b^2/4a + bx) \ exp(\sqrt{a}/2 \ x^2)$.

The solution to (2.9) is obtained by solving the auxiliary second order linear differential equation $(1/rz')' + sz = 0$, which reduces to

$$z'' + \sqrt{a}xz' - 1/4 \ (b^2/(4a) + bx) \ z = 0 \tag{2.10}$$

The functions

$$u(x) = exp(bx/(4\sqrt{a}))$$

$$\text{and} \quad v(x) = exp(bx/(4\sqrt{a}))\Phi(\sqrt{a}x + b/(2\sqrt{a})) \tag{2.11}$$

are two independent solutions to (2.10), which ultimately result in

$$\delta(X) = (1 + \sqrt{a})X + \frac{2(Au'(X) + Bv'(X))}{(Au(X) + Bv(X))} \tag{2.12}$$

as the general form of an estimator having the risk $a\theta^2 + b\theta + c$, i.e., the risk of the linear estimator $(1 + \sqrt{a})X + b/(2\sqrt{a})$.

The estimators in (2.12) can be simplified to the final form

$$\delta(X) = (1 + \sqrt{a})X + b/(2\sqrt{a}) + \frac{2\sqrt{a}\phi(\sqrt{a}X + b/(2\sqrt{a}))}{C + \Phi(\sqrt{a} \ X + b/(2\sqrt{a}))} \tag{2.13}$$

all of which have finite risks if $C \geq 0$ or $C \leq -1$.

Thus, (2.13) finally characterizes a class of estimators with risks matching the risk of the linear estimator $(1 + \sqrt{a})X + \frac{b}{2\sqrt{a}}$.

3. James-Stein Estimates

3.1. Basic Assumption. Because of historical importance and also for transparent presentation of the mathematics involved in arriving at the results, we will separately treat Stein estimates of the form

$$\delta_c(X) = (1 - c/t).X; \tag{3.1}$$

in the above, $0 \leq c \leq 2(p-2)$, and t denotes $|x|^2$. The case $c = 0$ corresponds to the problem of duplicating the risk of X; therefore, the estimator X will not be considered separately. For $p = 2$, the problem itself is uninteresting as the only relevant c is $c = 0$. Analogous to Assumption A in Section 2, if it will be assumed that EVERY estimate under consideration satisfies the following Assumption B:

Assumption B: Let $X \sim N_p(\theta, I)$; let $\delta(X)$ be an estimator of θ. Write $t = |X|^2$, and $\delta(X) = (1 - g(t)/t).X$. Assume $g^2(t)/t$ has a finite expectation under all θ, g is once continuously differentiable, and $g(t)\exp(-|x-\theta|^2/2) \to 0$ for all θ, as $t \to \infty$. In particular, the risk of $\delta(X)$ admits the expression

$$R(\theta, \delta) = p + E\{g^2(t)/t - 2(p-2)g(t)/t - 4g'(t)\} \qquad (3.2)$$

3.2. James-Stein Estimates. Theorem 3. Consider the estimator $\delta(X) = (1 - g(t)/t)X$; then $R(\theta, \delta) \equiv R(\theta, \delta_c)$ if and only if $g(.)$ has the form

$$g(t) = \frac{(cA - B(c - 4\alpha)t^{c/2-\alpha})}{(A + Bt^{c/2-\alpha})}, \qquad (3.3)$$

where $\alpha = p/2 - 1$, and A, B are arbitrary nonegative constants, not both zero.

In particular, $\delta(X)$ has constant risk p if and only if $g(.)$ has the form $g(t) = 2(p-2)/(1 + \gamma t^\alpha)$, where γ is an arbitrary nonnegative extended real number. If $c = p - 2$, there exist no estimators $\delta(X)$ such that $R(\theta, \delta) = R(\theta, \delta_c)$ for all θ.

Proof: The Stein estimator $\delta_c(X)$, by familiar computations, has the unbiased estimate of risk

$$c(t) = \frac{(c^2 - 4c\alpha)}{t} + 2(\alpha + 1) \qquad (3.4)$$

Hence, $\delta(X) = (1 - g(t)/t)X$ has the same risk if and only if

$$g^2(t)/t - 4\alpha g(t)/t - 4g'(t) = (c^2 - 4c\alpha)/t \quad \text{a.e.,} \qquad (3.5)$$

and hence for all t.

After this, the argument follows the usual path; the auxiliary second order linear differential equation is

$$t^2 z''(t) + (\alpha + 1)tz'(t) + (4c\alpha - c^2)/16 \; z(t) = 0, \qquad (3.6)$$

with the independent solutions

$$z_1(t) = t^{-c/4} \quad \text{and} \quad z_2(t) = t^{\frac{c-4\alpha}{4}}. \tag{3.7}$$

This then results in the stated solutions for the Riccati equation (3.5) and the first part of the Theorem follows. Note that the stated estimates indeed have finite risk. The case of the constant risk p follows immediately on simply substituting $c = 4\alpha = 2(p - 2)$ (or one can equivalently substitute $c = 0$) in the general formula.

The stated result for $c = p - 2$ is well known (see, e.g., Bock (1986), Brown (1986)) and hence a proof using our methods will be omitted.

3.3. Discussion of Theorem 3. First consider the estimators which have risk identical to that of X. As stated above, they have the form

$$\delta(X) = (1 - g(t)/t)X, \quad \text{where} \quad g(t) = 2(p - 2)/(1 + \gamma t^{p/2 - 1}), \quad \text{where} \quad \gamma \geq 0. \tag{3.8}$$

This is clearly interesting, because the function $g(t)$, although bounded between 0 and $2(p - 2)$, is actually monotone decreasing. The standard Baranchik class corresponds to g which are monotone increasing, and the condition was relaxed in Efron and Morris (1976). Efron and Morris (1976) gave the weaker condition:

$$f(t) = \frac{t^{p/2 - 1}g(t)}{(2(p - 2) - g(t))} \quad \text{is monotone nondecreasing.}$$

Very interestingly, if one sets $f = $ constant, then one exactly gets the estimators in (3.8). Thus, these estimators, in this sense, form the extreme points of the Efron-Morris (1976) minimax class.

b. Next observe that $\gamma = 0$ corresponds to the James-Stein estimator with $c = 2(p-2)$, while $\gamma = \infty$ corresponds to the usual estimator X. Therefore, an estimate with $0 < \gamma < \infty$ corresponds to a data dependent convex combination of these two estimators.

c. Since each $\gamma \geq 0$ provides a duplicate of X in risk, an estimate which is a fixed (free of X) mixture of the estimates, the mixture being on the free constant γ, is a strict improvement on X. The new estimate thus obtained is also spherically symmetric, and shares

other properties of the parent estimates; for instance, the corresponding $g(t)$ function is again monotone decreasing.

d. The estimates given in Theorem 3 are all improved, as usual, by their positive part versions. If one takes the positive parts of the estimates in Theorem 3 that duplicate X in risk, and then takes a mixture of the positive parts, something interesting emerges. The resultant estimate is not necessarily a positive part estimate; quite interestingly, it is of the form

$$\begin{aligned} \delta(X) &= (1-w)X && \text{if } t \le 2(p-2) \\ &= (1 - 2(p-2)w/t)X && \text{if } t > 2(p-2), \end{aligned} \tag{3.9}$$

if the mixing distribution G on $\gamma \ge 0$ is taken to be $G\{0\} = 1 - G\{\infty\} = w$. This estimate in (3.9) is what researchers in the area call a restricted risk Bayes estimate: an estimate that minimizes the Bayes risk with respect to a prior in the class of estimates with a specified upper bound on the risk.

e. Note that a result parallel to Theorem 3 cannot hold if one starts with the positive part versions $(\delta_c(X))t$ instead of $\delta_c(X)$ itself, by virture of results in Bock (1986), Brown (1986).

Thus, the general message of the points made in the discussion above is that some quite interesting and novel features of minimax multivariate estimation seem to come out of the duplicators in Theorem 3.

4. General Estimates in the Multivariate Normal Case

4.1 A General Inadmissibility Theorem. The methods outlined in the previous sections can be used to characterize, for a general estimator $\delta(X)$ satisfying certain conditions, all other estimators which duplicate $\delta(X)$ in risk. A usual argument will then imply that $\delta(X)$ is inadmissible, and our methods automatically produce infinitely many improvements in the process. We will give one general result to this effect below. The

A. DASGUPTA AND W. E. STRAWDERMAN

following notation will be used:

$$\delta_1(X) = (1 - g_1(t)/t)X \quad \text{is a given estimator;}$$

$$z_1(x) = e^{-\int \frac{g_1(x)}{4x}}$$

$$F(x) = \int \frac{x^{-(\alpha+1)}}{z_1^2(x)} \quad \text{where} \quad \alpha = p/2 - 1 \tag{4.1}$$

$$g(t) = g_1(t) - 4BtF'(t)/(A + BF(t))$$

$$\delta(X) = (1 - g(t)/t)X$$

(Where A, B are constants to be specified later)

We will make the following assumption:

Assumption C: Suppose there exist constants A, B, not both zero, such that

$$E\{\frac{\sqrt{t}F'(t)}{A + BF(t)}\}^2 < \infty, \quad \text{for all} \quad \theta$$

Theorem 4. Let $\delta_1(X)$ satisfy Assumption C. Then $\delta_1(X)$ is inadmissible.

Proof: We will outline the main steps.

Step 1. $R(\theta, \delta) = R(\theta, \delta_1)$ for all θ gives:

$$\frac{g^2(t)}{t} - 4\alpha\frac{g(t)}{t} - 4g'(t) = \frac{g_1^2(t)}{t} - 4\alpha\frac{g_1(t)}{t} - 4g_1'(t) \quad \text{for all } t \tag{4.2}$$

Step 2. On transforming to $f(t) = t^\alpha g(t)$, (4.2) reduces to the Riccati equation

$$f' - \frac{t^{-\alpha-1}f^2}{4} + t^\alpha v(t)/4 = 0 \quad \text{for all } t \tag{4.3}$$

where

$$v(t) = \frac{g_1^2(t)}{t} - 4\alpha\frac{g_1(t)}{t} - 4g_1'(t)$$

Step 3. The Riccati equation (4.3) is solved by solving the linear differential equation

$$(1/r \; z')' + sz = 0,$$

$$\text{where} \quad r(t) = -\frac{t^{-\alpha-1}}{4}$$

$$s(t) = \frac{t^\alpha v(t)}{4} \tag{4.4}$$

Step 4. Writing (4.4) as $z'' + b_1 z' + b_0 z = 0$, where

$$b_1 = \frac{(\alpha + 1)}{t}$$

$$b_0 = -\frac{v(t)}{16t} \tag{4.5}$$

it follows that if $z_1(t)$ is one particular solution of (4.4), then a second independent solution, by Abel's identity, is

$$z_2(t) = z_1(t) \cdot \int \frac{w(t)}{z_1^2(t)},$$

where $w(t) = e^{-\int b_1(t)}$ and $\int$ denotes a primitive; see Ross (1989) for Abel's identity which gives a second independent solution of a second order equation knowing one particular solution.

The particular solution $z_1(t)$ is actually available here; it is defined in (4.1).

Step 5. This now produces the general solution of (4.3) as

$$f(t) = \frac{1/r(t)(Az_1'(t) + Bz_2'(t))}{(Az_1(t) + Bz_2(t))} \tag{4.6}$$

where A, B are not both zero.

Step 6. Transforming back to $g(t) = t^{-\alpha} f(t)$ produces an estimator duplicating the risk of $\delta_1(X)$. Assumption C is used in verifying finiteness of risk of the new estimator.

Step 7. Since we can find an estimator different from $\delta_1(X)$ having the same risk as $\delta_1(X)$, it follows $\delta_1(X)$ must be inadmissible.

Theorem 4 will be used in the next section to give a more direct proof of a theorem in Brown (1971).

5. Inadmissibility of Sphericallly Symmetric Estimates

Consider estimating the mean of a multivariate normal distribution under squared error loss using a GB estimate with respect to a spherically symmetric prior G (note G may or may not have a Lebesgue density). Since G is spherically symmetric, it follows that the marginal of X is so as well, and indeed there is a marginal density $m(x)$ of the form $m(x) = f(t)$, where t denotes $|x|^2$. $m(x)$ can also be written obviously in the form

$f^*(r)$, where r denotes $|x|$. This is the notation used in section 6 in Brown (1971). Brown shows the following:

Theorem. Suppose G is such that

$$\int_1^\infty (r^{p-1} f^*(r))^{-1} dr < \infty; \tag{5.1}$$

then the GB estimate with respect to G is inadmissible.

We will give a new and in some sense a more direct proof of Brown's theorem using our Theorem 4. Note that our result is slightly stonger than Brown's, as we exhibit that Stein's unbiased estimate of the risk difference can be solved for equality to zero under condition (5.1). Brown (1971, 1986) shows that this can be done too, provided the inequality can be solved. Due to this subtle point, our result is a bit stronger than Brown's.

Theorem 5. Assume Brown's condition holds. Then the GB estimate with respect to G can be duplicated in risk by another spherically symmetric estimator and in particular, it is inadmissible.

Proof: We break the proof into several key steps; note that Brown works with $r = |x|$, while our duplicability theorem works with $t = |x|^2$. It is important to constantly keep this notational difference in mind.

Step 1. In our notation, the estimator is of the form $(1 - g(t)/t)X$, which by a familiar identity equals $X + \nabla m(X)/X$, where $\nabla(.)$ denotes gradient.

Using now the fact that $m(X) = f(t)$, and the same notation as in (4.1), one has:

$$f(t) = z_1^2(t).$$

Step 2. Again using the same notation as in (4.1), one sees

$$F'(t) = \frac{t^{-p/2}}{z_1^2(t)}.$$

Step 3. By making the change of variable from r to t in Brown's theorem, one also sees that Brown's hypothesis is equivalent to:

$$\int_1^\infty F'(t) dt < \infty$$

i.e., F is bounded on $[1, \infty]$. Let k be such that $F(t) < k$ for all t.

Step 4. To show that the GB estimate with respect to G can be duplicated in risk, it is sufficient to show according to our Theorem 4 that

$$E\{\frac{t(F'(t))^2}{(A + BF(t))^2}\} < \infty \quad \text{for all} \quad \theta,$$

for some constants A, B, not both zero.

By a standard convexity argument valid for Exponential families, this will follow if we can show

$$E\{\frac{t(F'(t))^2}{(A + BF(t))^2}\} < \infty,$$

where the expectation now is with respect to the marginal density of X.

Step 5. Using only careful algebra, one now sees that the finiteness of the stated expectation in Step 4 under the marginal is equivalent to

$$\int_0^\infty \frac{F'(t)}{(A + BF(t))^2} dt < \infty.$$

Step 6. If in particular we choose $B = -1$, and $A = k$, then this is the same as

$$\int_0^\infty \frac{F'(t)}{(k - F(t))^2} dt < \infty.$$

Step 7. Since $F'/(k - F)^2 = $ derivative of $1/(k - F)$, it follows that the estimator under consideration is duplicable in risk if $1/(k - F(0)) - 1/(k - F(t))$ is bounded.

Step 8. By Step 3, each of these terms is indeed bounded ($F(0)$ may be $-\infty$, which is of no concern, however).

This completes the proof of Theorem 5.

6. Discrete Exponential Family

In order to make the picture a bit more complete, in this brief section we will indicate that discrete problems show fundamentally different phenomena in this context; i.e., risk

duplicability is not as persuasive or even rare in the discrete Exponential family. We will state two results that contrast the earlier results of the normal case. The proof of Theorem 6 goes through for the general discrete Exponential family with pmf:

$$P_\theta(X = x)$$
$$= \theta^x \beta(\theta) h(x), \tag{6.1}$$
$$x = 0, 1, 2, 3, \ldots$$

Theorem 6. Let $X \sim$ Poisson (λ); let $\delta_i(X), i = 0, 1$ be two estimates of λ, each nonnegative. Then $R(\lambda, \delta_1)$ cannot be equal to $R(\lambda, \delta_0)$ for all values of λ under squared error loss.

Proof: Let $\delta_1(X) = \delta_0(X) + g(X)$. Using the fact (See Hwang (1982) and also Clevenson and Zidek (1975)) that for any function g with finite second moment, $E(\lambda g(X)) = E(Xg(X - 1))$, where $g(-1)$ is defined to be zero for purposes of specificity, one has, by completeness,

$$R(\delta_0, \lambda) = R(\delta_1, \lambda), \quad \text{for all } \lambda$$
$$\Longrightarrow g^2(X) + 2g(X)\delta_0(X) - 2Xg(X - 1) = 0. \tag{6.2}$$

For $x = 0$, (6.2) gives

$$g^2(0) + 2g(0)\delta_0(0) = 0, \tag{6.3}$$

which will imply $g(0) = 0$.

It follows by induction that $g(x) = 0$ for all $x \geq 0$, proving the theorem.

Theorem 7. Let $X \sim$ Binomial (n, p); let $\delta_i(X), i = 0, 1$ be two different estimates of p under squared error loss, each taking values in $[0, 1]$, and strictly positive if $x > 0$. Then $R(\delta_0, p)$ cannot equal $R(\delta_1, p)$ for all values of p.

We will omit the proof as it follows the same inductive reasoning of Theorem 6, although the proof is not really trivial.

7. Conclusion

We have presented a new method for explicit construction of dominating estimates by using the theory of ordinary differential equations. Our method has a very good like-

lihood of applying widely: for instance for estimating natural parameters in the general Exponential family. Hopefully, our results will lead to further work that will contribute to better scientific understanding of the general area of parametric estimation.

Acknowledgement

References

Bock, M. E. (1986). Shrinkage estimators : Pseudo-Bayes rules for normal mean vectors, Stat. Dec. Theory and Rel. Topics, IV, 1,281-298.

Brown, L. D. (1971). Admissible estimators,recurrent diffusions and insoluble boundary value problems, Ann. Math. Stat., 42, 855-903.

Brown, L. D. (1986). The differential inequality of a statistical estimation problem, Stat. Dec. Theory and Rel. Topics, IV, 1, 299-324.

Casella, G. and Strawderman, W. E. (1980). Estimating a bounded normal mean, Ann. Stat., 9, 870-878.

Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the means of independent Poisson laws, J. Amer. Stat. Assoc., 70, 698-705.

Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution, Ann. Stat., 4, 11-21.

Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with appilications to Poisson and negative Binomial cases, Ann. Stat., 10, 857-867.

James, W. and Stein, C. (1961). Estimation with quadratic loss, Proc. Fourth Berkeley Symp., 1, 311-319.

Ross, S. (1989). Introduction to Ordinary Differential Equations, Wiley, New York.

Stigler, S. M. (1990). 1988 Neyman Lecture: Galtonian perspective on Shrinkage estimators, Stat. Sc., 5, 1, 147-155.

Wahba, G. (1990). Spline Models for Observational Data, *SIAM*, Philadelphia.

8. Appendix

A. The Riccati Equation. The Riccati differential equation has the form

$$f'(x) + r(x)f^2(x) + s(x) = 0. \tag{8.1}$$

If $r(x) \neq 0$ and $r(x), s(x)$ are both continuous for all x under consideration, then the general solution of (8.1) has the form

$$f(x) = \frac{1}{r(x)} \cdot \frac{Au'(x) + Bv'(x)}{A\,u(x) + Bv(x)}, \tag{8.2}$$

where A, B are real, not both zero, and $u(x), v(x)$ are two linearly independent solutions of the linear differential equation

$$\frac{1}{r(x)} Z''(x) + (\frac{1}{r(x)})' Z'(x) + s(x)Z(x) = 0. \tag{8.3}$$

B. Suppose (8.3) can be transformed into a simpler differential equation with constant coefficients:

$$h''(x) + a_1 h'(x) + a_0 h(x) = 0. \tag{8.4}$$

Then (8.3) can be solved by using the following fact:

Fact

1. If the characteristic polynomial $t^2 + a_1 t + a_0$ has two distinct real roots t_1 and t_2, then two linearly independent solutions of (8.4) are $e^{t_i x}$, $i = 1, 2$.

2. If the characteristic polynomial has a double root t_0, then two linearly independent solutions of (8.4) are $x^i e^{t_0 x}$, $i = 0, 1$.

3. If the characteristc polynomial has complex roots $a \pm ib$, then two linearly independent solutions of (8.4) are $e^{ax} \sin bx$ and $e^{ax} \cos bx$.