EVALUATING ESTIMATORS CONDITIONALLY

By GEORGE CASELLA[1]

Cornell University

BU-797-M

November, 1982

[1] Research supported by National Science Foundation Grant No. MCS81-02541.

AMS 1980 subject classifications. Primary 62C99; secondary 62F10.

Key words and phrases. Multivariate normal mean; conditional properties.

Summary

Classical decision-theoretic procedures have often been criticized on the grounds that their conditional performance is undesirable. It is argued that this is not a deficiency of the classical approach, but rather a case of not considering all information when evaluating a procedure. For the case of estimating a multivariate normal mean, it is shown that a simple conditional criterion, when coupled with the classical criterion of (unconditional) risk evaluation, leads to a class of estimators which are good both conditionally and unconditionally. In particular, the ordinary James-Stein estimator is excluded from this class, while its positive-part version is included. This new conditional criterion is also related to more familiar criteria.

EVALUATING ESTIMATORS CONDITIONALLY

By GEORGE CASELLA[1]

Cornell University

1. Introduction. Classical statistical procedures have often been criticized on the grounds that, when evaluated conditionally, their performance is unreasonable. To a certain extent this criticism is justified but, on the other hand, it seems somewhat unreasonable to criticize a procedure for non-optimality against a criterion when the criterion played no role in the construction of the procedure. Classical procedures are constructed to be optimal against unconditional criteria; those that involve averaging over the sample space. In doing this, classical procedures guarantee a certain degree of repeatability, a guarantee which cannot be made by Bayesians or likelihoodists.

In order to evaluate a procedure conditionally, there must be a subset of the sample space (a 'recognizable' subset, as named by Fisher, 1956) in which an experimenter is interested, indeed, in which the observations are expected to lie. If there is no interest in a particular recognizable subset, then there is no sense in evaluating the procedure conditionally. Thus, it seems that if a conditional evaluation is expected, prior knowledge (in some sense) is indicated. If such prior knowledge is not taken into account when a procedure is constructed, this is a fault of the method, not of the procedure.

---

[1] Research supported by National Science Foundation Grant No. MCS81-02541.

AMS 1980 subject classifications. Primary 62C99; secondary 62F10.

Key words and phrases. Multivariate normal mean; conditional properties.

The purpose of this paper is not to criticize classical procedures, but rather to show that when procedures are constructed taking all information into account, the results are good both conditionally and unconditionally. Recently, Berger and Wolpert (1982), commenting on conditional evaluation of classical procedures, remarked, "Of course, many researchers in the field study the issue solely to point out inadequacies in the frequentist viewpoint, and not to recommend specific conditional frequentist measures." For the problem of estimating a multivariate normal mean, we propose a specific conditional measure which, when coupled with the unconditional property of minimaxity, yields a class of estimators that perform well both conditionally and unconditionally.

Many traditional classical procedures are constructed without any regard to prior information, for example, a classical estimator of location is usually translation invariant. Research beginning with the fundamental papers of Stein (1955) and James and Stein (1961) showed that the best translation invariant estimator of a multivariate normal mean can be beaten (unconditionally) by an estimator which is not translation invariant. If X is an observation from a p-variate normal distribution ($p \geq 3$) with mean vector θ and covariance matrix the identity, then the best translation invariant estimator, X, is beaten by

$$\delta(X) = \theta_0 + \left(1 - \frac{p-2}{|X-\theta_0|^2}\right)(X - \theta_0) \quad , \tag{1.1}$$

where θ_0 is any fixed constant. If $\delta(X)$ of (1.1) is to be used, θ_0 should represent the best prior guess at θ. If this is the case, then it is reasonable to require $\delta(X)$ to perform well on a recognizable subset of the form $\{X : |X - \theta_0| \leq t\}$, for some fixed t. The fact that $\delta(X)$ performs poorly near θ_0 ($\delta(X)$ has a singularity at $X = \theta_0$), coupled with the fact that, unconditionally, $\delta(X)$ is superior to X, may be viewed as a deficiency of classical decision theory. However, we do not view it as such. The problem, rather, is that the proper information was not taken into account when the estimator was evaluated. If, in fact, one requires both unconditional domination of X, and good conditional performance when X is near θ_0, one can work within classical decision theory to obtain such estimators. As will be seen, $\delta(X)$ of (1.1) is then quickly eliminated from consideration.

In Section 2 we define a simple conditional criterion and show that this criterion eliminates the ordinary Stein estimator (1.1) as an alternative to X, but establishes the positive-part Stein estimator as a reasonable alternative. In Section 3 this conditional criterion, which we call conditional minimaxity, is related to the more familiar notion of admissibility. In particular, it is shown that admissible estimators (hence proper Bayes estimators) are conditionally minimax. Furthermore, it is shown that failure of an estimator to be conditionally minimax is constructive in the sense that a uniformly better estimator can immediately be obtained. In Section 4 conditional minimaxity is related to another conditional criterion, the existence of relevant betting procedures. There it is shown that the absence of conditional minimaxity implies the existence of relevant betting procedures. The criticism is again constructive, for a relevant betting procedure can immediately be constructed if the estimator is not conditionally minimax. Section 5 contains some comments, and there is also an Appendix, which contains technical arguments needed in Section 2. The Appendix also contains a lemma giving an identity involving noncentral chi-squared expectations, and may possibly be of independent interest.

2. The Conditional Property. Let X be an observation from a p-variate ($p \geq 3$) normal distribution with mean θ and identity covariance matrix. For any estimator $\delta(X)$ of θ, the loss in estimating θ by $\delta(X)$ is

$$L(\theta,\delta) = |\theta - \delta(X)|^2 \quad ,$$

where $|\cdot|$ denotes Euclidean distance. Define $\Delta_\theta(\delta)$ to be the difference in loss between $\delta(X)$ and X at θ, i.e.,

$$\Delta_\theta(\delta) = |\theta - \delta(X)|^2 - |\theta - X|^2 \quad .$$

The difference in risks is then given by $E_\theta \Delta_\theta(\delta)$. Note that $\delta(X)$ is minimax if and only if $E_\theta \Delta_\theta(\delta) \leq 0$ for all θ.

Stein-type estimators of a multivariate normal mean pull the maximum likelihood estimator, X, toward a particular point (which, without loss of generality, can be taken to be zero). This selected point should be interpreted as an experimenter's best (prior) guess at the true mean, and will locate the portion of the parameter space in which the greatest risk improvement will be attained. It should be expected then, that given X is close to the prior guess, there should be good risk improvement in that portion of the parameter space. However, this is not always the case, and this criterion separates Stein-type estimators into two distinct classes.

We begin with two definitions:

DEFINITION 2.1. An estimator $\delta(X)$ of θ is said to favor θ_0 if $E_{\theta_0} \Delta_{\theta_0}(\delta) \leq E_\theta \Delta_\theta(\delta)$ for all $\theta \neq \theta_0$.

DEFINITION 2.2. An estimator $\delta(X)$ of θ is said to be conditionally minimax at θ_0 if

$$E_{\theta_0}\left(\Delta_{\theta_0}(\delta) I_{\{|X-\theta_0| \leq t\}}(X)\right) \leq 0 \qquad \text{for all } t \geq 0 \quad . \tag{2.1}$$

Given that an estimator favors θ_0, condition (2.1) seems a rather mild requirement for an estimator to satisfy. Indeed, such a condition can be viewed as a minimal requirement. The condition is strong enough, however, to separate the ordinary James-Stein estimator from its positive-part version.

Before proceeding, there is a technical matter which must be made clear. A distinction is being made between estimators which favor zero and estimators which shrink toward zero, i.e., estimators of the form $h(X)X$, $0 \le h(X) \le 1$. An estimator which favors zero need not shrink toward zero for all X (as in the case of the ordinary James-Stein estimator). However, since in both cases zero represents the best prior guess at θ, it is reasonable to require all these estimators to perform well at $\theta = 0$. The following theorem characterizes a class of estimators which favor zero.

THEOREM 2.1. _Let_ $\delta(X) = [1 - r(|X|^2)/|X|^2]X$, _where_ r _satisfies_

$$\text{i) } r(u) \text{ is nondecreasing ,}$$

$$\text{ii) } r(u)/u \text{ is nonincreasing ,}$$

$$\text{iii) } 0 \le r(u) \le 2(p-2) \quad .$$

_If_ r(u) _is a concave function of_ u, _then_ $\delta(X)$ _favors zero, i.e.,_ $E_0\Delta_0(\delta) \le E_\theta\Delta_\theta(\delta)$ _for all_ θ. _In fact,_ $E_\theta\Delta_\theta(\delta)$ _is a nondecreasing function of_ $|\theta|$.

PROOF. Given in the Appendix.

Although the concavity condition on r seems rather strong, most familiar estimators satisfy the condition. These include not only the ordinary and positive-part James-Stein estimators, but also the proper Bayes minimax estimators of Strawderman (1971). Also, it is easy to show that if r satisfies conditions i) – iii), and in addition $r(u) \ge p-2$ and is convex, then the resulting estimator does not favor zero. A simple transformation will show that if $\delta(X)$ favors zero, then $\theta_0 + \delta(X - \theta_0)$ favors θ_0.

The conditional property with which we are concerned is that an estimator be conditionally minimax at any θ that it favors, i.e., at any θ for which the infimum of the unconditional risk is attained. We will mainly be concerned with estimators which pull toward zero, and the reader should interpret the phrase 'conditionally minimax' as meaning 'conditionally minimax at the point $\theta = 0$.' If an estimator does not favor 0, whether or not it is conditionally minimax (i.e., conditionally minimax at the point $\theta = 0$) may be irrelevant. However, it will arise (as in the proof of Theorem 3.1) that the conditional risk at $\theta = 0$, of an estimator which may not favor zero, is helpful in evaluating the conditional risk of other estimators.

We now present the main result of this section, that the requirement of conditional minimaxity separates the ordinary James-Stein estimator from its positive-part version.

THEOREM 2.2. Let $\delta^s(X) = \left(1 - \frac{a}{|X|^2}\right)X$, and $\delta^+(X) = \left(1 - \frac{a}{|X|^2}\right)^+ X$. $\delta^s(X)$ is not conditionally minimax, but $\delta^+(X)$ is conditionally minimax.

PROOF. Let χ_p^2 denote a chi-squared random variable with p degrees of freedom. It is straightforward to calculate

$$E_0\left(\Delta_0(\delta^s)I_{[0,t]}(|X|)\right) = aP(\chi_p^2 \le t^2)\left[\frac{a}{p-2}\,\frac{P(\chi_{p-2}^2 \le t^2)}{P(\chi_p^2 \le t^2)} - 2\right] . \tag{2.2}$$

As $t^2 \to 0$, the ratio of probabilities in (2.2) approaches ∞, establishing the existence of a t for which $E_0\left(\Delta_0(\delta^s)I_{[0,t]}(|X|)\right) > 0$.

Actually, one can solve (numerically) for the unique t_0 (depending only on a), for which (2.2) is positive if and only if $t > t_0$. However, it is also straightforward to establish that $t^2 \le a/2$ is sufficient to insure that (2.2) is positive.

For the positive-part estimator, we consider two cases. If $t^2 \le a$, it is immediate that

$$E_0\left(\Delta_0(\delta^+)I_{[0,t]}(|X|)\right) = E_0\left(-|X|^2 I_{[0,t]}(|X|)\right) \le 0 \quad . \tag{2.3}$$

If $t^2 > a$, it is readily established that

$$E_0\left[\Delta_0(\delta^+)I_{[0,t]}(|X|)\right] = \frac{a^2}{p-2} P(a \le \chi^2_{p-2} \le t^2) - 2aP(a \le \chi^2_p \le t^2) - pP(\chi^2_{p+2} \le a) \quad . \tag{2.4}$$

Differentiating with respect to t^2 shows that $E_\theta\left[\Delta_0(\delta^+)I_{[0,t]}(|X|)\right]$ is decreasing for $t^2 > a/2$, establishing the result. ||

Therefore, although the ordinary Stein estimator dominates X in (unconditional) risk if $0 \le a \le 2(p-2)$, it fails to dominate X conditionally. Moreover, this failure occurs at the experimenter's best prior guess at θ, demonstrating that the ordinary Stein estimator is not a reasonable alternative to X. The positive-part Stein estimator, however, dominates X both conditionally and unconditionally, and is a reasonable alternative.

The shortcoming of the ordinary James-Stein estimator is its 'uncontrolled overshoot' near $|X| = 0$. While this deficiency is quite obvious from merely examining the estimator, it has never been related to a decision-theoretic quantity. It is our contention that the class of conditionally minimax estimators defines a class of estimators which contain all the reasonable alternatives to the maximum likelihood estimator. Any conditionally minimax estimator should be acceptable to a Bayesian, since its conditional performance is good (indeed, it will be shown that all Bayes estimators are conditionally minimax). Furthermore, any minimax conditionally minimax estimator should be acceptable to a frequentist, for if one requires minimaxity, one might as well choose an estimator with good conditional performance.

It is clear that an estimator which is conditionally minimax should be preferred over one that is not, but the question of what types of estimators are conditionally minimax has not yet been addressed. The next section shows that the class of conditionally minimax estimators is large enough to include estimators which are optimal against more familiar criteria.

3. Conditional Minimaxity and Admissibility. An estimator which is not conditionally minimax is not only undesirable conditionally, but also can be dominated unconditionally, as the following theorem shows.

THEOREM 3.1. *Suppose* $\delta(X) = [1 - r(|X|^2)/|X|^2]X$ *where* $r(u)$ *is nondecreasing and* $r(u)/u$ *is nonincreasing. If there exists* $t > 0$ *such that* $E_0\Delta_0(\delta)I_{[0,t]}(|X|) > 0$, *then* $\delta(X)$ *is inadmissible against squared error loss.*

PROOF. The theorem will be established by exhibiting an estimator $\delta^1(X)$ which uniformly dominates $\delta(X)$ in risk. Direct calculation shows

$$
\begin{aligned}
E_\theta\Delta_\theta(\delta)I_{[0,t]}(|X|) &= E_\theta\left\{\left(|\theta - \delta(X)|^2 - |\theta - X|^2\right)I_{[0,t]}(|X|)\right\} \\
&= E_\theta\left\{\left(\frac{2r(|X|^2)}{|X|^2}(\theta'X - |X|^2) + \frac{r^2(|X|^2)}{|X|^2}\right)I_{[0,t]}(|X|)\right\} \\
&= 2E_\theta\left\{\frac{r(|X|^2)}{|X|}|\theta|\cos\beta I_{[0,t]}(|X|)\right\} \\
&\quad + E_\theta\left\{r(|X|^2)\left(\frac{r(|X|^2)}{|X|^2} - 2\right)I_{[0,t]}(|X|)\right\} ,
\end{aligned} \tag{3.1}
$$

where $\cos\beta = \theta'X/|\theta||X|$.

Now

$$
\begin{aligned}
&E_\theta\left\{\frac{r(|X|^2)}{|X|}\cos\beta I_{[0,t]}(|X|)\right\} \\
&= K\int_0^t\int_0^\pi \frac{r(s^2)}{s^{\frac{1}{2}}}s^{p-1}(\sin\gamma)^{p-2}\cos\gamma\, e^{-\frac{1}{2}(s^2-2s|\theta|\cos\gamma)}d\gamma ds ,
\end{aligned} \tag{3.2}
$$

where $K = e^{-\frac{1}{2}|\theta|^2}\left[\pi^{1/2}2^{p/2}\Gamma\left((p-1)/2\right)\right]^{-1}$. Integrating the right-hand side of (3.2) by parts shows

$$E_\theta\left\{\frac{r(|X|^2)}{|X|}\cos\beta I_{[0,t]}(|X|)\right\}$$

(3.3)
$$= K|\theta|\int_0^t\int_0^\pi \frac{r(s^2)}{s^{\frac{1}{2}}}s^{p-2}\,\frac{(\sin\gamma)^{p-1}}{p-1}e^{-\frac{1}{2}(s^2-2s|\theta|\cos\gamma)}d\gamma ds \geq 0 \quad ,$$

which, combined with (3.1) establishes

(3.4)
$$E_\theta\Delta_\theta(\delta)I_{[0,t]}(|X|) \geq E_\theta\left\{r(|X|^2)\left(\frac{r(|X|^2)}{|X|^2}-2\right)I_{[0,t]}(|X|)\right\}$$

for all $t \geq 0$. We next show that there exists a t^* for which the right-hand side of (3.4) is positive for all θ. Once this fact is established, the inadmissibility of $\delta(X)$ is immediate, for it can be beaten by

(3.5)
$$\delta^1(X) = \begin{cases} X & \text{if } |X| \leq t^* \\ \delta(X) & \text{if } |X| > t^* \end{cases} .$$

Thus, to complete the proof, it must be shown that there exists $t^* > 0$ for which

(3.6)
$$E_\theta\left\{r(|X|^2)\left(\frac{r(|X|^2)}{|X|^2}-2\right)I_{[0,t^*]}(|X|)\right\} > 0 \qquad \text{for all } \theta \quad .$$

Since $E_0\Delta_0(\delta)I_{[0,t]}(|X|) > 0$ for some t, it follows that

(3.7)
$$E_0\left(\frac{r(|X|^2)}{|X|^2}-2\right)I_{[0,t]}(|X|) > 0 \qquad \text{for some } t \quad ,$$

otherwise, we have

$$E_0\Delta_0(\delta)I_{[0,t]}(|X|) = E_0 r(|X|^2)\left(\frac{r(|X|^2)}{|X|^2}-2\right)I_{[0,t]}(|X|)$$

(3.8)
$$\leq E_0\left\{r(|X|^2)I_{[0,t]}(|X|)\right\}E_0\left\{\left(\frac{r(|X|^2)}{|X|^2}-2\right)I_{[0,t]}(|X|)\right\}/P_0(|X|<t)$$

$$\leq 0 \quad ,$$

a contradiction. The inequality in (3.8) follows from the fact that $r(u)$ is nondecreasing and $r(u)/u$ is nonincreasing. Now, using (3.7) and the fact that $r(u)/u$ is nonincreasing, we have that there exists $t^* > 0$ such that

$$\frac{r(|X|^2)}{|X|^2} > 2 \qquad \text{for } |X| \le t^* \ , \tag{3.9}$$

establishing (3.6) and completing the proof. ||

The ordinary Stein estimator is an example of an estimator which is minimax but not conditionally minimax. It is also possible for an estimator to be conditionally minimax but not minimax. For example, the estimator $\delta(X) = cX$, $0 < c < 1$, which is proper Bayes, favors zero and, from Theorem 3.1, is conditionally minimax at zero. However, $\delta(X) = cX$ is not minimax.

It is interesting to examine more closely the estimator $\delta^1(X)$, which dominates $\delta(X)$ if $\delta(X)$ is not conditionally minimax. Let $\delta^s(X) = (1 - a/|X|^2)X$. It is easy to show that, if $t^* \le (a/2)^{\frac{1}{2}}$, the estimator

$$\delta^{t^*} = \begin{cases} X & \text{if } |X| < t^* \\ \left(1 - \dfrac{a}{|X|^2}\right)X & \text{if } |X| > t^* \end{cases}$$

dominates $\delta^s(X)$ in risk and is conditionally minimax at zero. Figure 1 is a graph of $E_0\left[\Delta_0(\delta) I_{[0,t]}(|X|)\right]$, as a function of t, for δ^s, δ^+ and δ^{t^*} using $p = 3$, $a = p - 2$, $t^* = [(p-2)/2]^{\frac{1}{2}}$. As the picture illustrates, controlling the 'overshoot' in an estimator results in improved conditional performance. Controlling 'overshoot' also results in improved unconditional performance as seen in Figure 2, a graph of the unconditional risk of δ^s, δ^+ and δ^{t^*}. The risk of δ^{t^*} is always between that of δ^s and δ^+.

It has not been established analytically that δ^{t^*} favors zero, but Figure 2 (and more extensive numerical calculations) indicate that this is the case.

Although δ^{t^*} is a vast improvement over δ^s, it still 'overshoots', although by a finite amount. Whether or not δ^{t^*} itself is a good estimator is not clear, but it appears to be reasonable when evaluated conditionally. In any case, the results of Efron and Morris (1973) show that δ^+ uniformly dominates δ^{t^*} in risk. Since δ^+ is conditionally minimax, it is clearly preferable to δ^{t^*}.

Thus, the property of conditional minimaxity does not necessarily imply that an estimator is totally desirable, but it is clear that the absence of conditional minimaxity indicates that the estimator is undesirable. Moreover, as the proof of Theorem 3.1 shows, the absence of conditional minimaxity is a constructive criticism, in that a better estimator can immediately be obtained. In the next section it is shown how conditional minimaxity relates to another constructive criticism, the existence of relevant betting procedures.

4. Conditional Minimaxity and Relevant Betting Procedures. The theory of relevant betting procedures has been studied by many authors, with the first step in a formal theory being taken by Buehler (1959). Many others have contributed to this theory (e.g., Stein, 1961; Brown, 1967; Olshen, 1973; Bondar, 1977), with the main focus being on conditional properties of interval estimators. More recently, Robinson (1979a,b) has given a detailed development of conditional properties of point estimators, which is what we are concerned with here.

In evaluating point estimators, Robinson gives details only for the case of one dimension, but indicates the obvious extentions to the multivariate problem. For completeness, we restate some of his definitions and theorems as they apply to the multivariate case.

DEFINITION 4.1. A function $s(X)$ is a betting procedure if it is bounded as a function of X. If $s(X)$ is unbounded but $E|s(X)|$ is bounded, $s(X)$ is called a wide-sense betting procedure.

If $s(X)$ is bounded, then without loss of generality the bound can be taken to be 1. The quantity $|s(x)|$ can then be interpreted as the probability that a bet of unit size is made given that $X=x$ is observed, hence the name 'betting procedure'.

DEFINITION 4.2. If $\delta(X)$ is a point estimator for the location vector θ, the betting procedure $s(X)$ is semirelevant if

$$E_\theta\{[\theta - \delta(X)]'s(X)\} \geq 0 \qquad \text{for all } \theta \quad , \tag{4.1}$$

with strict inequality for some θ, is relevant if, for some $\epsilon > 0$,

$$E_\theta\{[\theta - \delta(X)]'s(X)\} \geq \epsilon E\{|s(X)|\} \qquad \text{for all } \theta \tag{4.2}$$

with strict inequality for some θ, and is super-relevant if, for some $\epsilon > 0$,

$$E_\theta\{[\theta - \delta(X)]' s(X)\} \geq \epsilon \qquad \text{for all } \theta \quad . \tag{4.3}$$

The betting procedure $s(X)$ defines a direction in the Euclidean space along which one can expect positive gain by betting that the inner product $[\theta - \delta(X)]' s(X)$ is greater than zero. As Robinson notes, the existence of a super-relevant betting procedure is a severe criticism of an estimator, while the existence of a semi-relevant betting procedure is a mild criticism. However, the existence of a relevant betting procedure is about at the level to cause concern, and is, in fact, a constructive criticism in the following sense: If $s(X)$ is a relevant betting procedure for $\delta(X)$, then the estimator $\delta^*(X) = \delta(X) + \epsilon s(X)$ has uniformly smaller risk. The following theorem is from Robinson (1979a):

THEOREM 4.1 (Robinson). *For a point estimator $\delta(X)$, of a location vector θ, the absence of wide-sense semi-relevant betting procedures implies admissibility with respect to squared error loss, and admissibility with respect to squared error loss implies the absence of relevant betting procedures.*

An admissible estimator does not allow the existence of relevant betting procedures, but an estimator which does not allow the existence of relevant betting procedures is not necessarily admissible. However, estimators which do not allow relevant betting procedures are quite desirable from a conditional viewpoint, and the fact that such an estimator may not be admissible is of lesser concern.

The property of conditional minimaxity is slightly weaker than the absence of relevant betting procedures, but strong enough to eliminate the more undesirable estimators.

THEOREM 4.1. *Let* $\delta(X) = [1 - r(|X|^2)/|X|^2]X$ *where* r *satisfies the conditions of Theorem 3.1. If there exists a* $t^* > 0$ *such that* $E_0\Delta_0(\delta)I_{[0,t^*]}(|X|) > 0$, *then the betting procedure*

$$s(X) = \frac{r(|X|^2)X}{|X|^2} I_{[0,t^*]}(|X|) \tag{4.4}$$

is a wide-sense relevant betting procedure. $s(X)$ *is a relevant betting procedure if* $r^2(|X|^2)/|X|^2$ *is bounded.*

PROOF. From Theorem 3.1, $\delta(X)$ is dominated in risk by $\delta^1(X)$, given in (3.5). $\delta^1(X)$ can be written as

$$\delta^1(X) = \delta(X) + s(X) \quad ,$$

where $s(X)$ is defined by (4.4). We then have

$$\begin{aligned} 0 &\le E_\theta|\theta - \delta(X)|^2 - E_\theta|\theta - \delta^1(X)|^2 \\ &= E_\theta\{2[\theta - \delta(X)]'[\delta^1(X) - \delta(X)] - |\delta^1(X) - \delta(X)|^2\} \\ &= 2E_\theta\{[\theta - \delta(X)]'s(X) - \tfrac{1}{2}|s(X)|^2\} \quad . \end{aligned} \tag{4.5}$$

Recall from Theorem 3.1 that $r(|X|^2)/|X| > 2|X|$ for $|X| \le t^*$. Hence

$$\begin{aligned} \tfrac{1}{2}E_\theta|s(X)|^2 \ge \tfrac{1}{2}[E_\theta|s(X)|]^2 &= \tfrac{1}{2}\left[E_\theta \frac{r(|X|^2)}{|X|} I_{[0,t^*]}(|X|)\right]^2 \\ &\ge \left[E_\theta|X|I_{[0,t^*]}(|X|)\right] E_\theta|s(X)| \\ &\ge \left[E_0|X|I_{[0,t^*]}(|X|)\right] E_\theta|s(X)| \\ &> \epsilon E_\theta|s(X)| \end{aligned} \tag{4.6}$$

for $\epsilon < E_0|X|I_{[0,t^*]}(|X|)$. Thus, combining (4.5) and (4.6) yields

$$\begin{aligned} 0 &\le 2E_\theta\{[\theta - \delta(X)]'s(X) - \tfrac{1}{2}|s(X)|^2\} \\ &\le 2E_\theta\{[\theta - \delta(X)]'s(X) - \epsilon|s(X)|\} \qquad \text{for all } \theta \quad , \end{aligned} \tag{4.7}$$

and hence $s(X)$ is relevant for $\delta(X)$. If $r^2(|X|^2)/|X|^2$ is unbounded, then $s(X)$ is a wide-sense betting procedure, and is a betting procedure if $r^2(|X|^2)/|X|^2$ is bounded. ||

In general, given an estimator $\delta(X)$, it is a difficult task to verify the existence of a relevant betting procedure. It is quite simple, however, to verify conditional minimaxity (since one need only work with the central distribution). Thus, the property of conditional minimaxity once again serves to eliminate estimators which have undesirable conditional properties.

If $\delta(X) = [1 - r(|X|^2)/|X|^2]X$, where r satisfies the conditions of Theorem 3.1, then violation of conditional minimaxity implies the existence of a wide-sense relevant betting procedure. Although it has not been proven, it is conjectured that these conditions are equivalent (for this class of estimators). That is, an estimator is conditionally minimax if and only if there does not exist a wide-sense relevant betting procedure for it. Robinson's theorem shows that there exist conditionally minimax estimators which allow the existence of wide-sense semi-relevant betting procedures (otherwise all conditionally minimax estimators would be admissible), and the following theorem demonstrates that the property of minimaxity alone is sufficient to insure the absence of super-relevant betting procedures.

THEOREM 4.2. _The condition of minimaxity implies the absence of super-relevant betting procedures._

PROOF. Suppose $\delta(X)$ is a minimax estimator of θ, and $s(X)$ is a super-relevant betting procedure for $\delta(X)$. (Assume, without loss of generality, that $|s(X)| \leq 1$ for all X.) Let $\epsilon > 0$ satisfy

(4.8) $$E_\theta[\theta - \delta(X)]'s(X) \geq \epsilon \qquad \text{for all } \theta \quad .$$

We then have

$$
\begin{aligned}
E_\theta|\theta - [\delta(X) + \epsilon s(X)]|^2 &= E_\theta\{|\theta - \delta(X)|^2 - 2\epsilon(\theta - \delta)'s(X) + \epsilon^2|s(X)|^2\} \\
&\le E_\theta|\theta - \delta(X)|^2 - 2\epsilon^2 + \epsilon^2 \qquad (4.9) \\
&\le p - \epsilon^2 \qquad \text{for all } \theta \quad ,
\end{aligned}
$$

showing that the estimator $\delta(X) + \epsilon s(X)$ has risk strictly less than the minimax risk, a contradiction. ||

Therefore, the class of estimators which are both minimax and conditionally minimax is in between the class of estimators which does not allow super-relevant betting procedures, and the class that does not allow wide-sense relevant betting procedures.

5. Comments. The impact of the Stein effect has been far-reaching, and in this paper one of its (perhaps) more subtle implications is explored. Stein-type estimators allow the incorporation of prior information at no cost to the experimenter (since they are minimax), but the fact that the prior information is being used has never been incorporated into a decision-theoretic evaluation of the estimator. As we have seen, such an evaluation is necessary to insure reasonable conditional performance.

The criterion of conditional minimaxity was developed specifically to deal with estimators (such as Stein-type estimators) which give preference, in terms of unconditional risk, to certain regions of the parameter space. It is, in a sense, the weakest possible conditional requirement that one could put on such estimators, yet it is strong enough to eliminate some undesirable estimators. For evaluating procedures which use the Stein effect, and hence favor a region of the parameter space, a criterion such as conditional minimaxity seems more useful than the more general criterion of existence of relevant betting procedures. At the very least, we can say that it is much easier to check whether an estimator is conditionally minimax than it is to check whether any relevant betting procedures exist.

One might argue that the whole idea of prior information has no relevance in any classical decision-theoretic evaluation, but such an argument is now antiquated. The Stein effect is one of the few instances where one can truly get something for nothing; prior information can be used at no cost, and one would be just plain foolish not to do so. Once this idea has been accepted, conditional minimaxity becomes a desirable property for an estimator to have. Indeed, if one is to take the prior information seriously, conditional minimaxity is a requirement, and one might as well restrict attention to the class of minimax conditionally minimax estimators.

In this paper we have dealt only with estimators which favor a known point θ_0, corresponding to a prior point guess at θ . In practical applications, however, it is usually more advantageous to use an estimator which is adaptively centered, say at the grand mean. In terms of prior input, this corresponds to specifying a linear subspace (of the parameter space) in which θ is expected to lie. Such information can result in estimators of the form

$$\delta(X) = AX + \left(1 - \frac{r(|X-AX|^2)}{|X-AX|^2}\right)(X - AX) \quad , \tag{5.1}$$

where A is a known $p \times p$ matrix. With only minor modifications, the results of this paper can be made to apply to such estimators. For example, if the prior information specifies $H\theta = 0$ for some matrix H that is $r \times p$ of rank r, A would be chosen to satisfy $A = I - H'(HH')^{-1}H$, making AX the maximum likelihood estimator of θ under the restriction $H\theta = 0$. The resulting estimator should then be required to be conditionally minimax for all θ satisfying $H\theta = 0$ or, equivalently, $|\theta - A\theta| = 0$. An estimator of the form (5.1) has risk which depends on θ only through $|\theta - A\theta|$, so conditional evaluation at $|\theta - A\theta| = 0$ reduces the case considered in this paper.

APPENDIX: PROOF OF THEOREM 2.1

The proof of Theorem 2.1 follows quickly with the help of the following lemma, which may be of independent interest. In the following, let $\chi^2_{p,\lambda}$ denote a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter λ.

LEMMA 1. *Let* $h:[0,\infty)\to(-\infty,\infty)$ *be differentiable. Then, provided both sides exist,*

$$\frac{\partial}{\partial\lambda} E\left\{h\left(\chi^2_{p,\lambda}\right)\right\} = E\left\{\frac{\partial}{\partial\chi^2_{p+2,\lambda}} h\left(\chi^2_{p+2,\lambda}\right)\right\} . \tag{1}$$

PROOF. The lemma is established by equating the results of the well-known integration-by-parts technique with the results of some lesser-known identities for expectations of noncentral chi-squared random variables. We will proceed by evaluating the risk of the estimator $\delta(X)=\{1-[h(|X|^2)/|X|^2]\}X$, where $X\sim N_p(\theta,I)$. The usual integration-by-parts yields

$$E_\theta|\theta-\delta(X)|^2 = p-4E_\theta h'(|X|^2)+E_\theta\left\{\frac{h(|X|^2)}{|X|^2}[h(|X|^2)-2(p-2)]\right\} . \tag{2}$$

We can also write

$$\begin{aligned} E_\theta|\theta-\delta(X)|^2 &= E_\theta|[\theta-X]+[X-\delta(X)]|^2 \\ &= p+2E_\theta\{(\theta-X)'Xh(|X|^2)/|X|^2\}+E_\theta\{h^2(|X|^2)/|X|^2\} . \end{aligned} \tag{3}$$

We now employ the following identities, which can be found either in Bock (1975) or Casella (1980). If $h:[0,\infty)\to(-\infty,\infty)$, then provided the expectations exist,

$$(4)\qquad E_\theta\{Xh(|X|^2)\} = \theta E\left\{h\left(\chi^2_{p+2,|\theta|^2}\right)\right\} \quad ,$$

$$(5)\quad |\theta|^2 E\left\{h\left(\chi^2_{p+2,|\theta|^2}\right)/\chi^2_{p+2,|\theta|^2}\right\} = E\left\{h\left(\chi^2_{p-2,|\theta|^2}\right)\right\} - (p-2)E\left\{h\left(\chi^2_{p,|\theta|^2}\right)/\chi^2_{p,|\theta|^2}\right\} \quad ,$$

$$(6)\qquad \frac{\partial}{\partial|\theta|^2} E\left\{h\left(\chi^2_{p,|\theta|^2}\right)\right\} = \tfrac{1}{2}\left\{Eh\left(\chi^2_{p+2,|\theta|^2}\right) - Eh\left(\chi^2_{p,|\theta|^2}\right)\right\} \quad .$$

Now, using (4) and (5) on the first expectation in (3), and rearranging terms, we obtain

$$(7)\qquad \begin{aligned} E_\theta|\theta - \delta(X)|^2 &= p - 2\left\{Eh\left(\chi^2_{p,|\theta|^2}\right) - Eh\left(\chi^2_{p-2,|\theta|^2}\right)\right\} \\ &\quad + E_\theta\left\{\frac{h(|X|^2)}{|X|^2}[h(|X|^2) - 2(p-2)]\right\} \quad . \end{aligned}$$

We note in passing that, using the fact that the noncentral chi-squared distribution has monotone likelihood ratio in its degrees of freedom, equation (7) provides an immediate proof of the minimaxity of $\delta(X)$ provided h is nondecreasing and $0 \le h \le 2(p-2)$.

Now, equating (2) and (7), cancelling common terms, and using (6) establishes

$$(8)\qquad \frac{\partial}{\partial|\theta|^2} Eh\left(\chi^2_{p-2,|\theta|^2}\right) = E_\theta h'(|X|^2) = E\frac{\partial}{\partial\chi^2_{p,|\theta|^2}} h\left(\chi^2_{p,|\theta|^2}\right) \quad ,$$

proving the lemma. ||

THEOREM 2.1. *Let* $X \sim N_p(\theta, I)$, $\delta(X) = \{1 - [r(|X|^2)/|X|^2]\}X$, *where* r satisfies

i) $r(u)$ *is nondecreasing in* u ,

ii) $r(u)/u$ *is nonincreasing in* u ,

iii) $0 \le r(u) \le 2(p-2)$ *for all* u ,

iv) $r(u)$ *is concave in* u .

Then $(\partial/\partial|\theta|^2)E_\theta|\theta - \delta(X)|^2 \ge 0$ *for all* $|\theta|$.

PROOF. Assume, for the moment, that r is twice differtiable. Then, using (2), we have

$$E_\theta|\theta - \delta(X)|^2 = p - 4E_\theta r'(|X|^2) + E_\theta\left\{\frac{r(|X|^2)}{|X|^2}[r(|X|^2) - 2(p-2)]\right\} \quad . \tag{9}$$

Conditions i) – iii) on r insure that the function inside the second expectation in (9) is nondecreasing in $|X|^2$, hence the expectation is nondecreasing in $|\theta|^2$. Thus, we only need establish that $E_\theta r'(|X|^2)$ is nonincreasing in $|\theta|^2$. Using Lemma 1, we have

$$\frac{\partial}{\partial|\theta|^2} E_\theta r'(|X|^2) = \frac{\partial}{\partial|\theta|^2} Er'\left(\chi^2_{p,|\theta|^2}\right) = Er''\left(\chi^2_{p+2,|\theta|^2}\right) \leq 0 \tag{10}$$

by the fact that r is concave. Thus, the theorem is established if r is twice-differentiable. If r is not twice-differentiable, we can take a sequence $\{r_n\}$ of twice-differentiable concave functions which uniformly approach r. By carrying out the above calculations, and passing to the limit, the theorem is readily established. ||

REFERENCES

BERGER, J. O. and WOLPERT, R. L. (1982). The likelihood principle: a review and generalizations. Technical Report No. 82-33, Department of Statistics, Purdue University.

BOCK, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 3, 209-218.

BONDAR, J. V. (1977). A conditional confidence principle. Ann. Statist. 5, 881-891.

BROWN, L. D. (1967). The conditional level of Student's test. Ann. Math. Statist. 38, 1068-1071.

BUEHLER, R. J. (1959). Some validity criteria for statistical inference. Ann. Math. Statist. 30, 845-863.

CASELLA, G. (1980). Minimax ridge regression estimation. Ann. Statist. 8, 1036-1056.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors – an empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117-130.

FISHER, R. A. (1956). Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 361-379.

OLSHEN, R. A. (1973). The conditional level of the F-test. J. Amer. Statist. Assoc. 68, 692-698.

ROBINSON, G. K. (1979a). Conditional properties of statistical procedures. Ann. Statist. 7, 742-755.

ROBINSON, G. K. (1979b). Conditional properties of statistical procedures for location and scale parameters. Ann. Statist. 7, 756-771.

STEIN, C. (1955). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. **1**, 197-206.

STEIN, C. (1961). Approximation of prior measures by probability measures. Inst. Math. Statist., Wald Lectures. Unpublished manuscript.

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Math. Statist. **42**, 385-388.

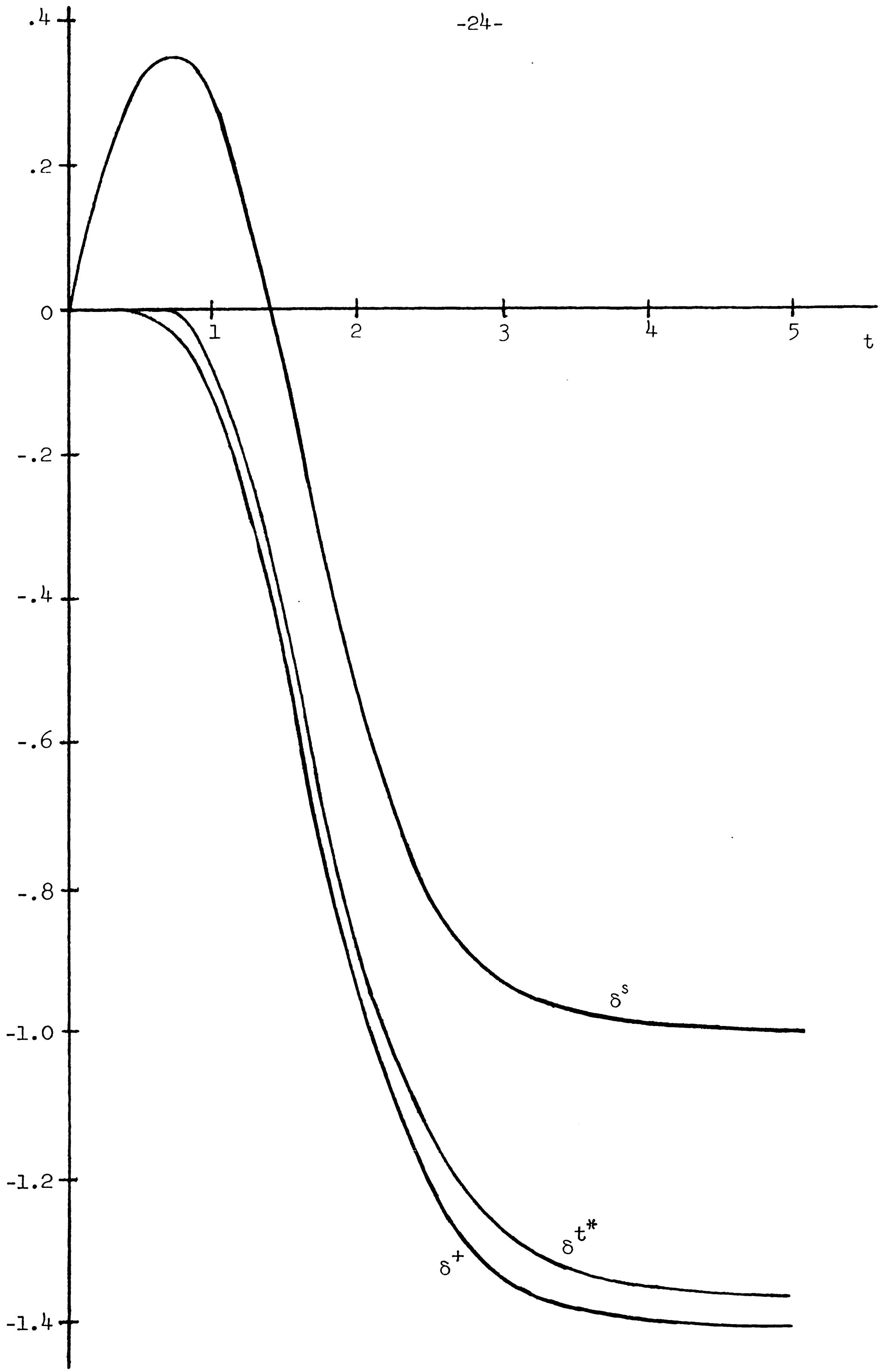

Fig. 1: Conditional difference in risks $E_0\Delta_0(\delta)I_{[0,t]}(|X|)$ for δ^s, δ^{t^*}, and δ^+ with $a = p-2$, $t^* = [(p-2)/2]^{\frac{1}{2}}$, $p = 3$

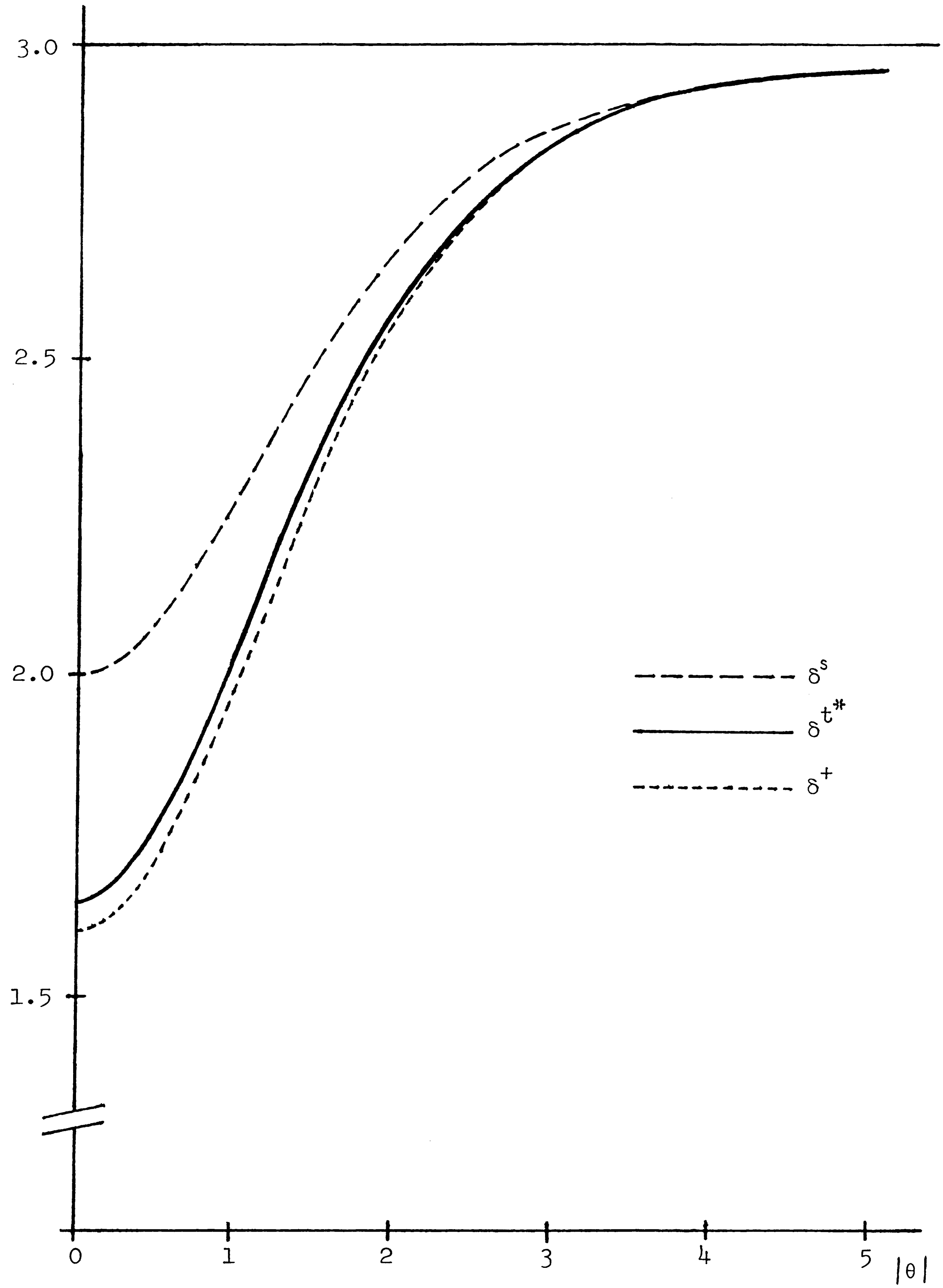

Fig. 2: Unconditional risks of δ^s, δ^{t^*}, δ^+ with

$a = p-2,\ t^* = [(p-2)/2]^{\frac{1}{2}},\ p = 3$