THE ISOTONIC REGRESSION PROBLEM AND ITS DUAL

BY

R. E. BARLOW and H. D. BRUNK

TECHNICAL REPORT NO. 51
AUGUST 3, 1970

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP 17172
Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE ISOTONIC REGRESSION PROBLEM AND ITS DUAL

by

R. E. BARLOW*

University of California, Berkeley

and

H. D. BRUNK

Oregon State University

TECHNICAL REPORT NO. 51
AUGUST 3, 1970

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT 17172

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

*This research was completed while the author was at Stanford University, 1969-70, and was supported in part by the National Science Foundation under Grant GP-17172 at Stanford and the Office of Naval Research under Contract N00014-69-A-0200-1036 with the University of California.
THE ISOTONIC REGRESSION PROBLEM AND ITS DUAL
by
R. E. Barlow and H. D. Brunk

1. INTRODUCTION

The isotonic regression solves many restricted maximum likelihood estimation problems [cf. Ayer, Brunk, Ewing, Reid, Silverman (1955), van Eeden (1956), (1957), Brunk (1965)]. Perhaps, the prototype problem leading to the isotonic regression problem is the following [cf. Bartholomew (1959)]. Given $k$ normal distributions, $N(\mu_i, \sigma_i^2)$ $(i = 1, 2, \ldots, k)$, let $\bar{x}_i$ be the sample mean from the $i$th distribution and $n_i$ the sample size. We suppose $\sigma_i$ $(i = 1, 2, \ldots, k)$ are known and consider the problem of maximizing the likelihood subject to an order restriction on the means; i.e.,

$$\text{Maximize} \quad \frac{k}{\prod_{i=1}^{n_i} \sqrt{2\pi\sigma_i}} \exp \left[ -\frac{1}{2\pi\sigma_i} (\bar{x}_i - \mu_i)^2 / \sigma_i^2 \right]$$

Subject to $\mu_i \leq \mu_j$ \hspace{1cm} when $i \leq j$

where "$\leq$" is a partial ordering on $\Omega = \{1, 2, \ldots, k\}$. The partial ordering is provided by some a priori knowledge concerning the means of the distributions.

This is equivalent to the problem

$$\text{Minimize} \quad \frac{k}{\prod_{i=1}^{n_i} \sqrt{2\pi\sigma_i}} \exp \left[ -\frac{1}{2\pi\sigma_i} (\bar{x}_i - \mu_i)^2 / \sigma_i^2 \right]$$

Subject to $\mu_i \leq \mu_j$ \hspace{1cm} when $i \leq j$

(1.1)

where $w_i = \frac{n_i}{\sigma_i^2} > 0$. We call this the isotonic regression problem since we seek
to minimize \( (1.1) \) subject to \( \mu \) being an isotonic (i.e., order preserving) function on \( \Omega \). The solution to this problem is well known and is characterized in the next section. What makes the solution to the isotonic regression interesting is that it also solves many other seemingly unrelated problems.

We formulate a generalization of the isotonic regression problem and calculate its Fenchel dual in Section 3. The solution to the isotonic regression problem also provides a solution to the dual problem. We prove this independently of Fenchel's duality theorem using only well-known properties of the isotonic regression. An inventory problem and a problem of estimating stochastically ordered distributions are identified as versions of the dual problem in Sections 4 and 5.
2. PROPERTIES OF THE ISOTONIC REGRESSION

The isotonic regression problem is to

Minimize \[ \sum_{i=1}^{k} (g_i - x_i)^2 w_i \]

Subject to \( x_i \leq x_j \) when \( i \leq j \)

where \( g_i \) and \( w_i > 0 \) \( (i = 1, 2, \ldots, k) \) are given and \( \leq \) is a specified partial ordering on \( \Omega = \{1, 2, \ldots, k\} \). We call the solution to this problem the isotonic regression. A vector \( x = (x_1, x_2, \ldots, x_k) \) is said to be isotonic or order preserving if \( i \leq j \) implies \( x_i \leq x_j \). Note that the set of isotonic vectors \( x \in R^k \) is a convex cone, say \( K \). This property is all that is needed for the following theorem.

Theorem 2.1:

Let \( g \) and \( w \) be given functions on \( \Omega = \{1, 2, \ldots, k\} \), \( w > 0 \). Let \( K \) be a convex cone of functions on \( \Omega \) (not necessarily the cone of isotonic functions with respect to a partial ordering). Then a function \( u \) in \( K \) solves the problem

\[(2.1) \quad \text{Minimize} \quad \sum_{x \in K} \sum_{i=1}^{k} (g_i - x_i)^2 w_i \]

if and only if

\[(2.2) \quad \sum_{i=1}^{k} (g_i - u_i)u_i w_i = 0 \]

and

\[(2.3) \quad \sum_{i=1}^{k} (g_i - u_i)x_i w_i \leq 0 \quad \text{for all} \ x \in K \]
There is at most one such solution.

If \( u \) satisfies these conditions, then also

\[
(2.4) \quad \sum_{i=1}^{k} (g_i - x_i)^2 w_i \geq \sum_{i=1}^{k} (g_i - u_i)^2 w_i + \sum_{i=1}^{k} (u_i - x_i)^2 w_i
\]

for all \( x \in K \).

A proof in a more general context may be found in Brunk (1965). Conditions (2.2) and (2.3) are essentially the Kuhn-Tucker conditions for this special convex programming problem. We shall show that if \( u \) solves (2.2) then \( g - u \) solves the dual problem. Condition (2.3) says that \( g - u \) belongs to the dual cone, \( K^* \), i.e., \( g - u \) is feasible for the dual problem, while (2.2) says that the solutions are orthogonal. [See Rockafellar, *Convex Analysis*, (1970), Theorem 31.4 for the general duality theorem.]

We denote the solution to (2.1) by \( g^* \) to indicate that it is the isotonic regression on \( g \) with respect to the weights \( w \). If \( K \) is the cone of linear functions, for example, then \( g^* \) becomes ordinary linear regression.

The convex cone of isotonic functions is also a lattice. This is the additional property used to prove Theorem 2.2.

**Theorem 2.2:**

Let \( g \) and \( w \) be given functions on \( \Omega = \{1, 2, \ldots, k\} \), \( w > 0 \). Let \( K \) be a convex cone of isotonic functions on \( \Omega \) with respect to some partial ordering "\( \preceq \)".

(2.5) The isotonic regression \( g^* \) of \( g \) solving (2.1) exists and is unique.

(2.6) If \( a_i \leq g_i \leq b_i \) (\( i = 1, 2, \ldots, k \)), then \( a_i \leq g_i^* \leq b_i \).

(2.7) For an arbitrary real valued function \( \psi \) on the reals, \( \sum_{i=1}^{k} (g_i - g_i^*) \psi(g_i^*) w_i = 0 \).
A proof in a more general context may be found in Brunk (1965).

If $K$ is a cone of isotonic functions, we can characterize $g^*$ in terms of upper and lower sets determined by the partial ordering $\leq$.

**Definition:**

A subset $L$ of $\Omega$ is a lower set with respect to $\leq$ if $i \in L$ and $j \leq i$ implies $j \in L$. A subset $U$ of $\Omega$ is an upper set if $i \in U$ and $i \leq j$ implies $j \in U$.

The solution to the isotonic regression problem can be characterized by

$$
(2.8) \quad g^*_i = \min_{L} \max_{U} \sum_{j \in \text{U \cap L}} g_j w_j / \sum_{j \in \text{U \cap L}} w_j .
$$

This algorithm is given in Brunk, Ewing and Utz (1957) and in Brunk (1955). Additional algorithms for computing $g^*$ are discussed in van Eeden (1956), Alexander (1970), and Thompson (1962).
3. THE GENERALIZED ISOTONIC REGRESSION PROBLEM AND ITS DUAL

Let \( g \) and \( w > 0 \) be given functions on \( \Omega = \{1, 2, \ldots, k\} \). Let \( \phi \) be a convex function on an interval \( I \) and \( K \) a convex cone of functions on \( \Omega \) having range in \( I \). We call the problem

\[
\min_{x \in K} \sum_{i=1}^{k} [\phi(x_i) - g_i x_i] w_i
\]

the generalized isotonic regression problem. If \( \phi(x_i) = x_i^2 \), then (3.1) is equivalent to the isotonic regression problem (2.1).

Theorem 3.1:

Let \( \phi \) be convex and finite on the interval \( I \) and \( \phi \) a determination of its derivative, also finite on \( I \). Let \( K \) be a convex cone of isotonic functions on \( \Omega \) having range in \( I \). Suppose that the range of \( g \) is in the domain of \( \phi^{-1} \) where \( g \) is a given function on \( \Omega \). Then

\[
x_i^0 = \phi^{-1}(g_i^*) \quad i = 1, 2, \ldots, k
\]

solves (3.1) and \( g^* \) is given by (2.8). The minimizing function is unique if \( \phi \) is strictly convex.

\[\text{Proof:}\]

Since \( \phi \) is convex,

\[
\phi(x_i) - \phi(\phi^{-1}(g_i^*)) - [x_i - \phi^{-1}(g_i^*)] g_i^* \geq 0
\]

for all \( x_i \in I \). Note that \( g_i^* \) is in the domain of \( \phi^{-1} \) by (2.6). Hence,
\[
\sum_{i=1}^{k} \left[ \phi(x_i) - g_i x_i - \phi^{-1}(g_i^*) + g_i \phi^{-1}(g_i^*) \right] w_i
\]
\[
= \sum_{i=1}^{k} \left[ \phi(x_i) - \phi^{-1}(g_i^*) - \left[ x_i - \phi^{-1}(g_i^*) \right] g_i \right] w_i
\]
\[
+ \sum_{i=1}^{k} \left[ x_i - \phi^{-1}(g_i^*) \right] (g_i - g_i^*) w_i.
\]

The first term is \( \geq 0 \) by (3.2). The second term is \( \geq 0 \) since \( x \in K \) implies
\[
\sum_{i=1}^{k} (g_i^* - g_i) x_i w_i \geq 0
\]
by (2.3) and
\[
\sum_{i=1}^{k} (g_i^* - g_i) \phi^{-1}(g_i^*) w_i = 0
\]
by (2.7). ||

Theorem 3.1 provides the solution to problems of restricted maximum likelihood estimation for ordered parameters of multinomial, Poisson and gamma distributions among many others.

**Fenchel Duality**

The problem dual to (3.1) can be obtained via Fenchel's duality theory [Rockafellar, Theorem 31.1, (1970)]. Let \( f \) be a convex function on \( R^k \) and \( h \) a concave function on \( R^k \). Define the inner product \((\cdot, \cdot)\) by
\[
(x, y) = \sum_{i=1}^{k} x_i y_i w_i.
\]

The convex conjugate of \( f \) is by definition
\[ f^*(y) = \sup_x [(x, y) - f(x)] \]

while the concave conjugate of \( h \) is by definition

\[ h^*(y) = \inf_x [(x, y) - h(x)] \].

Note that \( f^* \) is convex and \( h^* \) is concave.

The primal problem is

(I) \[ \inf_x [f(x) - h(x)] \]

while the dual problem is

(II) \[ \sup_x [h^*(x) - f^*(x)] \].

**Theorem 3.2:** (Fenchel's Duality Theorem)

Let \( f \) be a proper convex function on \( \mathbb{R}^k \), and let \( h \) be a proper concave function on \( \mathbb{R}^k \). Then

\[ \inf_x [f(x) - h(x)] = \sup_x [h^*(x) - f^*(x)] \]

if \( \text{ri} (\text{dom } f) \cap \text{ri} (\text{dom } h) \neq \emptyset \), where \( \emptyset \) is the empty set.

A concave function \( h \) is proper if \( h(x) > -\infty \) for at least one \( x \) and \( h(x) < +\infty \) for every \( x \). A proper convex function is defined using the fact that \( f \) is convex if and only if \(-f\) is concave. \( \text{ri} (\text{dom } h) \) means the relative interior of the domain of \( h \).

For a proof of Theorem 3.2, see Rockafellar (1970).

To obtain the dual of (3.1), let
\[ \delta(x \mid K) = \begin{cases} 0 & \text{if } x \in K \\ +\infty & \text{otherwise} \end{cases} \]

If we let \( h(x) = -\delta(x \mid K) \), we have

\[
(3.3) \quad \inf_{x} [f(x) - h(x)] = \inf_{x \in K} f(x).
\]

The dual problem can be defined in terms of the dual (polar) convex cone

\[ K^* = \{ y \mid (x, y) \leq 0 \text{ for all } x \in K \}. \]

The concave conjugate to \( h(x) = -\delta(x \mid K) \) is

\[ h^*(y) = -\delta(-y \mid K^*). \]

Hence, the problem dual to (3.3) is

\[
\sup_{y} [h^*(y) - f^*(y)] = \sup_{z} [-\delta(z \mid K^*) - f^*(-z)]
\]

\[ = \sup_{z \in K^*} -f^*(-z). \]

The primal problem becomes

(I) \[ \inf_{x \in K} f(x) \]

and the dual problem is

(II) \[ \sup_{y \in K^*} -f^*(-y). \]

Fenchel's duality theorem can be restated for this setup as follows.
Theorem 3.3: [Rockafellar, Theorem 31.4]

Let $f$ be a closed proper convex function on $\mathbb{R}^k$, and let $K$ be a closed convex cone in $\mathbb{R}^k$. Let $K^*$ be the dual of $K$, i.e.,

$$K^* = \{ y \mid (x, y) \leq 0 \text{ for all } x \in K \}.$$

Then

$$\inf f(x) = \sup_{x \in K} -f^*(-y)$$

\[ \text{if } \operatorname{ri} (\operatorname{dom} f) \cap \operatorname{ri} K \neq \mathbb{R} \].

In general, if $f$ is differentiable with gradient $\nabla f$, then $x^o$ and $y^o$ satisfy

$$f(x^o) = \inf f(x) = \sup_{x \in K} -f^*(-y) = f^*(-y^o)$$

if and only if

$$y^o = -\nabla f(x^o), \ x^o \in K, \ y^o \in K^* \text{ and } (x^o, y^o) = 0.$$

The proof can be found in Rockafellar (1970).

The Dual of the Generalized Isotonic Regression Problem

Returning to the generalized isotonic regression problem (3.1), let

$$f(x) = \sum_{i=1}^{k} [\phi(x_i) - g_i x_i] w_i.$$

Then

$$\phi^*(y) = \sup_{x} [(x, y) - f(x)] = \sum_{i=1}^{k} \left[ \phi^{-1}(g_i + y_i) - (g_i - y_i) \phi^{-1}(g_i + y_i) w_i \right].$$
where \( \phi \) is a determination of the derivative of \( \phi \). The problem dual to the isotonic regression problem is

\[
\sup_{y \in K^*} -f^*(-y) = \sup_{y \in K^*} \sum_{i=1}^{k} \left[ \phi^{-1}(s_i - y_i) \right] - (s_i - y_i) \phi^{-1}(s_i - y_i) w_i \\
= -\inf_{y \in K^*} \sum_{i=1}^{k} \theta(s_i - y_i) w_i
\]

where \( \theta(s_i - y_i) = (s_i - y_i) \phi^{-1}(s_i - y_i) - \phi\left(\phi^{-1}(s_i - y_i)\right) \). Note that \( \theta'(y) = \theta(y) = \phi^{-1}(y) \).

Recapitulating, the primal problem is

(I) \[ \text{Minimize} \sum_{x \in K, i=1}^{k} [\phi(x_i) - g_i x_i] w_i \]

and the dual problem is

(II) \[ \text{Minimize} \sum_{y \in K^*, i=1}^{k} \theta(s_i - y_i) w_i \]

where \( \theta(y) = \phi^{-1}(y) \). It follows from Theorem 3.3 that the solution of (II) satisfies

\[ y_i^o = -\phi(x_i^o) + s_i \]

where \( x_i^o \) solves (I). By Theorem 3.1, \( x_i^o = \phi^{-1}(s_i^*) \). Hence, \( y_i^o = s_i - s_i \) for \( i = 1, 2, \ldots, k \) and is independent of \( \theta \). This invariance result can be proved directly using the properties of the isotonic regression.

**Theorem 3.4:**

Let \( g \) and \( w \) be given functions on \( \Omega = \{1, 2, \ldots, k\} \), \( w_i > 0 \) for \( i \in \Omega \).
Let \( \theta \) be a convex function finite on an interval \( I \) containing the range of \( g \). Let \( \theta \) be a determination of its derivative, also finite on \( I \). Let \( K \) be the class of functions on \( \Omega \) isotonic with respect to a partial order on \( \Omega \), and let \( K^* \) be the dual cone. Then \( g^* \) solves the problem

\[
(3.4) \quad \text{Minimize } \sum_{i=1}^{k} \theta(y_i)w_i^*.
\]

The minimizing function is unique if \( \theta \) is strictly convex.

**Proof:**

Since \( g \in I \), also \( g^* \in I \) by (2.6). Since \( \theta \) is convex,

\[
\theta(y_i) \geq \theta(g_i^*) + \theta(g_i^*)[y_i - g_i^*] + (g_i^* - g_i^*)
\]

By (2.7),

\[
\sum_{i=1}^{k} \theta(g_i^*)[g_i^* - g_i^*]w_i = 0.
\]

Also, since \( g^* \) is isotonic and \( \theta \) is nondecreasing, \( \theta(g^*) \in K \). Then since \( g - y \in K^* \),

\[
\sum_{i=1}^{k} \theta(g_i^*)(g_i^* - y_i^*)w_i \leq 0.
\]

Thus, finally,

\[
\sum_{i=1}^{k} \theta(y_i)w_i \geq \sum_{i=1}^{k} \theta(g_i^*)w_i
\]

for all \( y \) such that \( g - y \in K^* \). The uniqueness follows from the strict inequality in
\[ \theta(u) \geq \theta(v) + (u - v)\theta'(v) \]

when \( \theta \) is strictly convex. ||

Arthur Veinott first noted this duality result in the case where \( K \) is determined by a complete ordering. He was motivated by an inventory application.
4. APPLICATIONS

Dantzig (1952) treats the case \( \theta(u) = \frac{u^2}{2} \) of Theorem 3.4, but with a more general class of restrictions. He also observed the following geometrical interpretation of a version of the dual problem.

The Taut String Solution

Let \( \theta(y) = [1 + y^2]^{\frac{k}{2}} \). Then the dual problem (3.4) becomes

\[
\text{Minimize } \sum_{i=1}^{k} \left[ 1 + y_i^2 \right]^{\frac{k}{2}} w_i.
\]

Let \( K \) be the convex cone of increasing functions, i.e., the ordering on \( \Omega \) is the natural ordering. The upper sets are sets of the form \( \{i, i+1, \ldots, k\} \) for \( i = 1, 2, \ldots, k \). If \( g - y \in K^* \), then

\[
(4.1) \quad \sum_{i=1}^{k} \left( g_i - y_i \right) x_i w_i \leq 0
\]

for all \( x \in K \) implies in particular that

\[
(4.2) \quad \sum_{j=1}^{i} g_j w_j \geq \sum_{j=1}^{i} y_j w_j \quad i = 1, 2, \ldots, k
\]

and

\[
\sum_{j=1}^{k} g_j w_j = \sum_{j=1}^{k} g_j w_j
\]

since the functions \( x \equiv 1 \) and \( x \equiv -1 \) belong to \( K \). Likewise, (4.2) implies (4.1). Hence, in this case, our dual problem becomes
Minimize \( \sum_{i=1}^{k} \left[ 1 + y_{i}^{2} \right]^{\frac{1}{k}} w_{i} \)

Subject to \( \sum_{j=1}^{i} y_{j}w_{j} \leq \sum_{j=1}^{i} g_{j}w_{j} \) \quad 1 \leq i \leq k

and \( \sum_{j=1}^{k} y_{j}w_{j} = \sum_{j=1}^{k} g_{j}w_{j} \).

Figure 4.1 illustrates the problem for \( k = 5 \). The cumulative sum diagram (CSD) of \( g \) is the set of points in the Cartesian plane consisting of the origin, \( P_{0} : (0,0) \) and the points \( P_{i} \) with coordinates \( \left( \sum_{j=1}^{i} w_{j}, \sum_{j=1}^{i} g_{j}w_{j} \right) \) \( i = 1,2, \ldots, k \). Imagine a string which joins \( P_{0} \) and \( P_{k} \) and passes nowhere above the CSD. For each \( i \in \Omega \) consider the slope of the segment of the string arriving at \( P_{i} \) from the left. In order for the string to be taut, this slope, as a function on \( \Omega \), must minimize the total length of the string,

\[ \sum_{i=1}^{k} \left[ 1 + y_{i}^{2} \right]^{\frac{1}{k}} w_{i}, \]

in the class of functions \( y \) satisfying the inequality constraints. The solution to this problem

\[ y_{i}^{o} = \min_{t \geq 1} \max_{s \leq i} \frac{t}{s} \sum_{j=s}^{t} g_{j}w_{j}/ \sum_{j=s}^{t} w_{j} \]

is geometrically evident. By Theorem 3.4, this is also the solution, \( g^{*} \), to the isotonic regression problem when \( K \) is the convex cone of increasing functions.

Remark:

When \( K \) is the cone of increasing vectors in \( R^{k} \), the dual cone \( K^{*} \) induces a cone ordering on \( R^{k} \), namely majorization. To be precise, a real \( k \) vector \( y \) is said to majorize the \( k \) vector \( x \) if the components of \( x \) and \( y \) can be reordered so that
\[ G_1 = \sum_{j=1}^{i} g_j w_j \]

**FIGURE 4.1: THE TAUT STRING SOLUTION**
(i) \[ x_1 \geq x_2 \geq \ldots \geq x_k, \ y_1 \geq y_2 \geq \ldots \geq y_k. \]

(ii) \[ \sum_{j=1}^{i} x_j \leq \sum_{j=1}^{i} y_j, \ i = 1, 2, \ldots, k-1, \]

(iii) \[ \sum_{j=1}^{k} x_j = \sum_{j=1}^{k} y_j. \]

[Cf. Hardy, Littlewood and Polya (1952), p. 45.]

**A Production Planning Problem**

Professor Arthur Veinott kindly pointed out the following production planning problem and its connection with isotonic regression.

The dual cone of the cone of isotonic functions on \( \Omega \) enters naturally in the following production planning problem. Suppose an amount \( y \) of goods is to be manufactured during the \( i \)th period, \( i = 1, 2, \ldots, k \). Suppose the cost of manufacturing an amount \( y \) is \( C(y) \) where \( C(\cdot) \) is convex. Suppose, furthermore, there is a demand for \( r_i \) goods at the \( i \)th period, \( i = 1, 2, \ldots, k \). The problem is to choose the production vector \( y = (y_1, y_2, \ldots, y_k) \) so as to minimize total production cost,

\[ \sum_{i=1}^{k} C(y_i) \]

subject to the requirement that demand never exceeds supply;

\[ \sum_{j=1}^{i} (y_j - x_j) \geq 0, \quad i = 1, 2, \ldots, k-1, \]

and that total demand is equal to total production:

\[ \sum_{j=1}^{k} (y_j - x_j) = 0. \]
If $K$ is the convex cone of vectors in $\mathbb{R}^k$ with decreasing coordinates, then $g - y \in K^*$ implies that
\[
\sum_{j=1}^i (g_j - y_j)w_j \leq 0 , \quad i = 1, 2, \ldots, k-1
\]
and
\[
\sum_{j=1}^k (g_j - y_j)w_j = 0 .
\]

If we let $g \equiv r$ and $w \equiv 1$, then we see that this is the dual problem of Theorem 3.4. The solution is
\[
y_i^* = \min_{s \leq i} \max_{t > i} \sum_{j=s}^t r_j / (t - s + 1) .
\]

The Isotonic Regression Problem and Its Dual

The isotonic regression problem is equivalent to
\[
\text{Minimize} \sum_{i=1}^k \frac{1}{2} [g_i - x_i]^2 w_i
\]
where $K$ is a convex cone. Its Fenchel dual is
\[
\text{Maximize} -\frac{1}{2} \sum_{i=1}^k \left[ y_i^2 - 2y_i g_i \right] w_i
\]
which is equivalent to the problem
\[
\text{Minimize} \sum_{i=1}^k \frac{1}{2} [y_i - g_i]^2 w_i
\]
Geometrically, the isotonic regression problem and its dual can be pictured as in Figure 4.2. The two-dimensional cone represents a convex cone of functions. The isotonic regression problem is to find that function in $K$ which is closest in the inner product norm to the given function $g$. The solution, $g^*$, is the foot of the perpendicular to $K$ which passes through $g$. It is geometrically evident that $g - g^*$ solves the dual problem. This geometrical interpretation was also pointed out by Sinden (1962) for a more general problem.

**Optimal Capacity Expansion**

Howard and Nemhauser (1968) consider the following production problem. The level of production capacity (or plant size) is $c_i$ for period $i$ ($i = 1, 2, \ldots, k$). Let $\Theta(c_i)$ be the cost of investment and cost of capacity shortages and excesses in the $i$th time period. We assume that $\Theta(c_i)$ is a strictly convex differentiable function of $c_i$. The problem is to

$$\begin{align*}
\text{Minimize} & \quad \sum_{i=1}^{k} \Theta_i(c_i) \\
\text{Subject to} & \quad 0 \leq c_1 \leq c_2 \leq \ldots \leq c_k.
\end{align*}$$

(4.3)

If $\Theta_i(c_i) = \phi(c_i) - g_i c_i$, then this would be the generalized isotonic regression problem in the simple ordering case. Howard and Nemhauser (1968) solve this problem and several variations of it. We generalize this problem by assuming a partial ordering ($\preceq$) instead of a complete ordering on $c = (c_1, c_2, \ldots, c_k)$. Let $K$ be the cone of isotonic vectors with respect to $\preceq$. Then (4.3) becomes

$$\begin{align*}
\text{Minimize} & \quad \sum_{i=1}^{k} \Theta_i(c_i) \\
\text{Subject to} & \quad c \in K.
\end{align*}$$

(4.4)
Primal

Minimize $\|g - x\|$

Subject to $x \in K$

Dual

Minimize $\|y\|$

Subject to $g - y \in K^*$

FIGURE 4.2: THE GEOMETRY OF ISOTONIC REGRESSION
One characterization of the solution resembles (2.8). Let $\mathcal{U}$ and $\mathcal{L}$ be the upper and lower sets respectively induced by the partial ordering $\preceq$. Let $z_{\mathcal{U} \cap \mathcal{L}}$ be the zero (assumed to exist) of

$$\sum_{j \in \mathcal{U} \cap \mathcal{L}} \theta_j(z_{\mathcal{U} \cap \mathcal{L}}) = 0$$

where $\theta_j$ is the derivative of $\theta_j$ and $U \in \mathcal{U}$, $L \in \mathcal{L}$. The solution to (4.4) can be expressed as

$$z^*_i = \min \max_{L \in \mathcal{L}} \max_{U \in \mathcal{U}} z_{\mathcal{U} \cap \mathcal{L}}.$$  

This min max formula is contained in Theorem 1.12 of Brunk and Johansen (1970).
5. MAXIMUM LIKELIHOOD ESTIMATION OF STOCHASTICALLY ORDERED DISTRIBUTIONS

Brunk et al (1966) solved the problem of maximum likelihood estimation of two stochastically ordered distributions. We show that the problem is equivalent to a version of the dual to the isotonic regression problem.

Let a random variable \( X \) be known to be stochastically smaller than a random variable \( Y \); i.e.,

\[
F(x) = P[X \leq x] \geq G(x) = P[Y \leq x]
\]

for all real \( x \). Suppose all \( m + n \) observations of random samples of size \( m \) from \( F \) and \( n \) from \( G \) are arranged in increasing order. Some may coincide; when \( x \)- and \( y \)-observations occur at the same point, list the \( x \)-observations first. Reading from left to right, let \( a_1 \) \( x \)-observations precede all \( y \)-observations. It will be convenient to assume \( a_1 > 0 \); the contrary case can easily be reduced to this one. Let \( b_1 \) \( y \)-observations precede the next group of \( x \)-observations; \( b_1 > 0 \). Let \( a_2 \) uninterrupted \( x \)-observations follow; \( a_2 > 0 \) (unless \( a_1 = m \)). Following these are \( b_2 \) \( y \)-observations; again \( b_2 > 0 \) unless \( b_1 = n \), etc. Suppose in this way the observations are separated into \( r \) groups. We assume again that \( b_r > 0 \); the contrary case can be reduced to this one. Let \( I_i \) be an interval containing only the first \( a_1 \) \( x \)-observations and the first \( b_1 \) \( y \)-observations; \( I_2 \) a contiguous interval containing only the second \( a_2 \) \( x \)-observations and the second \( b_2 \) \( y \)-observations, etc. The maximum likelihood estimate of the distribution function \( F \), assumed discrete, will assign equal (unknown) probabilities \( p_i \) to all \( x \)-observations in \( I_i \) \((i = 1, 2, \ldots, r)\). The maximum likelihood estimate of \( G \) will assign equal probabilities \( q_i \) to the \( y \)-observations in \( I_i \) \((i = 1, 2, \ldots, r)\).

The stochastic ordering requirement will be satisfied if and only if

\[
\sum_{j=1}^{i} a_j p_j \geq \sum_{j=1}^{i} b_j q_j \quad \text{for} \quad i = 1, 2, \ldots, r.
\]
We are thus led to formulate the problem of maximum likelihood estimation in terms of multinomial parameters:

\[ a_i p_i = P[X \in I_i], b_i q_i = P[Y \in I_i], \quad i = 1, 2, \ldots, r. \]

The problem is

(I)

\[ \text{Maximize} \quad \prod_{i=1}^{r} \frac{a_i b_i}{p_i q_i} \]

(5.1)

\[ \text{Subject to} \quad \sum_{j=1}^{i} a_j p_j \geq \sum_{j=1}^{i} b_j q_j \quad i = 1, 2, \ldots, r, \]

and

\[ \sum_{j=1}^{r} a_j p_j = \sum_{j=1}^{r} b_j q_j = 1. \]

(5.3)

If we let \( \rho_i = a_i p_i \) and \( \sigma_i = b_i q_i \) (\( i = 1, 2, \ldots, r \)), then Problem (I) is equivalent to

(II)

\[ \text{Maximize} \quad \prod_{i=1}^{r} \frac{a_i b_i}{\rho_i \sigma_i} \]

(5.4)

\[ \text{Subject to} \quad \sum_{j=1}^{i} \rho_j \geq \sum_{j=1}^{i} \sigma_j \quad i = 1, 2, \ldots, r, \]

and

\[ \sum_{j=1}^{r} \rho_j = \sum_{j=1}^{r} \sigma_j = 1. \]

(5.6)

If we replace conditions (5.5) by
(5.7) \[
\sum_{j=1}^{i} \rho_j = \sum_{j=1}^{i} \sigma_j \quad i = 1, 2, \ldots, r,
\]

we easily see that the solution is

\[
\hat{\rho}_i = \hat{\sigma}_i = \frac{(a_i + b_i)}{N}
\]

where \( N = \sum_{i=1}^{r} (a_i + b_i) \).

It is intuitively clear (and we prove it in Theorem 5.2) that Problem (II) is equivalent to

(III)

(5.8) \[
\text{Maximize} \quad \prod_{i=1}^{r} \frac{a_i b_i}{\rho_i \sigma_i}
\]

(5.9) \[
\text{Subject to} \quad \sum_{j=1}^{i} \rho_j \geq \sum_{j=1}^{i} \frac{(a_j + b_j)}{N} \quad i = 1, 2, \ldots, r
\]

\[
\sum_{j=1}^{i} \frac{(a_j + b_j)}{N} \geq \sum_{j=1}^{i} \sigma_j \quad i = 1, 2, \ldots, r,
\]

and

(5.10) \[
\sum_{j=1}^{r} \rho_j = \sum_{j=1}^{r} \sigma_j = 1.
\]

If we make the usual change of variable again, Problem (I) becomes
\[(I')\]

Maximize \[ \prod_{i=1}^{r} \frac{a_i b_i}{p_i q_i} \]

Subject to \[ \sum_{j=1}^{i} a_j p_j \geq \sum_{j=1}^{i} a_j g_j \]
\[ i = 1, 2, \ldots, r \]

\[ \sum_{j=1}^{i} b_j g_{j+r} \geq \sum_{j=1}^{i} b_j q_j \]
\[ i = 1, 2, \ldots, r \]

and

\[ \sum_{j=1}^{r} a_j p_j = \sum_{j=1}^{r} b_j q_j = 1 \]

where

\[ g_j = \frac{(a_j + b_j)}{Na_j} \]
\[ j = 1, 2, \ldots, r \]

\[ g_j = \frac{(a_{j-r} + b_{j-r})}{Nb_{j-r}} \]
\[ j = r+1, \ldots, 2r \]

Let

\[ g = (g_1, g_2, \ldots, g_r, g_{r+1}, \ldots, g_{2r}) \]

\[ z = (p_1, p_2, \ldots, p_r, q_1, q_2, \ldots, q_r) \]

\[ w = (a_1, a_2, \ldots, a_r, b_1, b_2, \ldots, b_r) \]

and \( (x, y) = \sum_{i=1}^{2r} x_i y_i w_i \). Then Problem \((I')\) is equivalent to

Minimize \[ \sum_{i=1}^{r} - [\log z_i] a_i + \sum_{i=r+1}^{2r} - [\log z_i] b_{i-r} \]

Subject to \( g - z \in \mathbb{R}^* \)
where $K^*$ is the dual cone to $K$ and $K$ is the cone of isotonic functions decreasing on $\{1, 2, \ldots, r\}$ and increasing on $\{r+1, \ldots, 2r\}$. (The restriction $g - z \in K^*$ is equivalent to $(g - z, x) \leq 0$ for all $x \in K$.) The solution follows from Theorem 3.4 where $\Theta(z_i) = -\log z_i$ and

$$
\hat{p}_i = g^*_i = \min_{1 \leq s \leq i \leq t < r} \max_{s \leq j \leq r} \frac{\sum_{j=s}^{t} (a_j + b_j)}{a_j} \quad i = 1, 2, \ldots, r
$$

and

$$
\hat{q}_{i-r} = g^*_i = \min_{i \leq t < 2r \leq s \leq i} \max_{s \leq j \leq r} \frac{\sum_{j=s}^{t} (a_{j-r} + b_{j-r})}{b_{j-r}} \quad i = r+1, \ldots, 2r.
$$

To prove that Problem (II) (i.e., (5.4) through (5.6)) is equivalent to Problem (III) (i.e., (5.8) through (5.10)), we need the following lemma.

**Lemma 5.1:**

Let $\{\hat{\rho}_i\}_{i=1}^{r}$, $\{\bar{\sigma}_i\}_{i=1}^{r}$, $\{\bar{z}_i\}_{i=1}^{r}$ maximize, respectively,

$$
\sum_{i=1}^{r} a_i \log \rho_i, \sum_{i=1}^{r} b_i \log \sigma_i, \sum_{i=1}^{r} (a_i + b_i) \log z_i
$$

subject to $\sum_{i=1}^{r} \rho_i = \sum_{i=1}^{r} \sigma_i = \sum_{i=1}^{r} z_i = 1$. If

$$
(5.11) \quad \sum_{j=1}^{i} \bar{\rho}_j \geq \sum_{j=1}^{i} \bar{\sigma}_j \quad i = 1, 2, \ldots, r-1,
$$

then
\[
\frac{1}{r} \sum_{j=1}^{r} \bar{\rho}_j \geq \frac{1}{r} \sum_{j=1}^{r} \bar{z}_j = \frac{1}{r} \sum_{j=1}^{r} \frac{(a_j + b_j)}{(a_j + b_j)} \geq \frac{1}{r} \bar{\sigma}_j \quad i = 1, 2, \ldots, r-1.
\]

**Proof:**

This is an easy consequence of the solutions

\[
\bar{\rho}_i = \frac{a_i}{\sum_{j=1}^{r} a_j} \quad \text{and} \quad \bar{\sigma}_i = \frac{b_i}{\sum_{j=1}^{r} b_j} \quad i = 1, 2, \ldots, r.
\]

**Theorem 5.2:**

The solution of Problem (II) satisfies

\[
(5.12) \quad \frac{1}{r} \sum_{j=1}^{r} \rho_j \geq \frac{1}{r} \sum_{j=1}^{r} \frac{(a_j + b_j)}{(a_j + b_j)} \geq \frac{1}{r} \sum_{j=1}^{r} \sigma_j \quad i = 1, 2, \ldots, r.
\]

Hence, the solution to Problem (II) also solves Problem (III).

**Proof:**

The proof is by induction on \( r \).

Let \( r = 2 \):

Let \( \{\hat{\rho}_i\}_{i=1}^{2} \) and \( \{\hat{\sigma}_i\}_{i=1}^{2} \) solve Problem (II).

**Case 1:**

Suppose \( \hat{\rho}_1 > \hat{\sigma}_1 \). Then \( \hat{\rho}_1 \) maximizes

\[
a_1 \log \rho_1 + a_2 \log (1 - \rho_1)
\]
and \( \hat{\sigma}_1 \) maximizes

\[ b_1 \log \sigma_1 + b_2 \log (1 - \sigma_1) \]

since otherwise (5.1) could be increased by perturbing \((\hat{\rho}_1, \hat{\sigma}_1)\) away from \((\hat{\rho}_1, \hat{\sigma}_1)\). Then (5.12) follows from Lemma 5.1.

**Case 2:**

Suppose \( \hat{\rho}_1 = \hat{\sigma}_1 \). Then \( (\hat{\rho}_1, 1 - \hat{\rho}_1) \) maximizes (5.1) also subject to \( \rho_1 = \sigma_1 \); i.e., maximizes

\[ (a_1 + b_1) \log z_1 + (a_2 + b_2) \log (1 - z_1) \]

and therefore \( \hat{\rho}_1 = \frac{a_1 + b_1}{N} = \hat{\sigma}_1 \).

We illustrate the induction with \( r = 4 \).

**Let** \( r = 4 \):

Let \( \{\hat{\rho}_i\}_{i=1}^4 \) and \( \{\hat{\sigma}_i\}_{i=1}^4 \) solve Problem (II).

**Case 1:**

\( \sum_{j=1}^{i} \rho_j > \sum_{j=1}^{i} \sigma_j \) \( i = 1, 2, 3 \).

Then \( \{\hat{\rho}_i\}_{i=1}^4 \) maximizes \( \sum_{i=1}^{4} a_i \log \rho_i \) given \( \sum_{i=1}^{4} \rho_i = 1 \) and \( \{\hat{\sigma}_i\}_{i=1}^4 \) maximizes \( \sum_{i=1}^{4} b_i \log \sigma_i \) given \( \sum_{i=1}^{4} \sigma_i = 1 \) for otherwise (5.1) could be increased by perturbing \( \hat{\rho} \) or \( \hat{\sigma} \) slightly. Then \( \{\hat{\rho}_i, \hat{\sigma}_i\}_{i=1}^4 \) satisfies (5.12) by Lemma 5.1.
Case 2:

Suppose \( \frac{3}{i} \hat{\rho}_i = \frac{3}{i} \hat{\sigma}_i \). Then \( \{\hat{\rho}_i, \hat{\sigma}_i\}_1^4 \) maximizes (5.1) subject to (5.2), (5.3) and

\[
\frac{3}{i} \hat{\rho}_i = \frac{3}{i} \hat{\sigma}_i.
\]

Make the change of variable

\[
\beta = \frac{3}{i} \hat{\rho}_i = \frac{3}{i} \hat{\sigma}_i
\]

and \( \xi_i = \frac{\hat{\rho}_i}{\beta} \), \( \eta_i = \frac{\hat{\sigma}_i}{\beta} \) \( i = 1, 2, 3 \). (5.1) becomes

\[
N - (a_4 + b_4) - \frac{a_4 + b_4}{1 - \beta} \prod_{i=1}^{3} \xi_i \eta_i
\]

while (5.2) becomes \( \xi_1 \geq \eta_1 \) and \( \xi_1 + \xi_2 \geq \eta_1 + \eta_2 \). (5.3) becomes

\[
\frac{3}{i} \xi_i = \frac{3}{i} \eta_i = 1.
\]

The solution is

\[
\hat{\beta} = \frac{N - (a_4 + b_4)}{N}
\]

\[
\hat{\xi}_i = \frac{\hat{\rho}_i}{\hat{\beta}}, \quad \hat{\eta}_i = \frac{\hat{\sigma}_i}{\hat{\beta}} \quad i = 1, 2, 3.
\]

By the inductive hypothesis,

\[
\frac{1}{i} \sum_{j=1}^{i} \hat{\xi}_j \geq \frac{1}{i} \sum_{j=1}^{i} \frac{a_j + b_j}{3} \geq \frac{1}{i} \hat{\eta}_i \quad i = 1, 2.
\]
Hence,

\[ \frac{1}{N} \sum_{j=1}^{n} \rho_j \geq \frac{1}{N} \left( \frac{1}{2} (a_j + b_j) \right) \geq \frac{1}{2} c_j \]

\[ i = 1, 2. \]

Other cases can be verified in a similar manner.

--

ACKNOWLEDGEMENT

The authors would like to acknowledge many helpful conversations with Professor Arthur Veinott.
REFERENCES


