Report Title:
UNIFORM IMPROVEMENTS ON THE CERTAINTY EQUIVALENT RULE
IN A STATISTICAL CONTROL PROBLEM

by
L. Mark Berliner

Project Title:
PROBLEMS OF STATISTICAL INFERENCE WITH SPECIAL APPLICATIONS

Jagdish S. Rustagi, P.I.

DEPARTMENT OF THE NAVY
Office of Naval Research
Arlington, Virginia 22217

Contract No. N00014-78-C-0543

December, 1981
**Uniform Improvements on the Certainty Technique: Equivalent Rule in a Statistical Control Problem**

**Authors:** L. Mark Berliner

**Performing Organization:**
Dept. of Statistics, Ohio State University, 1358 Neil Ave., Columbus, Ohio 43210

**Controlling Office:** Office of Naval Research, Department of Navy, Arlington, Virginia 22207

**Report Date:** December 1981

**Number of Pages:** 12

**Distribution Statement:** Approved for public release; distribution unlimited.

**Supplementary Notes:**

**Key Words:** Control problem, differential inequalities, inadmissibility, multivariate normal distribution.

**Abstract:**
The problem considered arises when it is desired to choose values for design variables in a linear model so that the resulting dependent random variable will be close to a prescribed constant. A decision theoretic analysis is considered. Control procedures which dominate in risk, the commonly used certainty equivalent rule are presented.
UNIFORM IMPROVEMENTS ON THE CERTAINTY EQUIVALENT RULE IN A STATISTICAL CONTROL PROBLEM

by

L. MARK BERLINER

1Research supported in part by the National Science Foundation under Grant MEC-7802300 and by the Office of Naval Research under Contract N00014-78-C-0543(NR042-403).

Technical Report No. 237
Ohio State University
Department of Statistics
Columbus, Ohio 43210
December, 1981
UNIFORM IMPROVEMENTS ON THE CERTAINTY EQUIVALENT RULE IN A STATISTICAL CONTROL PROBLEM

L. Mark Berliner
Department of Statistics
Ohio State University
Columbus, Ohio, U.S.A.

1. The Control Problem. The problem considered here arises when it is desired to choose values for design variables in a linear model so that the resulting dependent random variable will be close to a prescribed constant. For discussions and applications of this problem, see Aoki [1], Dunsmore [6], Lindley [7], and Zellner [12].

The analysis here concerns the following transformed version of the problem as given in Basu [2] and Zaman [11]. Consider the linear model

\[ y = z^T \beta + \varepsilon, \]

where \( \beta \) is an unknown \( p \)-vector, \( z \) is an arbitrary \( p \)-vector, and \( \varepsilon \sim N(0, \sigma^2); \)

\( \sigma^2 \), possibly unknown. Furthermore, an estimate (independent of \( \varepsilon \)) of \( \beta \), say \( \hat{\beta} \), is available. Assume that \( \hat{\beta} \sim N_p(\beta, \Lambda) \), where \( \Lambda \) is a known, positive definite matrix. The goal is to choose a controller \( z_c(\hat{\beta}) \) which performs well with respect to a control risk function \( R_c \) given by

\[
R_c(z_c, \beta) = E(y-y^*)^2 \\
= E[(z_c(\hat{\beta}))^T \beta + \varepsilon - y^*]^2,
\]

where \( y^* \) is a non-zero constant, and the expectation is taken over both \( \hat{\beta} \) and \( \varepsilon \). Computation in (1.1) yields

\[
R_c(z_c, \beta) = \sigma^2 + (y^*)^2 [z_c(\hat{\beta})^T \beta - 1]^2.
\]

Research supported in part by the National Science Foundation under Grant MSC-7807300, and by the Office of Naval Research under Contract N00014-78-C-0543(NR042-403).
Now, consider the following transformation:

\[
\begin{align*}
X &= A^{-1/2} \beta \\
\theta &= A^{-1/2} \beta \\
\delta &= (y^*)^{-1} A^{-1/2} z_c(\hat{\beta}).
\end{align*}
\]

Application of this transformation in (1.2) yields the equivalent problem of choosing a decision rule \(\delta(x) = (\delta_1(x), \ldots, \delta_p(x))^T\), based on an observation \(X \sim N_p(\theta, I)\), \(\theta\) unknown, subject to incurring a loss,

\[
L(\delta, \theta) = (\theta^T \delta - 1)^2.
\]

This version of the control problem will be considered below.

The approach here is decision theoretic; control rules are evaluated in terms of their risk (expected loss) functions, \(R(\delta, \theta) = E_\theta L(\delta(X), \theta)\).

Our interest is the proposition of rules which dominate, in risk, the rule \(\delta_m(x) = |x|^{-2} x\). \(|x|^2 = \sum_{i=1}^p x_i^2\). This rule is the certainty equivalent controller. For brevity, discussion of the "certainty equivalent principle" (also known as the "separation principle") is omitted. The reader is referred to Aoki [11], Basu [21], and Zellner [12] for such discussions.

However, in the spirit of this principle, note that the loss function implies that \(\delta\) should be an estimator, in some sense, of the quantity \(|\theta|^{-2} \theta\). Clearly, \(\delta_m\) is the maximum likelihood estimator of this quantity.

As will be seen, a theorem of Zaman [11] implies that \(\delta_m\) is inadmissible in all dimensions. (In fact, note that \(\delta_m\) has infinite risk when \(p\) is 1 or 2.)

2. Previous Results. Most previous results are concerned with spherically symmetric (s.s.) rules, i.e., rules of the form \(\delta(x) = \Psi(|x|) x\).

Many inadmissibility results are based on the asymptotic behavior of s.s. rules as \(|x| \to \infty\). In particular, suppose that for some constant \(c\),

2
(2.1) \[ \delta(x) \approx (|x|^2 + c)^{-1}x \]

for \(|x|\) sufficiently large. The essence of the available results is that if \(c > 5 - p\), then \(\delta\) is inadmissible. (See Zaman [10] and Berger, Berliner, and Zaman [4] for precise theorems.) Related work of Zaman [10] and Takeuchi [9] also concerns rules of the form (2.1). In particular, Zaman [10] showed that the value \(c = 5 - p\) is asymptotically optimal (i.e., for all \(|\theta|\) sufficiently large, \(\delta(x) \approx (|x|^2 + 5 - p)^{-1}x\) has the smallest risk of all rules of the form (2.1)). Berger, Berliner, and Zaman [5] studied the admissibility of s.s. generalized Bayes controllers. Of special interest in this paper, it was shown that, under suitable regularity conditions on generalized prior measures, the corresponding generalized Bayes rules, that are given by \(\delta(x) \approx (|x|^2 + 5 - p)^{-1}x\) for \(|x|\) large, are admissible.

Three other results are directly applied in the discussion below. These results are paraphrased here. The reader is referred to the indicated references for precise statements.

RESULT 1. (Zaman [11]). If \(\delta(x) = \psi(|x|)x\) is admissible, then

i) \(0 < \psi(r) < 1.\)

ii) \(\lim_{r \to 0} \psi(r) = 1.\)

iii) \(\psi(r)\) is a \(C^{(\infty)}\) function.

Note that Result 1; i), ii), imply that \(\delta_m\) is inadmissible.

In Result 2, and the rest of the paper, the following notation is used:

DEFINITION: For any function \(\psi\), let

\[ T_1(\psi(r)) = \min \{1; \psi(r)\}. \]

Result 1 can be viewed as a partial motivation of Result 2.
RESULT 2. (Zaman [11].) Let \( \delta_0(x) = \varphi_0(|x|)x \). If \( \varphi_0(|x|) > 1 > 0 \), then \( \delta \) given by \( \delta(x) = \int_1^\infty \frac{\varphi_0(|x|)}{x} \) dominates \( \delta_0 \).

The next domination result was given by Berliner [5]. The derivation of this result is based on an integration by parts technique for risk analysis first introduced by Stein [8]. See Berliner [5] for a discussion and recent references.

For any differentiable function \( \psi(r) \) let

\[
\psi'(r) = \frac{d\psi(r)}{dr}.
\]

Also, for all real \( \eta \) define the quantities \( \xi(\psi) \) and \( \zeta(\psi) \) by

\[
\xi(\psi) = \int_0^\infty \left[ \frac{d}{dr} \left[ r(\psi + 1) \psi^2(r) \exp(-\frac{1}{2}r^2) \right] \right] \exp(r \eta) dr
\]

and

\[
\zeta(\psi) = \left[ r(\psi + 1) \psi^2(r) \exp(-\frac{1}{2}r^2) \exp(r \eta) \right]_0^\infty.
\]

RESULT 3. (Berliner [5].) Let \( \delta(x) = \varphi(|x|)x \) and \( \delta_0(x) = \varphi_0(|x|)x \).

Suppose that both \( \varphi \) and \( \varphi_0 \) are continuous, piece-wise differentiable functions on \((0,\infty)\) such that, for all real \( \eta \), the following conditions hold: (i) \( \xi(\psi) < \infty \), (ii) \( \xi(\psi_0) < \infty \), (iii) \( |\zeta(\psi)| < \infty \), and (iv) \( |\zeta(\psi_0)| < \infty \). If

\[
(2.2) \quad \varphi(r)(2r\varphi(r) + (p+1-r^2)\varphi(r) + 2) \geq \varphi_0(r)(2r\varphi_0(r) + (p+1-r^2)\varphi_0(r) + 2)
\]

for all \( r > 0 \), and with strict inequality on a set of positive Lebesgue measure, then \( \delta \) dominates \( \delta_0 \).

3. Main Results. Theorems 1 and 2 below present classes of procedures which dominate \( \delta_m \). These theorems are direct applications of Result 3.
THEOREM 1. Assume $p > 6$. Let $\delta(x) = \psi(|x|)x$ where $\psi(r) = (r^2-g(r))^{-1}$.

Suppose that

i) $g(r)$ is continuous, piece-wise differentiable, and non-decreasing,

ii) $g(r)$ is not identically zero,

and iii) $0 < g(r) \leq \min(2(p-5); h(r^2,p))$, where $h(r^2,p)$ is given by

$$h(r^2,p) = \frac{r^2[r^2+3(p-3) - ((r^2+1-p)^2 + 32(p-5))^{1/2}]}{2(r^2+p-3)}.$$ 

Then $\delta$ dominates $\delta_m$.

Proof: First, note that $\delta_m$ clearly satisfies the requirements of Result 3 (when $p \geq 3$). It is also easy to verify that $\delta$ satisfies these requirements.

Step 1. To apply Result 3, let $\psi_0(r) = r^{-2}$ and $\psi(r) = (r^2-g(r))^{-1}$.

Computing (2.2) in this case yields

$$\frac{(r^2-g(r))^{-1}[-2r(2r-g'(r))(r^2-g(r))^{-2} + (p+1-r^2)(r^2-g(r))^{-1} + 2]}{r^2(2r-2r^2) + (p+1-r^2)r^{-2} + 2}. 

Simplification of (3.1) yields

$$\frac{(r^2-g(r))^{-3}[2r(g'(r)-2r) + [(p+1-r^2) + 2(r^2-g(r)))r^2-g(r))]}{r^2[(p-3)r^{-2} + 1]}. 

It is easy to check that $g(r) < h(r^2,p)$ implies that $g(r) < r^2$ for all $r^2 > 0$. Using this fact, an algebraic manipulation of (3.2) implies the equivalent inequality.
\[(3.3) \quad 2rg'(r) + r^{-4} g(r) \left[ (2(p-5)-g(r))r^4 ight. \\
\left. + [g(r)-3(p-3)]g(r)r^2 + (p-3)g^2(r) \right] \geq 0.\]

Since \( g'(r) \geq 0 \) it is sufficient to verify that
\[(3.4) \quad [2(p-5) - g(r)] r^4 + [g(r) - 3(p-3)]g(r)r^2 + (p-3)g^2(r) \geq 0.\]

Step 2: Let \( Q(g) \) denote the L.H.S. of (3.4). Note that \( Q(g) \) can be written as (suppressing the dependence of \( g \) on \( r \))
\[(3.5) \quad Q(g) = (r^2 + p-3)g^2 - r^2(r^2 + 3(p-3))g + 2(p-5)r^4.\]

Clearly \( Q(g) \) is an upward opening quadratic function of \( g \). Consider the smallest root, say \( \gamma \), of \( Q(g) \). The quadratic formula implies that \( \gamma = h(r^2,p) \). Next, inspection of \( h(r^2,p) \) implies that \( h(r^2,p) > 0 \) when \( r^2 > 0 \) (and, of course, \( p \geq 6 \)). Hence, \( Q(g) \geq 0 \) for all \( g \leq \gamma \).

Finally, the dominance assertion then follows directly. ||

The remainder of this section is devoted to the development of a readily applicable version of Theorem 1. The motivation of this discussion is twofold. First, the function \( h(r^2,p) \) in Condition iii) of Theorem 1 is rather cumbersome. Second, the specific \( g \) functions to be discussed in Section 4 are easily bounded by linear functions of \( r^2 \).

**THEOREM 2.** Assume \( p \geq 6 \). Let \( \delta \) be defined as in Theorem 1. Suppose that the corresponding \( g \) function satisfies Conditions i) and ii) of Theorem 1. If
\[0 \leq g(r) \leq (p-5)\min(1,p^2/y^*)\]
where \( y^* > 0 \) satisfies \( h(y^*,p) = p-5 \), then \( \delta \) dominates \( \delta_m \).
The proof is based on the following lemma.

**Lemma 1.** For \( p > 6 \) and \( y > 0 \), \( h(y,p) \) is an increasing, concave function of \( y \).

The proof of Lemma 1 is a straightforward differentiation argument and is omitted.

**Proof of Theorem 2:** By Theorem 1 it is sufficient to verify that 
\[
g(r) \leq \min\{(p-5); h(r^2,p)\}.
\]

First, by Lemma 1, \( h(r^2,p) \leq p-5 \) iff \( r^2 \leq y^* \). Now consider a graph of \( h(r^2,p) \) as a function of \( r^2 \). The formula for the line connecting the origin and the point at which \( h(r^2,p) = p-5 \) is \([(p-5)/y^*]r^2 \). Therefore, by concavity, \([(p-5)/y^*]r^2 \leq h(r^2,p) \) for all \( r^2 \leq y^* \).

We close this section with a few remarks. First, the limit as \( r^2 \to \infty \) of \( h(r^2,p) \) is \( 3(p-3)/2 \). Lemma 1 implies that if \( p > 11 \), Condition iii) of Theorem 1 simply reduces to \( 0 \leq g(r) \leq h(r^2,p) \). (Since \( 3(p-3)/2 < 2(p-5) \) when \( p > 11 \).)

Second, Theorem 2 can be generalized in the following sense. Theorem 2 is an application of Theorem 1 where the implicit upper bound on \( g \) is lowered from \( 2(p-5) \) to \( p-5 \). The same analysis could be performed for other upper bounds (moderated by the remark immediately above) if desired. However, Theorem 2 does include the important asymptotically optimal case.

Finally, the computation of \( y^* \) is required. For convenience, a partial list of the values of \( y^* \) (computed numerically) is given in Table 1.
Table 1. Selected Values of \( y^* \)

|\( \begin{array}{c|c|c|c}
| \( p \) & \( y^* \) & \( p \) & \( y^* \) \\
|\hline
| 6 & 7.505551 & 15 & 20.000000 \\
| 7 & 9.123106 & 16 & 21.280110 \\
| 8 & 10.582576 & 17 & 22.549834 \\
| 9 & 12.000000 & 18 & 23.810250 \\
| 10 & 13.385165 & 19 & 25.062258 \\
| 11 & 14.744563 & 20 & 26.306624 \\
| 12 & 16.082763 & 25 & 32.433981 \\
| 13 & 17.403124 & 30 & 38.440307 \\
| 14 & 18.708204 & 40 & 50.206556 \\
|\end{array}\) |

4. Proposed Control Rules. First, the following limitation in the application of Result 3 is noted. Substitution of \( \psi_0 = \psi_m \) in (2.2) forces the R.H.S. of the inequality to be (positive) unbounded as \( r \to 0 \) (see (3.2)). Meanwhile, the L.H.S. of (2.2) remains bounded as \( r \to 0 \) for any admissible (or nearly admissible, in the sense of Result 1; i), ii)) rule. Hence, the differential inequality (2.2) cannot yield admissible alternatives to \( \delta_m \). In the case of Theorems 1 and 2, this fact is reflected in the implicit restriction that the \( \psi \) functions satisfying the conditions of these theorems are bounded from below by \( K|x|^{-2} \), for some constant \( K \), as \( |x| \to 0 \). However, the combination of Result 2 and these theorems leads to reasonable alternatives to \( \delta_m \) for \( p \geq 6 \).

PROPOSITION 1. Assume \( p \geq 6 \). Let \( \delta(x) = \psi(|x|)x \). Suppose \( \delta \) satisfies the assumptions of Theorem 1 (or 2). Then \( \delta_T \) given by

\[
\delta_T(x) = T_1(\psi(|x|))x
\]

dominates \( \delta_m \).
**Proof:** Obvious.||

**REMARK:** Unfortunately, the rules $\delta_T$ are also inadmissible as they violate the smoothness requirement of Result 1.

The final point of this discussion is the suggestion of functions $g(|x|)$ for actual use in the application of Theorem 2 and Proposition 1. The functions described below arise naturally as parts of a certain class of generalized Bayes rules. The reader is referred to Berger [3] and Berliner [5] for discussions. Only the required facts are given here.

Let $n > 1/2$. For $v > 0$ define the function $r_n(v)$ by

$$r_n(v) = \frac{1}{n} \int_0^v \frac{\exp\left(- \frac{1}{2}v\lambda\right) d\lambda}{\int_0^v (n-1) \exp\left(- \frac{1}{2}v\lambda\right) d\lambda}.$$

The following facts concerning $r_n$ are needed here:

**LEMMA 2.** (Berger [3].) If $n > 1/2$, then

i) $0 < r_n(v) < 2n$.

ii) $r_n'(v) > 0$.

iii) $\lim_{v \to \infty} r_n(v) = 2n$.

iv) $\lim_{v \to \infty} r_n(v) = n/(n+1)$.

v) $r_n(v)/v \leq n/(n+1)$.

Now, let $v = a|x|^2$ for some constant $a > 0$. Next, for $p > 6$, define $r^*(v)$ by $r^*(v) = r_n(v)$ for $n = (p-5)/2$. Then, clearly, by Lemma 2, $r^*(a|x|^2) \to p-5$ as $|x|^2 \to \infty$. Hence, the rule $\delta(x) = (|x|^2 - r^*(a|x|^2))^{-1}x$ is asymptotically optimal.
Also, by Lemma 2, note that
\[ r^*(a|x|^2) \leq [(p-5)/(p-3)]a|x|^2. \]
Then, to apply Theorem 2, we require that
\[ a \leq (p-3)/y^*. \]

**PROPOSITION 2.** Assume \( p \geq 6 \). Let \( \delta^* \) be given by
\[ \delta^*(x) = T_1(1/2 - r^*(a|x|^2))^{-1}x. \]
Then for any constant \( a \) such that \( 0 < a \leq (p-3)/y^* \), \( \delta^* \) dominates \( \delta_m^* \).

**Proof:** The proof is a direct application of the above arguments, Lemma 2, and Proposition 1.

5. **Comments.** i) The rules \( \delta^* \) proposed above display desirable properties:
   a) They are relatively easy to compute.
   b) Their behavior for \( |x| \) large is similar to that of the generalized Bayes, admissible rules discussed in Berliner [5].
   c) They are asymptotically optimal.

However, they are not admissible since the smoothness requirement of Result 1 is violated.

ii) A common criterion for choosing among decision rules is minimaxity. We simply note here that \( \delta_m \) is minimax when \( p \geq 3 \). Hence, \( \delta^* \) is also minimax. See Berliner [5] for proofs and discussion.

iii) Another natural control procedure often considered is the uniform measure, generalized Bayes rule \( \delta_u(x) = (1+|x|^2)^{-1}x \). This rule is admissible for \( p \leq 4 \), but inadmissible for \( p \geq 5 \). Several authors have shown that \( \delta_u \) is dominated by \( \delta_m \) when \( p \geq 5 \). Hence, the rules \( \delta^* \)
proposed above also dominate $\delta_u$ when $p \geq 6$. For further discussion, references for the above results, and another class of rules which dominate $\delta_u$, see Berliner [5].

Acknowledgments. I wish to express my appreciation to Jim Berger for his help and comments. My thanks also go to Jayne Sage for programming the computations for Table 1.

REFERENCES


