AN APPLICATION OF THE CAUCHY-SCHWARZ INEQUALITY TO AN EXTREMAL PROBLEM

BY

M. L. EATON and I. OLKIN

University of Chicago and Stanford University

TECHNICAL REPORT NO. 39
AUGUST 5, 1969

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT 6681

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
AN APPLICATION OF THE CAUCHY-SCHWARZ INEQUALITY TO AN EXTREMAL PROBLEM

BY

M. L. EATON and I. OLKIN
University of Chicago and Stanford University

TECHNICAL REPORT NO. 39
AUGUST 5, 1969

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT 6681

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
1. **Introduction**

The efficiency of certain statistical procedures frequently requires the solution of an extremal problem, which usually may be solved by a calculus of variations argument. A simpler and more elegant argument may, in some instances, be provided by using an inequality for which the case of equality yields the solution. Although this idea has been used before, the present result permits a wide class of applications. A discrete version is discussed by Bush and Olkin (1959); the use of Jensen's inequality in a similar context is given by Gastwirth and Wolff (1968).

2. **An extremal problem**

Let $C[0,1]$ denote the space of continuous functions on the interval $[0,1]$, and introduce the bilinear functional $(\cdot,\cdot)$ on $C[0,1]$ by

$$
(\ell_1,\ell_2) = \int_0^1 \ell_1(x) \ell_2(x) \, dx.
$$

The extremal problem considered is as follows:

$$
\min \quad \{(\ell,\ell) + (\ell,g)^2\},
\quad (\ell,h)=1
$$

where $g$ and $h$ are two fixed elements of $C[0,1]$.

In order to solve problem (2), we consider the Hilbert space $(H,(\cdot,\cdot))$ of square integrable functions on $[0,1]$, with $(\cdot,\cdot)$
defined by (1). A simplification may be achieved by defining a linear
transformation $F_1$ on $H$ to $H$ by

$$F_1 f = f + g(f, g) = f + F g,$$

so that

$$F_1 = I + F g,$$

where $I$ denotes the identity operator. In terms of the transformation
$F_1$, we want

$$\text{Min}_{(f, h) = 1} (f, F_1 f).$$

Note that $(f_1, F_1 f_2) = (F_1 f_1, f_2)$, and $\|F_1\| = 1 + (g, g)$, where $\|F_1\|$
denotes the operator norm of $F_1$.

We now require a transformation $T_g$ such that $(I - T_g) = (I + F g)^{-1}$.
Motivated by the vector result

$$(I + u'v)^{-1} = I - \frac{u'v}{1 + uv'v},$$

where $u$ and $v$ are row vectors, define $T_g$ by

$$T_g(f) = g \frac{(f, g)}{1 + (g, g)}.$$

It is then straightforward to verify
\[(I-T_g)(I+F_g) = (I+F_g)(I-T_g) = I,\]

so that \(I-T_g = (I+F_g)^{-1}\). Since \(I-T_g\) is positive definite, \(I-T_g\) has a self adjoint positive definite square root, say \(B = (I-T_g)^{1/2}\).

Clearly, \(B\) has an inverse and \((B^{-1})^2 = I+F_g\), e.g., see Lorch (1962). The Cauchy-Schwarz inequality yields

\[(B^{-1} f_1, B^{-1} f_1) (B f_2, B f_2) \geq (f_1, f_2)^2,\]

which together with the symmetry of \(B\) becomes

\[(4) \quad ((B^{-1})^2 f_1, f_1) (B^2 f_2, f_2) = (f_1, f_1) ((I-T_g) f_2, f_2) \geq (f_1, f_2)^2.\]

Equality is achieved if and only if

\[f_1 = c(I-T_g)^{-1} f_2.\]

Recall the extremal problem (3): \(\min(F_1 f_1, f_1)\) subject to \((f_1, h) = 1\).

Setting \(f_2 = h\) in (4), we obtain

\[(F_1 f, f) \geq \frac{1}{((I-T_g)h, h)},\]

with equality if and only if \(f = c(I-T_g)h\). However, the condition
\((f,h) = 1\) implies that \(c = 1/((I-T_g)h,h)\). Thus the minimizing function is

\[
(f^*,h) = \frac{(I-T_g)h}{((I-T_g)h,h)} = \frac{h[1+(g,g)]-g(h,g)}{(h,h)+(g,g)(h,h)-(g,h)^2},
\]

and the minimum is

\[
(f,f^*,f^*) = \frac{1+(g,g)}{(h,h)+(g,g)(h,h)-(g,h)^2}.
\]

Remark. Recall that the initial extremal problem requires a minimizer over the space \(C[0,1]\). From the above argument, it is clear that we could have started with any set \(S \subseteq H\) and determined the solution of \((1)\) where \(f \in S\). If the minimizer \(f^*\) given by \((5)\) satisfies \(f^* \in S\), then the minimization problem is solved. For example, \(S\) could be the space of bounded measurable functions. Obviously, the space \([0,1]\) does not play an essential role in these arguments.

3. An extension

Consider a Hilbert space \((H,\langle \cdot, \cdot \rangle)\) and let \(F\) be a positive definite self-adjoint linear transformation on \(H\) to \(H\) such that \(F\) has a bounded inverse. Let \(z_1, \ldots, z_k\) be \(k\) linearly independent vectors in \(H\) and consider the problem: minimize \((x,Fx)\) subject to the side conditions \((x,z_i) = 1\) for \(i = 1, \ldots, k\).

Since \(F\) is positive definite, \(F\) has a square root (e.g., see Lorch (1962)) and arguing as before, we have
\[(x, F_x) (y, F^{-1}y) \geq (x, y)^2.\]

Let \( W = \{x| (x, z_i) = 1 \text{ for } i = 1 \ldots k \} \) and set \( Y = \{y| y = \sum_{i=1}^k \lambda_i z_i, \sum_{i=1}^k \lambda_i = 1 \} \). For any \( x \in W \) and \( y \in Y \), it follows that
\[
(x, Fx) \geq \frac{1}{(y, F^{-1}y)},
\]
with equality if and only if \( x = \pm F^{-1}y/(y, F^{-1}y) \). To solve the minimization problem, it is sufficient to produce a \( y \in Y \) such that \( F^{-1}y_0/(y_0, F^{-1}y_0) \in W \). Let \( A = (a_{ij}) \) be the \( k \times k \) real matrix with \( a_{ij} = (F^{-1}z_i, z_j) \) and note that \( A \) is positive definite and symmetric. Define \( \lambda^* = eA^{-1}/eA^{-1}e' \), and let \( y_0 = \sum_{i=1}^k \lambda_i z_i \). Then \( y_0 \in Y \) and it is easy to verify that
\[
(7) \quad x_0 = F^{-1}y_0/(y_0, F^{-1}y_0) \in W.
\]

Hence \( x_0 \) is the minimizing vector and the minimum is
\[
(8) \quad (x_0, Fx_0) = \frac{1}{(y_0, F^{-1}y_0)} = \frac{1}{\sum_{i=1}^k \lambda_i \lambda_i^*(z_i, F^{-1}z_i)}
\]
\[
= \frac{1}{\lambda^*A\lambda^*} = eA^{-1}e'.
\]
An application

In comparing the Wilcoxon test versus the Student's t-test, Hodges and Lehmann (1956) require the solution of the extremal problem

\[ \text{Min } \int f^2(x) \, dx, \]

subject to the conditions: \( f(x) \geq 0 \) for all \( x \),

\[ \int f(x) \, dx = \int (x+1) f(x) \, dx = \int x^2 f(x) \, dx = 1. \]

Although the calculus of variations argument is quite straightforward, we show how the solution is obtained from Section 3. Here \( F \) is the identity operator,

\[ z_1 = 1, \quad z_2 = (x+1), \quad z_3 = x^2, \]

and we let \( f(x) = 0 \) for \( |x| \geq a \). Subsequently, we obtain that value of \( a \) which minimizes (9). Then for \( a_{ij} = (z_i, z_j) \), \( A = (a_{ij}) \), we obtain

\[
A = 2a \begin{pmatrix}
1 & 1 & a^{2/3} \\
1 & 1+a^{2/3} & a^{2/3} \\
a^{2/3} & a^{2/3} & a^{1/5}
\end{pmatrix}, \quad A^{-1} = \frac{3}{8a^5} \begin{pmatrix}
4a^2+3a^4 & -4a^2 & -5a^2 \\
-4a^2 & 4a^2 & 0 \\
-5a^2 & 0 & 15
\end{pmatrix},
\]

\[
\lambda = \frac{eA^{-1}}{eA^{-1}e} = \frac{1}{t} (3a^4, 5a^2, 15-5a^2), \quad t = \frac{3a^4 - 10a^2 + 15.}
\]

6
Hence the minimizing function \( f(x) = c[\lambda_1 + \lambda_2 (1+x) + \lambda_3 x^2] \) becomes
\[ f(x) = c[(3a^4 - 5a^2) + (15-5a^2)x^2] \] for \( x^2 \leq a^2 \), and \( f(x) = 0 \), otherwise. The constant \( c \) is easily determined from the constraints to be \( c = 3/8a^5 \). The requirement \( f(x) \geq 0 \) for all \( x \) places a restriction on \( a \), namely \( 5/3 \leq a^2 \leq 5 \). Hence
\[
\int f^2(x) \, dx \geq ceA^{-1} e' = 3(3a^4 - 10a^2 + 15)/8a^5 = g(a).
\]
Since \( g \) is monotone decreasing in \([\sqrt{5/3}, \sqrt{5}]\), the minimum is attained at \( a = \sqrt{5} \), and \( f^*(x) = c_1 - c_2 x^2 \), for \( x^2 \leq 5 \), and \( f^*(x) = 0 \), otherwise, where \( c_1 = 3\sqrt{5}/20 \) and \( c_2 = 3/20\sqrt{5} \). To complete the argument, let \( f \) be any non-negative function satisfying (10). Since \( c_1 - c_2 x^2 < 0 \) for \( x^2 > 5 \),
\[
\int f(x) f^*(x) = \int_{\sqrt{5}}^{\infty} f(x) (c_1 - c_2 x^2) \, dx \geq \int_{-\infty}^{\infty} f(x) (c_1 - c_2 x^2) \, dx
\]
\[
= (c_1 - c_2) = \int (f^*)^2 \, dx.
\]
Consequently
\[
0 \leq \int (f-f^*)^2 = \int f^2 - 2 \int ff^* + \int (f^*)^2 \leq \int f^2 - \int (f^*)^2,
\]
so that \( f^* \) is the minimizer.
4. A problem of Chernoff and Savage

Let \( f \) be a strictly positive probability density and let

\[ F(x) = \int_{-\infty}^{x} f(y) \, dy. \]

When \( f \) is the normal density function, Chernoff and Savage (1958) considered the problem of minimizing

\[ I(G) = \int_{-\infty}^{\infty} [f(F^{-1}(G(x)))]^{-1} g^2(x) \, dx, \]

where \( F^{-1} \) is the inverse function of \( F \), and the minimum is taken over all distribution functions \( G \) with density function \( g \), mean 0, and variance 1. Rather than assume \( f \) is the normal density, we assume that \( f \) is twice continuously differentiable and

\[ f(x) f''(x) - (f'(x))^2 \leq 0, \]

that is, \( \log f(x) \) is a concave function. It is also assumed that

\[ \int_{-\infty}^{\infty} f(x) \left| \log f(x) \right| \, dx < +\infty. \]

Now, let \( \mathcal{B} \) be the set of distribution functions which possess a density \( g \) and satisfy

\[ \int_{-\infty}^{\infty} \left| \log f(x) \right| g(x) \, dx < +\infty, \]

and

\[ \int_{-\infty}^{\infty} (\log f(x)) g(x) \, dx = \int_{-\infty}^{\infty} (10g f(x)) f(x) \, dx. \]

The remainder of this section is devoted to showing that \( I(G) \geq I(F) = 1 \) for all \( G \in \mathcal{B} \) and, when \( \log f \) is strictly concave, then
\[ I(G) > I(F) = 1, \]

for all \( G \in \mathcal{G} \), \( G \neq F \). This result will include the result of Chernoff and Savage (1958).

First note that the Cauchy-Schwarz inequality implies

\[
1 = (\int g(x) \, dx)^2 \leq \int \frac{g^2(x)}{f^{-1}(G(x))} \, dx \cdot \int f^{-1}(G(x)) \, dx
\]

\[
= I(G) \int f^{-1}(G(x)) \, dx,
\]

so that \( I(G) \geq [\int f^{-1}(G(x)) \, dx]^{-1} \). Now

\[
\frac{d^2}{du^2} (f^{-1}(u))) = \frac{f(F^{-1}(u))f''(F^{-1}(u)) - [f'(F^{-1}(u))]^2}{f(F^{-1}(u))^2} \leq 0,
\]

so that \( f^{-1}(u)) \) is concave for \( u \in (0,1) \). For \( G \in \mathcal{G} \), let

\[
\psi(G) = \int f(F^{-1}(G(x))) \, dx,
\]

so that \( \psi(F) = 1 \). For \( \alpha \in [0,1] \), \( G \in \mathcal{G} \) and fixed real \( x \), define \( v(\alpha) \) by

\[
v(\alpha) = f(F^{-1}[1-\alpha] F(x) + \alpha G(x)).
\]

Since \( f(F^{-1}(u)) \) is concave, \( v \) is concave on \([0,1]\). Hence
\[(11) \quad \nu(\alpha) \leq \nu'(0) \alpha + \nu(0),\]

for \(\alpha \in [0,1]\). But \(\nu(0) = f(x)\) and

\[\nu'(0) = \frac{f'(x)}{f(x)} (G(x) - F(x)).\]

Integrating both sides of \((11)\) with respect to \(x\), we have

\[\psi((1-\alpha) F + \alpha G) \leq 1 + \alpha \int_{-\infty}^{\infty} \frac{f'(x)}{f(x)} (G(x) - F(x)) \, dx,\]

for \(\alpha \in [0,1]\). To show that \(I(G) \geq 1\), it suffices to show \(\psi(G) \leq 1\), for which it is sufficient to show that

\[\int_{-\infty}^{\infty} \frac{f'(x)}{f(x)} (G(x) - F(x)) \, dx = 0.\]

Since \(\log f(x)\) is concave, \(\log f(x)\) is eventually monotone (see Appendix) so that from Theorem 1,

\[\int_{-\infty}^{\infty} (\log f(x)) \, dG(x) = -\int_{-\infty}^{0} \frac{f'(x)}{f(x)} G(x) \, dx + \int_{0}^{\infty} \frac{f'(x)}{f(x)} (1-G(x)) \, dx\]

\[+ \log f(0),\]

and an analogous equation with \(G(x)\) replaced by \(F(x)\). Thus
\[ 0 = \int_{-\infty}^{0} \frac{f'(x)}{f(x)} (G(x)-F(x)) \, dx + \int_{0}^{\infty} \frac{f'(x)}{f(x)} (G(x)-F(x)) \, dx \]

\[ = \int_{-\infty}^{\infty} \frac{f'(x)}{f(x)} (G(x)-F(x)) \, dx, \]

and hence, \( I(G) \geq 1 \).

Now, suppose \( \log f \) is strictly concave and \( G \neq F \). Then for \( x \) in a non-empty open set \( \mathcal{O} \), \( G(x) \neq F(x) \); for \( x \in \mathcal{O} \) and \( \alpha \in (0,1] \),

\[ (12) \quad \nu(\alpha) < \nu'(0) \alpha + \nu(0). \]

Integrating both sides of (12), we have, as above, that \( \psi((1-\alpha) F + \alpha G) < 1 \) for \( \alpha \in (0,1] \) and \( F \neq G \). The conclusion then follows.

**Appendix**

**Definition:** A function \( g \) on \( R \) to \( R \) is **eventually monotone** if there exists an \( A > 0 \) such that \( g \) is monotone on \((A,\infty)\) and on \((\infty,-A)\).

Note that if \( g \) is eventually monotone, then \(|g|\) is eventually monotone.

**Theorem 1:** Let \( F \) be a continuous distribution function and suppose \( g \) is continuously differentiable and eventually monotone such that \( \int_{-\infty}^{\infty} |g(x)| dF(x) < +\infty \). Then for each real number \( b \), we have
\[ \int_{-\infty}^{\infty} g(x) \, dF(x) = -\int_{-\infty}^{b} F(x) \, g'(x) \, dx + \int_{b}^{\infty} (1-F(x)) g'(x) \, dx + g(b), \]

where the two integrals \( \int_{-\infty}^{b} F(x) \, g'(x) \, dx \) and \( \int_{b}^{\infty} (1-F(x)) g'(x) \, dx \) converge absolutely.

**Proof:** We first claim that

\[ \lim_{T \to \infty} |g(T)|(1-F(T)) = 0. \]

If \( |g| \) is eventually decreasing, the claim is obvious. If \( |g| \) is eventually increasing, then for \( T \) sufficiently large, we have

\[ 0 = \lim_{T \to \infty} \int_{T}^{\infty} |g(x)| \, dF(x) \geq \lim_{T \to \infty} |g(T)|(1-F(T)). \]

In exactly the same way, it may be shown that

\[ \lim_{T \to -\infty} |g(T)|F(T) = 0. \]

Now,

\[ \int_{-\infty}^{\infty} g(x) \, dF(x) = \int_{-\infty}^{b} g(x) \, dF(x) + \int_{b}^{\infty} g(x) \, dF(x). \]

However,

\[ \int_{-\infty}^{b} g(x) \, dF(x) = \lim_{T \to -\infty} \int_{T}^{b} g(x) \, dF(x) \]

\[ = \lim_{T \to -\infty} \left[ g(x) \, F(x) \right]_{x=T}^{x=b} - \int_{T}^{b} F(x) \, g'(x) \, dx \]

\[ = g(b) \, F(b) - \lim_{T \to -\infty} \int_{T}^{b} F(x) \, g'(x) \, dx, \]

12
since \( \lim_{T \to -\infty} g(T) F(T) = 0 \). Thus \( \lim_{T \to -\infty} \int_T^b F(x) g'(x) \, dx \) exists and is finite. That \( \int_{-\infty}^b F(x) g'(x) \, dx \) converges absolutely follows from the fact that \( g' \) is eventually positive or eventually negative. Thus,

\[
\int_{-\infty}^b g(x) \, dF(x) = g(b) F(b) - \int_{-\infty}^b F(x) g'(x) \, dx.
\]

In the same way, we have

\[
f(b) \, dF(x) = - \int_{-\infty}^\infty g(x) \, d(1-F(x))
\]

\[
= g(b)(1-F(b)) + \int_{b}^{\infty} (1-F(x)) g'(x) \, dx,
\]

and the conclusion follows. \( \| \)
References


