KHINČIN-KULLBACK-LEIBLER ESTIMATION WITH INEQUALITY CONSTRAINTS

by

A. Charnes
W.W. Cooper
J. Tyssedal*

November 1981

*University of Bergen, Norway

This research was partly supported by ONR Contract N00014-75-C-0569 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the United States Government.
ABSTRACT

In this paper we extend our grasp of statistical theory, duality theory, and computational convenience to the general linear inequality situation in $K^2L$ (or minimum discrimination information) estimation by exhibiting it as the limit of a simple one parameter sequence of equality problems. Only the finite discrete distribution case is treated here.

KEY WORDS

Khinchin-Kullback-Leibler Estimation
Inferential Distribution
Minimum Discrimination Information
Inequality Constrained Distributions
INTRODUCTION

Frequently in hypothesis testing or estimation of an "inferential" distribution (in Akaike's terminology) by the Khinmin-Kullback-Leibler (or Minimum Discrimination Information) method one has information about the possible candidate distributions in the form of linear inequalities on the components of the distribution in addition to equality (moment) constraints. In MOFS 1978, Charnes, Cooper, and Seiford [1] developed a convex programming duality theory for the $K^2L$ method with linear inequality constraints. It is especially incisive and convenient for constraints in equality (or moment) form which, further, have connections with established statistical theory as well as the analytic and computational facility of an unconstrained extremal problem in simple smooth functions.

In this paper we extend our grasp of statistical theory, duality theory, and computational convenience to the general linear inequality situation by exhibiting it as the limit of a simple one parameter sequence of equality problems. We treat only the finite discrete distribution case here, reserving the general distribution case to a forthcoming paper which also simplifies the duality theory of the "equality" case for general distributions as developed by Ben-Tal and Charnes in [2].

INEQUALITY AND APPROXIMATING EQUALITY FORMS

The dual programs for the inequality form as presented in Charnes-Cooper-Seiford [1] are
\[
\max \ v(\delta) = -\delta^T \ln \left[ \frac{\delta}{\delta c} \right]
\]

(I) s.t. \begin{align*}
\delta_T A^1 &= b^1_T \\
\delta_T A^2 + \gamma^T &= b^2_T
\end{align*}
\[\delta, \gamma \geq 0\]

and

(II) \[\inf \ z(z) = c \epsilon x^1 \epsilon + \epsilon^2 z^2 - b^1_T z^1 - b^2_T z^2\]

s.t. \[z^2 \geq 0\]

To obtain the dual programs for our approximating (weighted) equality form, we employ the procedure in [1]. Thus, consider

(1) \[K(x,y,\delta,\gamma) \triangleq \sum_{i \in I^1} (c_i \epsilon x^1 - \delta_i x^1) + \sum_{i \in I^2} (\epsilon c_i \epsilon (y^1_i/\epsilon) - y^1_i y^1_i)\]

where \(c_i > 0\) for all \(i\) and \(\epsilon > 0\).

Minimizing with respect to \(x^1\) and \(y^1\) gives:

(2) \[x^*_1 = \ln \left( \frac{\delta_i}{c_i} \right) , \ i \in I^1 , \ y^*_1 = \epsilon \ln \left( \frac{y^1_i}{\epsilon} \right) , \ i \in I^2\]

Hence,

(3) \[-\sum_{i \in I^1} \delta_i \ln \left( \frac{\delta_i}{c_i} \right) - \epsilon \sum_{i \in I^2} y^1_i \ln \left( \frac{y^1_i}{\epsilon c_i} \right) \leq K(x,y,\delta,\gamma)\]

To decouple, as in [1], we obtain

(4) \[\delta^T x + \gamma^T y = b^1_T z^1 + b^2_T z^2\]

and

(5) \[(\delta^T, \gamma^T) \begin{bmatrix} A^1 & A^2 \\ 0 & I \end{bmatrix} \begin{bmatrix} z^1 \\ z^2 \end{bmatrix} = (b^1_T, b^2_T) \begin{bmatrix} z^1 \\ z^2 \end{bmatrix}\]
when we choose

(6) \( x = A_1 z + A_2 z^2, y = z^2 \)

Thereby, we have the dual problems

\[
\max \ v(\delta, \gamma, \varepsilon) = -\sum_{i \in I_1} \delta_i \ln \left( \frac{\delta_i}{c_i} \right) - \varepsilon \sum_{i \in I_2} \gamma_i \ln \left( \frac{\gamma_i}{\varepsilon c_i} \right)
\]

(III) s.t.

\[
\begin{align*}
\delta^T A_1 &= b_1^T \\
\delta^T A_2 + \gamma^T &= b_2^T \\
\delta, \gamma &> 0
\end{align*}
\]

and

(IV) \( \inf \xi(z) = c^T e A_1 z + A_2 z^2 - b_1^T z^1 - b_2^T z^2 + \varepsilon c^T e (z^2/\varepsilon) \)

with \( z \) unconstrained

(Note that the \( c_i, i \in I_2 \), may be chosen arbitrarily.)

The duality theory of (III) and (IV) is precisely that of the equality case in [1], as may be seen by making the change of variables to

\[
\tilde{\gamma}_i := \varepsilon \gamma_i, \quad \tilde{c}_i := \begin{bmatrix} c_i, i \in I_1^1 \\ \varepsilon c_i, i \in I_2^1 \end{bmatrix}.
\]

Our present form is, however, more convenient for our arguments.

We now define

(7.1) \( f(z) = c^T e A_1 z + A_2 z^2 - b_1^T z^1 - b_2^T z^2, z \leq 0 \)

(7.2) \( g(z, \varepsilon) = c^T e A_1 z + A_2 z^2 - b_1^T z^1 - b_2^T z^2 + \varepsilon c^T e (z^2/\varepsilon) \)

We will then have
Theorem 1:

For some $c > 0$, $\inf_{z^2 > 0} f(z) \exists$ if and only if $\inf_{z} g(z,c) \exists$.

Further,

$$\lim_{c \to 0} \inf_{z^2 \leq 0} g(z,c) = \inf_{z^2 \leq 0} f(z)$$

Proof:

Consider $\frac{\partial g}{\partial z^2}$. It may be written as

$$\frac{\partial g}{\partial z^2} = (c^1 k) e A^1 z^1 + A^2 z^2 - b_1^1 + z^2 c_1^e \left( \frac{z_1^2}{\epsilon} \right), \quad \epsilon \in \mathbb{R}^2$$

where $k_1 \triangleq \frac{\partial}{\partial z^2} (r^2 z^2)$ and $(c^1 k)^T \triangleq (c_{i_1}^1 k_1, \ldots, c_{i_m}^1 k_m)$.

For $z^2 > 0$ we see that by choosing $\epsilon$ small enough we can make $g(z,c)$ an increasing function in that direction. Thus, in seeking a minimum we need only consider $z^2 \leq 0$. For notational simplicity in the following and "inf" with respect to $z$ shall always also entail $z^2 \leq 0$.

For $z^2 < 0$ and all $\epsilon > 0$ we have

$$f(z) \leq g(z,c) \leq f(z) + \epsilon \sum_{i \in I^2} c_i$$

If $\inf f(z) \exists$, then, by (9), $\inf f(z) \leq g(z,c)$ so that $\inf g(z,c) \exists$ for all $\epsilon > 0$. Further, $\inf f(z) \leq \inf g(z,c)$. On the other hand, if $\inf g(z,c) \exists$, then, by (9), $\inf g(z,c) \leq f(z) + \epsilon \sum_{i \in I^2} c_i$ so that $\inf f(z)$ exists. Further, too, $\inf g(z,c) \leq \inf f(z) + \epsilon \sum_{i \in I^2} c_i$. 
Hence

\[(10.1) \quad \inf f(z) \leq \inf g(z, \varepsilon) \leq \inf f(z) + \varepsilon \sum_{i \in I} c_i\]

and

\[(10.2) \quad \lim_{\varepsilon \to 0} \inf g(z, \varepsilon) = \inf f(z)\]

When \(f(z)\) has a minimum at \(z^*\), again from (9) we can conclude

\[(11) \quad f(z^*) = \lim_{\varepsilon \to 0} g(z^*, \varepsilon)\]

Q.E.D.

So much for the minimization side of the duality. For the \(K_L^2\) side we have

**Theorem 2:**

The maximum in problem (I) and, as \(\varepsilon \to 0\), in problem (III) is the same.

**Proof:**

For \(0 \leq \gamma_1 \leq e\gamma_1\), we have \(-c_i \leq \gamma_1 \ln \left(\frac{\gamma_1}{e\gamma_1}\right) \leq 0\), whereas for \(\gamma_1 > e\gamma_1\), \(\gamma_1 \ln \left(\frac{\gamma_1}{e\gamma_1}\right) > 0\). Therefore,

\[(12) \quad -\varepsilon \sum_{i \in I^2} \gamma_1 \ln \left(\frac{\gamma_1}{e\gamma_1}\right) \leq \varepsilon \sum_{i \in I^2} c_i \quad \text{for all} \quad \gamma_1 \geq 0.\]

Thus,

\[(13) \quad v(\delta, \gamma, \varepsilon) \leq v(\delta) + \varepsilon \sum_{i \in I^2} c_i \leq f(z) + \varepsilon \sum_{i \in I^2} c_i\]

From (13) we conclude
\[ (14) \quad \sup v(\delta, \gamma, \epsilon) \leq \sup v(\delta) + \epsilon \sum_{i \in I} c_i \leq \inf f(z) + \epsilon \sum_{i \in I} c_i \]

But by the theory of [1],
\[ \sup v(\delta, \gamma, \epsilon) = \max v(\delta, \gamma, \epsilon) = \inf g(z, \epsilon) \geq \inf f(z) \text{ by (10.1). Hence} \]
\[ (15) \quad \inf f(z) \leq \max v(\delta, \gamma, \epsilon) \leq \sup v(\delta) \leq \inf f(z) + \epsilon \sum_{i \in I} c_i \]

Thus letting \( \epsilon \to 0 \), we have
\[ (16) \quad \inf f(z) \leq \lim_{\epsilon \to 0} \max v(\delta, \gamma, \epsilon) \leq \sup v(\delta) \leq \inf f(z) \]

and
\[ (17) \quad \lim_{\epsilon \to 0} \max v(\delta, \gamma, \epsilon) = \sup v(\delta) = \inf f(z) \]

Q.E.D.

REFERENCES


Khinchin-Kullback-Leibler Estimation With Inequality Constraints

A. Charnes, W.W. Cooper, J. Tyssedal

Center for Cybernetic Studies, UT Austin
Austin, Texas 78712

Office of Naval Research (Code 434)
Washington, D.C.

November 1981

Unclassified

This document has been approved for public release and sale; its distribution is unlimited.

In this paper we extend our grasp of statistical theory, duality theory, and computational convenience to the general linear inequality situation in $K^2L$ (or minimum discrimination information) estimation by exhibiting it as the limit of a simple one parameter sequence of equality problems. Only the finite discrete distribution case is treated here.