ON RECURSIVE FORMULAS FOR ISOTONIC REGRESSION USEFUL FOR STATISTICAL INFERENCE UNDER ORDER RESTRICTIONS

by

Prem S. Puri*
Department of Statistics
Purdue University
West Lafayette, IN 47907 USA

Harshinder Singh
Department of Statistics
Panjab University
Chandigarh-160014, INDIA

Technical Report # 88-51

Department of Statistics
Purdue University

October 1988
Revised July 1989

*Research supported in part by the U.S. National Science Foundation Grant No.: DMS-8504319 at Purdue University.

△

ON RECURSIVE FORMULAS FOR ISOTONIC REGRESSION USEFUL FOR STATISTICAL INFERENCE UNDER ORDER RESTRICTIONS

by

Prem S. Puri* and Harshinder Singh

Prem S. Puri*
Department of Statistics
Purdue University
West Lafayette, IN 47907 USA

Harshinder Singh
Department of Statistics
Panjab University
Chandigarh-160014, INDIA

Abstract

New recursive formulas have been obtained for the isotonic regression of a given function, defined on a finite set, with given weight function with respect to the simple ordering on the set. These formulas directly give the isotonic regression without the use of any algorithm. Since isotonic regressions often arise in problems of statistical inference under order restrictions these formulas are found useful. We demonstrate their usefulness through an application to the problem of nonparametric estimation of a distribution function dominating stochastically a known distribution function.

AMS Subject Classification: Primary 62J99; Secondary 62G05.

Key words and phrases: Statistical inference under order restrictions; maximum likelihood estimation; isotonic regression; antitonic regression.

*Research supported in part by the U.S. National Science Foundation Grant No.: DMS-8504319 at Purdue University.

1. Introduction

The problems involving inferences about a distribution or its parameters when it is known apriori that these are subject to certain restrictions has attracted the attention of quite a few researchers in the past. For further study in this area the reader may refer to Ayer, Brunk, Ewing, Reid and Silverman (1955), Brunk (1955, 1958), Eden (1956, 1957), Marshal and Proschan (1965), Brunk, Franck, Hanson and Hogg (1966), Barlow, Bartholomew, Bremner and Brunk (1972), Robertson and Wright (1974), Dykstra (1982), Feltz and Dykstra (1985), Schoenfeld (1986), Sampson and Whitaker (1987), Robertson, Wright and Dykstra (1988). In most of the problems of statistical inference under order restrictions, there is an underlying basic mathematical concept of isotonic regression (cf. Barlow et. al. (1972)) defined below.

Let $S = \{x_1, x_2, \ldots, x_k\}$ be a finite set with the simple ordering $x_1 \prec x_2 \prec \ldots \prec x_k$. A real valued function $f(\cdot)$ on S is said to be isotonic if $x, y \varepsilon S, x \prec y$ implies $f(x) \leq f(y)$. Let $g(\cdot)$ be a given real valued function on S and $w(\cdot)$ a given positive function on S. An isotonic function $g^*(\cdot)$ on S is said to be *isotonic regression* of g with weights $w_i = w(x_i), i = 1, 2, \ldots, k$, with respect to the simple ordering $x_1 \prec x_2 \prec \ldots \prec x_k$, if it minimizes in the class of all isotonic functions f on S the sum

$$\sum_{x \in S} [g(x) - f(x)]^2 w(x). \tag{1.1}$$

A real valued function $f(\cdot)$ on S is said to be antitonic if $x, y \varepsilon X, x \prec y$ implies $f(x) \geq f(y)$. An antitonic function $\tilde{g}(\cdot)$ is said to be *antitonic regression* of g with weights $w_i, i = 1, 2, \ldots, k$, with respect to the simple ordering $x_1 \prec x_2 \prec \ldots \prec x_k$, if it minimizes in the class of all antitonic functions f on S the sum (1.1).

Let $W_j = \sum_{i=1}^{j} w_i$ and $G_j = \sum_{i=1}^{j} w_i g(x_i), j = 1, 2, \ldots, k$. The plot of the points $P_j = (W_j, G_j), j = 0, 1, 2, \ldots, k$, with $P_0 = (0, 0)$, in the Cartesian plane constitutes what is known as the cumulative sum diagram (CSD). Using this plot Reid (see Brunk (1956), Barlow et. al. (1972)) established the following theorem for finding the isotonic regression.

Theorem 1.1. *The value of the isotonic regression g^* at the point x_j is just the slope of the greatest convex minorant (GCM) of the CSD at the point P_j^* with abcissa $W_j, j = 1, 2, \ldots, k$. In particular if P_j^* is a corner of the GCM, then $g^*(x_j)$ is the slope of the segment of the GCM extending to the left.*

Ayer et. al. (1955) have expressed $g^*(x_i)$ in the form of the following max-min formulas.

$$g^*(x_i) = \max_{s \leq i} \min_{t \geq i} AV(s,t) = \min_{t \geq i} \max_{s \leq i} AV(s,t) \tag{1.2}$$

$$= \max_{s \leq i} \min_{t \geq s} AV(s,t) = \min_{t \geq i} \max_{s \leq t} AV(s,t), \tag{1.3}$$

where

$$AV(s,t) = \left(\sum_{r=s}^{t} g(x_r) w_r \right) / \left(\sum_{r=s}^{t} w_r \right). \tag{1.4}$$

Unfortunately in practice these max-min formulas are not easy to apply for calculating the isotonic regression. Consequently the above authors developed the so called "Pool Adjacent Violators" algorithm for finding the isotonic regression. Later Kruskal (1964) wrote a program to carry out this algorithm and developed essentially another version of this, called "Up-and-Down Block" algorithm.

In Section 2 of this paper we give rather simple (recursive) formulas for the isotonic and antitonic regressions which help to calculate them without using any algorithm. We believe that these formulas are not hitherto known in literature.

In Section 3 we demonstrate the usefulness of our recursive formulas (2.1) – (2.2) and (2.4) – (2.5) by applying these to the problem of maximum likelihood estimation of a distribution dominating stochastically a known distribution.

2. Formulas for isotonic and antitonic regressions

The following theorem gives the (recursive) formulas for the isotonic regression g^* of Theorem 1.1. Its proof is based on a rather simple observation leading to these formulas which have hitherto been apparently overlooked.

Theorem 2.1. *The values of isotonic regression g^* at the points x_j (denoted by g_j^*), $j = 1, 2, \ldots, k$, are given by the recursive formulas*

$$g_1^* = \min_{1 \le i \le k} \left[G_i / W_i \right], \tag{2.1}$$

$$g_j^* = \min_{j \le i \le k} \left[\left(G_i - \sum_{r=1}^{j-1} w_r g_r^* \right) / \sum_{r=j}^{i} w_r \right], j = 2, 3, \ldots, k. \tag{2.2}$$

Proof. We observe that for any convex function $h(\cdot)$, its right hand derivative $h'(x+)$ at point x is given by

$$\begin{aligned} h'(x+) &= \inf_{x' > x} [(h(x') - h(x))/(x' - x)] \\ &= \inf_A [(y' - h(x))/(x' - x)], \end{aligned} \tag{2.3}$$

where $A = \{(x', y') : x' > x, y' \ge h(x')\}$.

Now by Theorem 1.1, g_1^* is the slope of the first line segment starting from $(0, 0)$ of the GCM of the CSD. Thus by (2.3), g_1^* is the minimum of the slopes of the k lines joining the points $(0, 0)$ and the points P_i for $i = 1, 2, \ldots, k$. These slopes being $G_i/W_i, i = 1, 2, \ldots, k$, we have

$$g_1^* = \min_{1 \le i \le k} \left[G_i / W_i \right].$$

The remaining g_j^*'s are thereafter obtained recursively. More specifically, suppose we have obtained $g_1^*, g_2^*, \ldots, g_{j-1}^*$ for fixed $j \varepsilon \{2, 3, \ldots, k\}$, then as before by (2.3) g_j^*, being the slope of GCM at the point P_j^*, is the minimum of the slopes of the $k - j + 1$ lines joining the point $P_{j-1}^* = \left(W_{j-1}, \sum_{r=1}^{j-1} w_r g_r^*\right)$ and the points $P_i = \left(W_i, G_i\right), i = j, j+1, \ldots, k$. These slopes being

$$\left(G_i - \sum_{r=1}^{j-1} w_r g_r^*\right) / \left(\sum_{r=j}^{i} w_r\right), i = j, j+1, \ldots, k,$$

we have (2.2) establishing the theorem. □

The following important remark helps greatly in further reducing the calculations for finding isotonic regressions in practical situations.

Remark. In case the minimum in the expression (2.1) for g_1^* occurs at $i = r_1 \geq 2$, i.e., $g_1^* = G_{r_1}/W_{r_1}$, then it can be easily shown that $g_2^* = g_3^* = \ldots = g_{r_1}^* = g_1^*$, so that calculations for $g_j^*, j = 2, 3, \ldots, r_1$ are avoided. Next we calculate $g_{r_1+1}^*$ using the expression (2.2). If the minimum in this expression occurs at $i = r_2 > r_1 + 1$, then as before we have $g_{r_1+2}^* = g_{r_1+3}^* \cdots = g_{r_2}^* = g_{r_1+1}^*$. The process can be continued till one gets all g_j^*'s. This procedure is made clear through the following example taken from Barlow et. al. (1972) (Table 1.2, page 14).

Calculation of an isotonic regression with recursive formulas.

j	1	2	3	4	5
$w(x_j)$	1	2	1	3	1
$g(x_j)$	-1	1	4	$-(4/3)$	2
W_j	1	3	4	7	8
G_j	-1	1	5	1	3

Using (2.1), we have

$$g_1^* = \min_{1 \leq i \leq 5} \left[G_i/W_i\right] = \min[-1, 1/3, 5/4, 1/7, 3/8] = -1.$$

Since the minimum has occurred at the first term itself (i.e., $r_1 = 1$), so we need to calculate g_2^* from (2.2).

$$g_2^* = \min_{2\leq i\leq 5}\left[\left(G_i - w_1 g_1^*\right) / \sum_{r=2}^{i} w_r\right]$$
$$= \min[1, 2, 1/3, 4/7] = 1/3.$$

Since the minimum in expression for g_2^* is attained at $i = 4$, we have $g_3^* = g_4^* = g_2^* = 1/3$ and finally again from (2.2) we have

$$g_5^* = \left(G_5 - \sum_{r=1}^{4} w_r g_r^*\right) / w_5 = 2.$$

The antitonic regression $\tilde{g}$ at x_j is given by the slope at P_j^*, having abcissa W_j, of the least concave majorant (LCM) of the CSD, $j = 1, 2, \ldots, k$, (ref. Barlow et. al. (1972)). The following theorem, having similar proof as that of Theorem 2.1, gives (recursive) formulas for the antitonic regression.

Theorem 2.2. *The values of the antitonic regression $\tilde{g}$ at the points x_j (denoted by $\tilde{g}_j$), $j = 1, 2, \ldots, k$, are given by the recursive formulas*

$$\tilde{g}_1 = \max_{1\leq i\leq k}\left[G_i/W_i\right], \tag{2.4}$$
$$\tilde{g}_j = \max_{j\leq i\leq k}\left[\left(G_i - \sum_{r=1}^{j-1} w_r \tilde{g}_r\right) / \sum_{r=j}^{i} w_r\right], j = 2, 3, \ldots, k. \tag{2.5}$$

A similar remark, as was made following the Theorem 2.1, also holds in these recursive formulas for calculating the antitonic regression. This as before helps in reducing the calculations.

The following Theorem, to be used later, easily follows from our Theorem 2.1 and a corollary to Proposition 1.1 of Barlow et. al. (1972), (page 51).

Theorem 2.3. *Let ϕ be a convex, finite real valued function defined on an interval I of the real line. Let $\beta_1, \beta_2, \ldots, \beta_k$ be k real numbers, $w_i > 0, i = 1, 2, \ldots, k$ and*

$$D = \{(x_1, x_2, \ldots, x_k) : x_i \varepsilon I, i = 1, 2, \ldots, k, \sum_{i=1}^{j} w_i x_i \leq \beta_j, 1 \leq j \leq k-1, \sum_{i=1}^{k} w_i x_i = \beta_k\},$$

assumed to be a nonempty set. Then subject to $(x_1, x_2, \ldots, x_k) \varepsilon D$, the sum

$$\sum_{i=1}^{k} \phi(x_i) w_i \tag{2.6}$$

is minimized at a point $(\tau_1, \tau_2, \ldots, \tau_k) \varepsilon D$, where τ_i's are given by

$$\tau_1 = \min_{1 \leq i \leq k} \left[\beta_i \Big/ \left(\sum_{r=1}^{i} w_r \right) \right], \tag{2.7}$$

$$\tau_j = \min_{j \leq i \leq k} \left[\left(\beta_i - \sum_{r=1}^{j-1} w_r \tau_r \right) \Big/ \sum_{r=j}^{i} w_r \right], j = 2, 3, \ldots, k. \tag{2.8}$$

The above minimizing solution is unique if ϕ is strictly convex.

3. Maximum likelihood estimation of a cumulative distribution function F subject to $F(x) \leq F_0(x)$

Dykstra (1982) discussed as a special case the problem of maximum likelihood estimation of a survival function $P(t)$ subject to $P(t) \geq P_0(t)$, for the noncensoring case, where $P_0(t)$ denotes a given survival function. He developed an algorithm to calculate maximum likelihood estimator (M.L.E.) of $P(t)$. In this section we discuss, in the light of the Theorems 2.1 and 2.3, the maximum likelihood estimation of a cumulative distribution function (c.d.f.) $F(x)$ subject to the restriction $F(x) \leq F_0(x)$, $\forall\, x$, where $F_0(x)$ is a known c.d.f. Based on a random sample $y_1, y_2, \ldots, y_n$ of size n drawn from a population with c.d.f. F, let the ith ordered distinct value v_i occur m_i times, $i = 1, 2, \ldots, s$, with $\sum_{i=1}^{s} m_i = n$. We assume $F_0(v_1) > 0$. The problem is to obtain an M.L.E. of F subject to $F(x) \leq F_0(x)$, $\forall\, x$, based on the observation vector $(v_i, m_i, i = 1, 2, \ldots, s)$. Basically this is a nonparametric problem where it is not clear how to define what could be realistically called a likelihood function. This is because of the absence of a common σ–finite measure dominating every measure induced by F on $(\mathbf{R}, \mathcal{B})$, as F varies subject to $F(x) \leq F_0(x)$, $\forall x$, where $\mathcal{B}$ is the Borel σ–field on the real line $\mathbf{R}$. Instead we follow the "method of maximum likelihood" suggested by Scholz (1980), which bypasses the intermediate step of first defining the so called likelihood function. For this we shall need the following definitions as given by Scholz (1980).

Let $\mathcal{Y}$ be a metric space with metric d and let $\mathcal{P}$ be a family of probability measures on the Borel sets of $\mathcal{Y}$. For any (data) point $y \in \mathcal{Y}$ let $\mathcal{N}_y$ denote the family of all measurable sets N_y which contain y as an interior point. Let $D(N_y)$ denote the diameter of the set N_y.

Definition 1. *For $P, Q \in \mathcal{P}$ write $P \overset{y}{\geq} Q$ if*

$$\underline{\lim}\; P(N_y)/Q(N_y) \geq 1, \tag{3.1}$$

where $\underline{\lim}$ is to be interpreted here as

$$\underline{\lim}\; P(N_y)/Q(N_y) = \liminf_{\varepsilon \to 0} \{P(N_y)/Q(N_y) \colon N_y \in \mathcal{N}_y \text{ with } D(N_y) \leq \varepsilon\}, \tag{3.2}$$

with 0/0 to be taken as one by convention.

Definition 2. *$P, Q \in \mathcal{P}$ are said to be equivalent at y (written as $P \overset{y}{=} Q$) whenever $P \overset{y}{\geq} Q$ and $Q \overset{y}{\geq} P$. Thus $P \overset{y}{=} Q$ if and only if* $\underline{\lim}$ of 3.1 *exists and equals* 1.

Definition 3. *The statistic $P_0 \in \mathcal{P}$ is a maximum likelihood estimator (M.L.E.) with respect to $y \in \mathcal{Y}$ and $\mathcal{P}$ if for every $Q \in \mathcal{P}$ such that $Q \overset{y}{\geq} P_0$ it follows that $Q \overset{y}{=} P_0$. That is,*

P_0 is an M.L.E. if and only if there does not exist a $Q \in \mathcal{P} - \{P_0\}_y$ such that $Q \overset{y}{\geq} P_0$ or equivalently

$$\underline{\lim}\, Q(N_y)/P_0(N_y) < 1, \text{ for all } Q \in \mathcal{P} - \{P_0\}_y, \tag{3.3}$$

where $\{P\}_y = \{Q \in \mathcal{P} : Q \overset{y}{=} P\}$.

Before applying the above method to our problem we shall treat the problem heuristically and shall thereby come up with an "M.L.E." of F subject to $F(x) \leq F_0(x)$ by using Theorem 2.3. We shall then show that this estimate is indeed an M.L.E. also in the sense of Definition 3 of Scholz (1980). Heuristically we argue as follows:

If F is in the class of distributions which are absolutely continuous (with respect to Lebesgue measure) then the likelihood function of our observation vector based on the density function of F can be made arbitrarily large by choosing an F with a density which has arbitrarily large heights at the observation points v_i's. This essentially means (through a limiting argument) that we could restrict ourselves to the class of those F's which have positive (discrete) probability mass at each of the points $v_1, v_2, \ldots, v_s$, so that the "likelihood function" based on the probability of the observation vector $(v_i, m_i, i = 1, 2, \ldots, s)$ is given by

$$\prod_{i=1}^{s} [dF(v_i)]^{m_i}, \tag{3.4}$$

where $dF(x) = F(x) - F(x-)$. Clearly (3.4) is zero if any v_i is a point of continuity of F, while it is positive if $dF(v_i) > 0,\ \forall\, 1 \leq i \leq s$. One can now easily see that subject to $F(x) \leq F_0(x),\ \forall x$, whenever, *either* $dF(y) > 0$ for some $y \notin \{v_i, v_2, \ldots, v_s\}$ but with $y \leq v_s$ *or* there is some left–over absolutely continuous component of F over the interval $(-\infty, v_s]$, in either case there always exists another c.d.f. say F_1, with $F_1(x) \leq F_0(x),\ \forall x$, which is discrete over the interval $(-\infty, v_s]$ with strictly positive jumps *only* at the points $v_1 < v_2 < \ldots < v_s$ and with

$$\prod_{i=1}^{s}[dF_1(v_i)]^{m_i} > \prod_{i=1}^{s}[dF(v_i)]^{m_i}. \tag{3.5}$$

Consequently for finding a c.d.f. F that maximizes (3.4) subject to $F(x) \leq F_0(x)$, $\forall x$, we can assume that the maximizing F is of discrete type over the interval $(-\infty, v_s]$ and has positive jumps there only at the points $v_1 < v_2 < \ldots < v_s$, with their sum not exceeding $F_0(v_s)$. Note that for $x > v_s$ we are open for F except for the restriction $F(x) \leq F_0(x)$. Thus our task is reduced to finding the values of θ_i's defined by

$$\theta_i = m_i^{-1}[dF(v_i)],\ i = 1, 2, \ldots, s, \tag{3.6}$$

which maximize $\prod_{i=1}^{s}(m_i\theta_i)^{m_i}$ or equivalently

$$\prod_{i=1}^{s}(\theta_i)^{m_i} \tag{3.7}$$

subject to the restrictions

$$\sum_{i=1}^{j} m_i\theta_i \leq F_0(v_j),\ j = 1, 2, \ldots, s. \tag{3.8}$$

Note since w_i's of Theorem 2.3 correspond to the m_i's here, we have conveniently chosen to define θ_i's as in (3.6) (with a factor $1/m_i$) in order to fit the problem to the framework of our Theorem 2.3. It is now easily seen that the above problem further reduces to finding the values of θ_i's which minimize

$$\sum_{i=1}^{s} m_i(-\ln\theta_i) \tag{3.9}$$

subject to the restrictions

$$\begin{cases} \sum_{i=1}^{j} m_i\theta_i \leq F_0(v_j), & j - 1, 2, \ldots, s-1, \\ \sum_{i=1}^{s} m_i\theta_i = F_0(v_s). \end{cases} \tag{3.10}$$

Finally taking $\phi(x) = -\ln x$, $I = (0,1)$, $w_j = m_j$ and $\beta_j = F_0(v_j)$, $j = 1, 2, \ldots, s$, it follows using Theorem 2.3 that the minimizing θ_i's are *uniquely* given by

$$\hat{\theta}_1 = \min_{1 \leq i \leq s} \left[F_0(v_i) / \left(\sum_{r=1}^{i} m_r \right) \right] \tag{3.11}$$

$$\hat{\theta}_j = \min_{j \leq i \leq s} \left[\left(F_0(v_i) - \sum_{r=1}^{j-1} m_r\hat{\theta}_r \right) / \left(\sum_{r=j}^{i} m_r \right) \right], \; j = 2, 3, \ldots, s, \tag{3.12}$$

their uniqueness following from the strict convexity of the $\phi(x)$. From these and the fact that $F_0(v_1) > 0$, it easily follows that $\theta_i > 0$, $\forall i$, as expected. Thus an "M.L.E." $\hat{F}_n$ of F, subject to $F(x) \leq F_0(x)$, $\forall x$, is given by

$$\hat{F}_n(x) = \begin{cases} 0 & x < v_1, \\ \sum_{j=1}^{r} m_j\hat{\theta}_j, & v_r \leq x < v_{r+1}, r = 1, 2, \ldots, s-1, \\ F_0(x) & x \geq v_s, \end{cases} \tag{3.13}$$

where for $x > v_s$ we have conveniently taken its value to be $F_0(x)$.

Let P_n be the probability measure on $(\mathbb{R}, \mathcal{B})$ induced by the c.d.f. $\hat{F}_n$ of (3.13). We now prove, as per Definition 2, that P_n (or equivalently $\hat{F}_n$) is an M.L.E. with respect to the sample values $y_1, y_2, \ldots, y_n$ and the family $\mathcal{P}$ defined by

$$\mathcal{P} = \{P_F : F(x) \leq F_0(x), \; \forall x\}, \tag{3.14}$$

where P_F denotes the probability measure on $(\mathbb{R}, \mathcal{B})$ corresponding to the c.d.f. F. This requires showing equivalent of (3.3) namely

$$\underline{\lim}\; \vartheta_F(N_y)/\vartheta_{\hat{F}_n}(N_y) < 1, \; \forall F \text{ with } P_F \in \mathcal{P} - \{P_n\}_y, \tag{3.15}$$

where

$$\vartheta_F(N_y) = \int_{N_y} \prod_{i=1}^{n} F(du_i). \tag{3.16}$$

Note since $\hat{\theta}_i > 0$, $\forall i$, as $D(N_y) \to 0$, the limit of the integral $\vartheta_{\hat{F}_n}(N_y)$ is always positive. On the other hand, if F is such that it is continuous at v_i for some i, it is easy to see that the limit of $\vartheta_F(N_y)$, as $D(N_y) \to 0$, is zero, so that (3.15) is satisfied for the corresponding P_F. Let us now consider the case of an F with $F(x) \le F_0(x)$, $\forall x$, $F \ne \hat{F}_n$ and having positive jumps $dF(v_i)$, $\forall i$. For this the <u>lim</u> of (3.15) becomes equal to

$$\prod_{i=1}^{s} [dF(v_i)]^{m_i} / \prod_{i=1}^{s} (m_i \hat{\theta}_i)^{m_i} = \prod_{i=1}^{s} (\theta_i)^{m_i} / \prod_{i=1}^{s} (\hat{\theta}_i)^{m_i}, \tag{3.17}$$

which is strictly less than one, since the solution $\{\hat{\theta}_i\}$ to the related minimization problem of (3.9)–(3.10) is unique. This proves (3.15) and hence that P_n (or equivalently $\hat{F}_n$) is an M.L.E. in the sense of Scholz (1980). However note that as an M.L.E. the estimator $\hat{F}_n$ is not unique since, subject to $F(x) \le F_0(x)$, we could have taken other possible values for $\hat{F}_n$ for $x > v_s$.

Finally with the same notations as above, an M.L.E. $\tilde{F}_n$ of F, subject to the restriction $F(x) \ge F_0(x)$, $\forall x$, can be obtained in a similar fashion yielding

$$\tilde{F}_n(x) = \begin{cases} F_0(x) & x < v_1 \\ \sum_{j=1}^{r} m_j \hat{\lambda}_j, & v_r \le x < v_{r+1}, r = 1, 2, \ldots, s-1, \\ 1 & x \ge v_s, \end{cases} \tag{3.18}$$

where

$$\hat{\lambda}_1 = \max\left[\max_{1 \le i \le s-1} \left\{ F_0(v_i) / \sum_{r=1}^{i} m_r \right\}, 1 / \sum_{r=1}^{s} m_r \right], \tag{3.19}$$

$$\hat{\lambda}_j = \max\left[\max_{j \le i \le s-1} \left\{ \left(F_0(v_i) - \sum_{r=1}^{j-1} m_r \hat{\lambda}_r \right) / \sum_{r=j}^{i} m_r \right\}, \left(1 - \sum_{r=1}^{j-1} m_r \hat{\lambda}_r \right) / \sum_{r=j}^{s} m_r \right], \quad j = 2, 3, \ldots, s-1, \tag{3.20}$$

$$\hat{\lambda}_s = \left(1 - \sum_{r=1}^{s-1} m_r \hat{\lambda}_r \right) / m_s. \tag{3.21}$$

Theorem 3.1. *Let $F(x) \le F_0(x) \ \forall\ x$. Then $\hat{F}_n(x)$ defined by (3.13) is a consistent estimator of $F(x)$.*

Proof. Let $D_n = \sup_y |F_n(y) - F(y)|$. We first show that

$$0 \le F_n(x) - \hat{F}_n(x) \le D_n, \quad \forall\ x, \tag{3.22}$$

where F_n is the empirical c.d.f. The inequality (3.22) is trivially seen to be true for $x < v_1$ and $x \ge v_s$. For $v_1 \le x < v_2$, we have

$$F_n(x) - \hat{F}_n(x) = (m_1/n) - m_1\hat{\theta}_1. \tag{3.23}$$

From (3.11) and the fact that $F_0(v_s) \le 1$, it follows that

$$m_1\hat{\theta}_1 \le m_1 F_0(v_s)/n \le m_1/n.$$

Thus

$$F_n(x) - \hat{F}_n(x) \ge 0, \text{ for } v_1 \le x < v_2. \tag{3.24}$$

Also from (3.23) we have for $v_1 \le x < v_2$,

$$\begin{aligned} F_n(x) - \hat{F}_n(x) &= \max_{1\le i\le s}\left[\left(m_1/n\right) - \left(m_1 F_0(v_i)/\sum_{r=1}^{i} m_r\right)\right] \\ &= \max_{1\le i\le s}\left[\left(m_1/\sum_{r=1}^{i} m_r\right)\left(n^{-1}\sum_{r=1}^{i} m_r - F_0(v_i)\right)\right]. \end{aligned}$$

Using $F(x) \le F_0(x) \ \forall\ x$, we have for $v_1 \le x < v_2$

$$\begin{aligned} & F_n(x) - \hat{F}_n(x) \le \max_{1\le i\le s}\left[n^{-1}\sum_{r=1}^{i} m_r - F(v_i)\right] \\ \Longrightarrow & F_n(x) - \hat{F}_n(x) \le D_n, \text{ for } v_1 \le x < v_2. \end{aligned} \tag{3.25}$$

Thus (3.22) is true for $v_1 \le x < v_2$.

Assuming that (3.22) is true for $v_j \le x < v_{j+1}, 1 \le j \le s-2$, it is easily seen that (3.22) is true for $v_{j+1} \le x < v_{j+2}$. Using induction it follows that

$$0 \le F_n(x) - \hat{F}_n(x) \le D_n, \text{ for } v_1 \le x < v_s.$$

Thus (3.22) is true for all x. The consistency of $\hat{F}_n$ follows from (3.22) by using the well known Glivenko Cantelli Theorem. □

Similarly $\tilde{F}_n(x)$ given by (3.18) is also seen to be consistent estimator of F subject to $F(x) \ge F_0(x), \ \forall \ x$.

4. Concluding Remarks

(a) In the above we have considered an example of maximum likelihood estimation under order restrictions to demonstrate the usefulness of our recursive formulas. These formulas can very well be applied to the other problems of statistical interest already considered in literature, for instance see Barlow et. al. (1972) for the (i) estimation of ordered means of k normal distributions (page 98) (ii) estimation of ordered binomial parameters (page 38) (iii) maximum likelihood estimation of two stochastically ordered distributions (page 105) (iv) the geometric extremum problem (page 42) (v) Poisson extremum problem (page 43) (vi) gamma extremum problem (page 45) (vii) Taut string problem (page 50) (viii) the maximum likelihood estimation for distribution with monotone failure rate (page 231), etc. Needless to add that our Theorems 2.1–2.3 may also be useful for optimization problems arising in areas of Operation Research involving convex programming.

(b) Apart from the usefulness of recursive formulas for finding isotonic regressions without the use of any algorithm, they may at times enable one to prove certain theoretical results more easily as seen in Section 3. For a recent similar use of these formulas, the reader may refer to Liang (1989), Liang and Panchapakesan (1989) and Gupta and Liang (1989).

(c) The problem of obtaining the limiting distribution of $\hat{F}_n$ will form a topic of further study and will be reported elsewhere.

Acknowledgement

This research was carried out at the Indian Statistical Institute, New Delhi, where Prem S. Puri spent his leave from Purdue University for the sessions 1986-88. Both the authors are grateful to the Institute and its staff for providing a stimulating atmosphere to carry out this research.

References

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955), An Empirical Distribution Function for Sampling with Incomplete Information. *Annals of Mathematical Statistics*, **26**, 641–647.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972), *Statistical Inference Under Order Restrictions*, New York: John Wiley.

Brunk, H.D. (1955), Maximum Likelihood Estimates of Monotone Parameters, *Annals of Mathematical Statistics*, **26**, 607–616.

Brunk, H.D. (1956), On an Inequality for Convex Functions, *Proceedings of American Mathematical Society*, **7**, 817–824.

Brunk, H.D. (1958), On the Estimation of Parameters Restricted by Inequalities, *Annals of Mathematical Statistics*, **29**, 437–454.

Brunk, H.D., Franck, W.E., Hanson, D.L. and Hogg, R.V. (1966), Maximum Likelihood Estimation of the Distributions of Two Stochastically Ordered Random Variables, *Journal of the American Statistical Association*, **61**, 1067–1080.

Durbin, J. (1973), Distribution Theory for Tests Based on the Sample Distribution Function, *Regional Conference Series in Applied Mathematics*, No. 9, SIAM, Philadelphia.

Dykstra, R.L. (1982), Maximum Likelihood Estimation of the Survival Functions of Stochastically Ordered Random Variables, *Journal of the American Statistical Association*, **77**, 621–628.

Eeden, C. Van (1956), Maximum Likelihood Estimation of Ordered Probabilities, *Indag. Math.*, **18**, 444–455.

Eeden, C. Van (1957), Maximum Likelihood Estimation of Partially Or Completely Ordered Parameters, *Indag. Math.*, **19**, 201–211.

Feltz, C.J. and Dykstra, R.L. (1985), Maximum Likelihood Estimation of the Survival Functions of N Stochastically Ordered Random Variables, *Journal of the American Statistical Association*, **80**, 1012–1019.

Gupta, S.S. and Liang, T. (1989), On the Asymptotic Optimality of Certain Empirical Bayes Simultaneous Testing Procedures, *Technical Report* #89–22C, Dept. of Statistics, Purdue University.

Kruskal, J.B. (1964), Nonmetric Multidimensional Scaling: A Numerical Method, *Psychometrika*, **29**, 115–129.

Liang, T. (1989), Empirical Bayes Selection for the Highest Probability of Success in Negative Binomial Distributions, *Technical Report* #89–18C, Dept. of Statistics, Purdue

University.

Liang, T. and Panchapakesan, S. (1989), On a Monotone Empirical Bayes Test Procedure in a Geometric Model, *I.M.B. Bulletin* **18**, No. 3, Abstract 210–15, p. 247.

Marshal, A.W. and Proschan, F. (1965), Maximum Likelihood Estimation for Distributions with Monotone Failure Rate, *Annals of Mathematical Statistics*, **36**, 69–77.

Robertson, T. and Wright, F.T. (1974), On the Maximum Likelihood Estimation of Stochastically Ordered Random Variates, *Annals of Statistics*, **2**, 528–534.

Robertson, T., Wright, F.T. and Dykstra, R.L. (1988), *Order Restricted Statistical Inference*, New York: John Wiley.

Sampson, A.R. and Whitaker, L.R. (1987), Estimation of Multivariate Distributions Under Stochastic Ordering, *Bulletin of the Institute of Mathematical Statistics*, **16**, 108.

Schoenfeld, D.A. (1986), Confidence Bounds for Normal Means Under Order Restrictions with Applications to Dose Response Curves, Texicology Experiments and Low Dose Extrapolation, *Journal of the American Statistical Association*, **81**, 186–195.

Scholz, F.W. (1980), Towards a Unified Definition of Maximum Likelihood, *The Canadian Journal of Statistics*, **8**, 193–203.