The Annals of Statistics
1998, Vol. 26, No. 5, 1878–1893

MAXIMUM LIKELIHOOD ESTIMATES VIA DUALITY FOR LOG-CONVEX MODELS WHEN CELL PROBABILITIES ARE SUBJECT TO CONVEX CONSTRAINTS

BY HAMMOU EL BARMI AND RICHARD DYKSTRA

Kansas State University and University of Iowa

The purpose of this article is to derive and illustrate a method for fitting models involving both convex and log-convex constraints on the probability vector(s) of a (product) multinomial distribution. We give a two-step algorithm to obtain maximum likelihood estimates of the probability vector(s) and show that it is guaranteed to converge to the true solution. Some examples are discussed which illustrate the procedure.

1. Introduction. Log linear models have proven to be useful and powerful tools for analyzing categorical data and several excellent sources are available to the interested reader [see, e.g., Agresti (1990), Bishop, Fienberg and Holland (1975) and Haberman (1978)]. However, there are many instances when one may wish to combine log-convex constraints with those that place convex constraints directly on the cell probabilities. For example, ROC (receiver operating characteristic) curves are often employed to evaluate the merits of a proposed test for detecting whether a particular condition is present. Thus, if large values of a measure T indicate the condition's presence, $P(T \geq t \mid \text{condition present})$ is referred to as the sensitivity of the test and $1 - P(T \geq t \mid \text{condition absent})$ is its specificity. The graph of $(P(T \leq t \mid \text{condition present}), P(T \leq t \mid \text{condition absent}))$ as t varies results in a nondecreasing curve in the unit square and is referred to as the ROC curve. The closer the curve is to the left and upper boundaries of the unit square, the more appealing the test. If one assumes a discrete setting with $p_i = P(T = t_i \mid \text{condition present})$ and $q_i = P(T = t_i \mid \text{condition absent})$ and independent observations of T for both situations, a natural way to model the measurements is as independent multinomial distributions. The point is, both convexity assumptions such as $\sum_{i=1}^{j} p_i \geq \sum_{i=1}^{j} q_i$ for all j (which implies that the ROC curve never falls below the diagonal line) and log-convexity assumptions such as p_i/q_i nonincreasing in i (which implies that the ROC curve is concave) are naturally of interest. When multiple measures are considered, identifying possible structure is often a key concern. Maximum likelihood estimation and associated tests can often be carried out even when various convex and log-convex assumptions are combined using the approach in this paper.

Received June 1995; revised July 1998.

AMS 1991 *subject classifications*. Primary 62F30; secondary 62G05.

Key words and phrases. Convex constraints, log-convex constraints, maximum likelihood, multinomial, iterative algorithm, I-projection, duality, convex cones.

Haber and Brown (1986) have addressed this type of a problem for the particular situation where $\mathbf{p}$ is a vector of multinomial probabilities. They consider the problem of finding $\mathbf{p}$ when $\ln \mathbf{p}$ is expressible as a linear combination of the column vectors of a known matrix X and $\mathbf{p}$ also satisfies $C^T\mathbf{p} = \mathbf{d}$ for known C and $\mathbf{d}$. Haber and Brown propose a clever iterative procedure based on solving two separate systems of equations which yields vectors that converge to the MLEs under appropriate restrictions.

The essence of their procedure can be described as follows. A vector $\boldsymbol{\lambda}$ satisfying $C\boldsymbol{\lambda} = \mathbf{r}^0 = \mathbf{1}$ is selected as an initial guess for the linear constraint Lagrange multipliers after incorporating the log-subspace constraints into the kernel of the likelihood. Treating $\boldsymbol{\lambda}$ as the correct Lagrange multiplier vector, the subspace parameters $\boldsymbol{\beta}$ are selected to optimize the objective function (Lagrangian) by the method of Darroch and Ratcliff (1972).

For step 2, the estimated values from step 1 are treated as the observed frequencies and plugged into the linear restriction equations. This equation is then solved for the vector $\boldsymbol{\lambda}$ and these values are used as the new estimates of the Lagrange multipliers. This two-step procedure is repeated until adequate convergence is obtained.

Though very ingenious, this procedure will not work in the setting of this paper. It may not be possible to incorporate the log-cone constraints into the likelihood kernel by a finite-dimensional vector. If so, this would imply that the differentiability characterization of the optimal value need not be available. Even if this could be done, it may be impossible to express the convexity constraint Lagrange multipliers as a finite vector.

In this paper, an iterative procedure is proposed which works much more generally than that of Haber and Brown. In particular: (a) The subspace $\{X\boldsymbol{\beta};\ \boldsymbol{\beta} \in R^p\}$ can be replaced by a translated closed, convex cone $\overline{K} + \mathbf{r} = \{\mathbf{x} + \mathbf{r};\ \mathbf{x} \in \overline{K}\}$ which is a subset of the cartesian product of the extended reals $\overline{R}^k$. We do assume that K (those vectors in $\overline{K}$ with all finite coordinates) contains the constant vectors (as did Haber and Brown for their subspace) although they need not belong to $K + \mathbf{r}$ (we say that $\overline{K}$ is closed in the sense that if $\mathbf{v} \in \overline{K}$, there exists $\{\mathbf{v}^{(n)}\} \subset K$ such that $v_i^{(n)} \to v_i$ as $n \to \infty$ for all i). (b) The linear variety $\{\mathbf{p};\ C^T\mathbf{p} = \mathbf{d}\}$ can be replaced by an arbitrary, closed, convex set in $R^k \cap P$ (where P is the set of probability vectors in R^k). Note that since these regions need not be polyhedral, they may correspond to infinitely many linear inequality constraints. (c) Zero cells are allowed in $\hat{\mathbf{p}}$. Since Haber and Brown's results require that $m_t^{(\nu)}$ and $q_t^{(\nu)}$ be bounded away from 0, this essentially rules out zero-count cells. (d) Dual formulations of both cycle steps are given which greatly simplify and illuminate the procedure. An overall dual formulation is given which is especially appealing. However, the dual formulation is not in terms of the conjugate function of (1.1) in any usual sense, and hence Theorem 2.1 is not an application of any type of Fenchel duality theorem of which we are aware. In particular, there is not the one-to-one correspondence between the solutions of the primal and dual problems that is typical in Fenchel dual formulations.

Historically, maximum likelihood estimates (MLEs) under linearity assumptions have been obtained by direct iterative algorithms [Darroch and Ratcliff (1972)] or by iteratively reweighted least squares [Haberman (1974)]. Both of these techniques can be incorporated into the algorithm suggested in this paper if the problem format is of the appropriate form.

We note that the negative of the log-likelihood of a multinomial probability mass function is (except for an additive constant) equal to

$$f(p) = -\sum_{i=1}^{k} \hat{p}_i \ln p_i \tag{1.1}$$

over P, the set of probability vectors in R^k, where $\hat{\mathbf{p}}$ is the probability vector of the observed relative frequencies and $0 \ln 0$ is taken to be 0. We define f to be $+\infty$ elsewhere (0 times the log of a negative number is taken to be $-\infty$) and take multiplication and division of vectors to be done coordinatewise. Of course, minimizing f over C, a closed, convex subset of P, is equivalent to minimizing f over K_C where $K_C = \{\alpha \mathbf{p};\ \alpha \geq 0,\ \mathbf{p} \in C\}$. El Barmi and Dykstra (1994) showed that $\mathbf{p}^*$ minimizes f over C if and only if $\mathbf{y}^* = \hat{\mathbf{p}}/\mathbf{p}^* - \mathbf{1}$ solves

$$\sup_{y \in K_C^*} \sum_{i=1}^{k} \hat{p}_i \ln(1 + y_i) \tag{1.2}$$

(with appropriate definitions for indeterminate forms) where

$$K_C^* = \left\{ \mathbf{y} \in R^k;\ \sum_{i=1}^{k} p_i y_i \leq 0,\ \forall\ \mathbf{p} \in K_C \right\}$$

is the dual cone of K_C. For example, if C should be the region $\{\mathbf{p};\ \sum_{i=1}^{k} r_i p_i = 0\} \cap P$, then $K_C^* = \{\alpha \mathbf{r};\ \alpha \in R\}$ is a one-dimensional region and the optimal value $\hat{\alpha}\mathbf{r}$ that solves (1.2) is obtained from the unique value of α that solves $\sum_{i=1}^{k} \hat{p}_i r_i/(1 + \alpha r_i) = 0$. This value can usually be found very quickly by a Newton–Raphson scheme. In the event that $\leq$ replaces $=$ in the definition of C, the optimal value is $\hat{\alpha}^+ = \max(\hat{\alpha}, 0)$. It is rather amazing that for any closed convex set C contained in P, maximizing $-f(p)$ over C is an equivalent problem (though with a different solution) to maximizing the same function over $K_C^* + \mathbf{1}$.

Dykstra and Lemke (1988) considered the problem of minimizing (1.1) subject to $\ln \mathbf{p} = (\ln p_1, \ln p_2, \ldots, \ln p_k) \in \overline{K}$, a closed, convex cone in $\overline{R}^k$ which contains the constant vectors. In particular, they proved that minimizing (1.1) subject to this constraint has the same solution as the problem

$$\min_{\hat{\mathbf{p}} - \mathbf{p} \in K^*} \sum_{i=1}^{k} p_i \ln p_i, \tag{1.3}$$

where $K^* = \{\mathbf{y} \in R^k;\ \sum_{i=1}^{k} p_i y_i \leq 0,\ \forall\ \mathbf{p} \in \overline{K} \cap R^k\}$ is the dual cone of $\overline{K} \cap R^k$. Note that this is equivalent to an I-projection of the uniform distribution onto $\hat{\mathbf{p}} - K^*$ [Csiszar (1975)]. The optimal vector obtained in K^* is the additive adjustment that must be made to $\hat{\mathbf{p}}$ to obtain the desired solution.

This paper concerns the much more general situation of incorporating both general convexity constraints on the multinomial probability vector and translated cone constraints on the log of the probability vector. Of course, as a corollary, the results of this paper must contain the earlier results of both of the aforementioned papers. However, this paper addresses a much more difficult task than either of the aforementioned problems since now these constraints are simultaneous (note that this is not necessarily a convex programming problem since $\{(\exp(r_1 + x_1), \dots, \exp(r_k + x_k)); \mathbf{x} \in K\}$ need not be convex). It is a rather remarkable fact that the two types of constraints "separate" in the manner discussed in Section 2. The rest of the paper is organized as follows. In Sections 2 and 3, we give the main duality result and describe the algorithm. Section 4 addresses the accuracy and the rate of convergence of the suggested algorithm and Section 5 generalizes when it is applicable. Section 6 discusses some applications.

2. Main duality result. Suppose that the observed vector $(n_1, n_2, \dots, n_k)$ is a realization from a multinomial distribution with parameters n and $\mathbf{p} = (p_1, p_2, \dots, p_k)$. We wish to consider the general problem of finding the MLE of $\mathbf{p}$ subject to the constraint that $\mathbf{p} \in C$ and $\ln \mathbf{p} \in \overline{K} + \mathbf{r}$, where C is a closed, convex subset of P, the set of probability vectors in R^k, and K is a closed, convex cone containing the constant vectors. The basic problem is thus to find a solution of

$$\sup_{\mathbf{p} \in C,\, \ln \mathbf{p} \in \overline{K} + \mathbf{r}} \prod_{i=1}^{k} p_i^{n_i}. \tag{2.1}$$

Although we will assume simple multinomial sampling, one can reduce the case of product multinomial (and often product Poisson sampling with the same type of restrictions on the parameters) to a problem equivalent to (2.1). Interestingly, an equivalent, dual problem formulation of problem (2.1) can be obtained which phrases the constraints in terms of the dual cones K_C^* and K^*. To be specific, we state the following theorem.

THEOREM 2.1. *If* $\mathbf{y}^* > -1$ $(y_i^* > -1, \forall\, i)$ *and* $\mathbf{z}^*$ *minimize*

$$\sum_{i=1}^{k} (\hat{p}_i - z_i) \ln\left(\frac{\hat{p}_i - z_i}{(1 + y_i)\exp r_i} \right) \tag{2.2}$$

subject to $y \in K_C^*$ *and* $z \in K^*$, *then the vector* $\mathbf{p}^*$ *whose ith component is*

$$p_i^* = \frac{\hat{p}_i - z_i^*}{1 + y_i^*}, \qquad i = 1, 2, \dots, k,$$

attains (2.1) *and satisfies the associated constraints.*

PROOF. If $\mathbf{y}^* \in K_C^*$ and $\mathbf{z}^* \in K^*$ minimize (2.2), then by concavity

$$\frac{d}{d\alpha} - \sum_{i=1}^{k} (\hat{p}_i - z_i^* - \alpha(z_i - z_i^*)) \times \ln\left(\frac{\hat{p}_i - z_i^* - \alpha(z_i - z_i^*)}{(1 + y_i^* + \alpha(y_i - y_i^*))\exp r_i}\right)\Bigg|_{\alpha=0} \leq 0 \tag{2.3}$$

for all $\mathbf{y} \in K_C^*$ and $\mathbf{z} \in K^*$. Hence,

$$\sum_{i=1}^{k} (z_i^* - z_i)\ln\left(\frac{\hat{p}_i - z_i^*}{(1 + y_i^*)\exp r_i}\right) + \sum_{i=1}^{k} (z_i - z_i^*) + \sum_{i=1}^{k} \frac{(\hat{p}_i - z_i^*)(y_i - y_i^*)}{(1 + y_i^*)} \leq 0. \tag{2.4}$$

Since $\overline{K}$ contains the constant vectors, we know that $\sum_{i=1}^k z_i = \sum_{i=1}^k z_i^* = 0$ and (2.4) becomes

$$\sum_{i=1}^{k} (z_i - z_i^*)\ln\left(\frac{\hat{p}_i - z_i^*}{(1 + y_i^*)\exp r_i}\right) + \sum_{i=1}^{k} \left(\frac{(\hat{p}_i - z_i^*)(y_i - y_i^*)}{(1 + y_i^*)}\right) \leq 0 \tag{2.5}$$

for all $\mathbf{y} \in K_C^*$ and $\mathbf{z} \in K^*$. If $\mathbf{z} = \mathbf{z}^*$, we obtain

$$\sum_{i=1}^{k} \frac{(\hat{p}_i - z_i^*)(y_i - y_i^*)}{(1 + y_i^*)} \leq 0 \tag{2.6}$$

for all $\mathbf{y} \in K_C^*$ and if $\mathbf{y} = \mathbf{y}^*$ we have

$$\sum_{i=1}^{k} (z_i - z_i^*)\ln\left(\frac{\hat{p}_i - z_i^*}{(1 + y_i^*)(\exp r_i)}\right) \leq 0 \tag{2.7}$$

for all $\mathbf{z} \in K^*$. Moreover, if we take $\mathbf{y} = \mathbf{y}^*/2$ and $\mathbf{y} = 2\mathbf{y}^*$ in (2.6), we obtain

$$\sum_{i=1}^{k} \frac{(\hat{p}_i - z_i^*)y_i^*}{(1 + y_i^*)} = 0 \tag{2.8}$$

and

$$\sum_{i=1}^{k} \frac{(\hat{p}_i - z_i^*)y_i}{(1 + y_i^*)} \leq 0 \tag{2.9}$$

for all $\mathbf{y} \in K_C^*$ which implies that $(\hat{\mathbf{p}} - \mathbf{z}^*)/(1 + \mathbf{y}^*) \in K_C^{**} = K_C$. In a similar manner we can take $\mathbf{z} = \mathbf{z}^*/2$ and $\mathbf{z} = 2\mathbf{z}^*$ in (2.7) and conclude that

$$\sum_{i=1}^{k} z_i^* \ln\left(\frac{\hat{p}_i - z_i^*}{(1 + y_i^*)\exp r_i}\right) = 0 \tag{2.10}$$

and

$$\sum_{i=1}^{k} z_i \ln\left(\frac{\hat{p}_i - z_i^*}{(1 + y_i^*)\exp r_i}\right) \leq 0 \tag{2.11}$$

for all $\mathbf{z} \in K^*$ which implies that

$$\ln\left(\frac{\hat{\mathbf{p}} - \mathbf{z}^*}{(1+\mathbf{y}^*)\exp\mathbf{r}}\right) \in \mathrm{cl}(K^{**}) = \overline{K},$$

or, equivalently,

$$\ln\left(\frac{\hat{\mathbf{p}} - \mathbf{z}^*}{1+\mathbf{y}^*}\right) \in \overline{K} + \mathbf{r}.$$

To conclude that $\mathbf{p}^*$ gives the optimal value, note that if $\mathbf{p} \in C$ is such that $\ln \mathbf{p} \in \overline{K} + \mathbf{r}$, we have

$$\begin{aligned}
-\sum_{i=1}^{k} \hat{p}_i \ln p_i &\geq -\sum_{i=1}^{k} \hat{p}_i \ln p_i + \sum_{i=1}^{k} z_i^* (\ln p_i - r_i) \\
&= -\sum_{i=1}^{k} (\hat{p}_i - z_i^*) \ln p_i - \sum_{i=1}^{k} z_i^* r_i \\
&\geq -\sum_{i=1}^{k} (\hat{p}_i - z_i^*) \ln\left(\frac{\hat{p}_i - z_i^*}{1+y_i^*}\right) \\
&\quad - \sum_{i=1}^{k} (1+y_i^*)\left(p_i - \frac{\hat{p}_i - z_i^*}{1+y_i^*}\right) - \sum_{i=1}^{k} z_i^* r_i
\end{aligned}$$

by expanding $\ln p_i$ about p_i^*.

However,

$$\sum_{i=1}^{k} (1+y_i^*)\left(p_i - \frac{\hat{p}_i - z_i^*}{1+y_i^*}\right) = \sum_{i=1}^{k} p_i y_i^* \leq 0$$

by (2.8) and the fact that

$$\sum_{i=1}^{k} \frac{\hat{p}_i - z_i^*}{1+y_i^*} = 1.$$

Hence,

$$\begin{aligned}
-\sum_{i=1}^{k} \hat{p}_i \ln p_i &\geq -\sum_{i=1}^{k} (\hat{p}_i - z_i^*) \ln\left(\frac{\hat{p}_i - z_i^*}{1+y_i^*}\right) - \sum_{i=1}^{k} z_i^* r_i \\
&= -\sum_{i=1}^{k} \hat{p}_i \ln\left(\frac{\hat{p}_i - z_i^*}{1+y_i^*}\right) + \sum_{i=1}^{k} z_i^* \ln\left(\frac{\hat{p}_i - z_i^*}{(1+y_i^*)\exp r_i}\right) \\
&= -\sum_{i=1}^{k} \hat{p}_i \ln\left(\frac{\hat{p}_i - z_i^*}{1+y_i^*}\right),
\end{aligned}$$

where the final equality holds by (2.10). The desired conclusion now easily follows. □

3. Algorithm. We now present an iterative procedure to minimize (2.2) subject to $\mathbf{y} \in K_C^*$ and $\mathbf{z} \in K^*$. This procedure is guaranteed to converge to the true solution and encompasses the algorithm given in Haber and Brown (1986).

Each cycle of the algorithm consists of two steps; the first associated with the constraint set $\ln \mathbf{p} \in \overline{K} + \mathbf{r}$ and the second with the constraint set $\mathbf{p} \in C$. We initiate the algorithm by setting $\mathbf{q}^{(0)} = \hat{\mathbf{p}}$, $\mathbf{z}^{(0)} = \mathbf{0}$, $\mathbf{y}^{(0)} = \mathbf{0}$ and $n = 1$. At the nth cycle, step 1, we calculate the vector $\mathbf{q}^{(n)}$ which solves

$$\min_{\hat{\mathbf{p}} - \mathbf{q} \in K^*} \sum_{i=1}^{k} q_i \ln\left(\frac{q_i}{\left(1 + y_i^{(n-1)}\right) \exp r_i} \right)$$

and then set $\mathbf{z}^{(n)} = \hat{\mathrm{p}} - \mathbf{q}^{(n)}$. The vector $\mathbf{q}^{(n)}$ is the I-projection of the (properly scaled) probability vector $(1 + \mathbf{y}^{(n-1)}) \exp \mathbf{r}$ onto the set of probability vectors contained in $-K^* + \hat{\mathbf{p}}$ [Csiszar (1975)]. This is equivalent to finding the multinomial MLE of $\mathbf{p}$, subject to the constraint that $\ln \mathbf{p} \in \overline{K} + \mathbf{r} + \ln(1 + \mathbf{y}^{(n-1)})$.

The second step of the algorithm (in the nth cycle) amounts to deriving the MLE of $\mathbf{p}$, say $\mathbf{p}^{(n)}$, subject to $\mathbf{p} \in C$ where $\mathbf{q}^{(n)}$ serves as the vector of observed relative frequencies [El Barmi and Dykstra (1994)]. Equivalently, we let $\mathbf{y}^{(n)}$ denote the solution to the problem

$$\max_{\mathbf{y} \in K_C^*} \sum_{i=1}^{n} q_i^{(n)} \ln(1 + y_i)$$

and then set the nth cycle estimate of $\mathbf{p}$ to be the vector $\mathbf{p}^{(n)}$ whose ith component is given by

$$p_i^{(n)} = \frac{q_i^{(n)}}{1 + y_i^{(n)}} = \frac{\hat{p}_i - z_i^{(n)}}{1 + y_i^{(n)}}, \qquad i = 1, \dots, k.$$

This two-step procedure is repeated for $n = 1, 2, \dots$ until sufficient accuracy is attained. (Additional discussion on accuracy will be given in the following section.) Note that we only need to solve problems of the form $\sup_{\mathbf{p} \in C} \sum_{i=1}^{k} \hat{p}_i \ln p_i$ (with updated values for $\hat{\mathbf{p}}$) or problems of the form $\sup_{\ln \mathbf{p} \in \overline{K} + \mathbf{r}} \sum_{i=1}^{k} \hat{p}_i \ln p_i$ (with updated values for $\mathbf{r}$).

One would hope that $(\mathbf{z}^{(n)}, \mathbf{y}^{(n)})$ would converge to $(\mathbf{z}^*, \mathbf{y}^*)$ (implying that $\mathbf{p}^{(n)}$ converges to $\mathbf{p}^*$) and this is indeed the case. The nature of that convergence is the essence of Theorem 3.2. We now state and prove a few lemmas preliminary to that theorem. The proofs are given in the Appendix. We assume that there exists at least one $\mathbf{p} \in C$ such that $\ln \mathbf{p} \in \overline{K} + \mathbf{r}$.

LEMMA 3.1. *After the nth cycle of the algorithm described previously, we have*

$$\sum_{i=1}^{k} q_i^{(n)} = 1$$

and

$$\sum_{i=1}^{k} p_i^{(n)} = 1.$$

LEMMA 3.2. *The sequence* $\sum_{i=1}^{k} q_i^{(n)} \ln(p_i^{(n)}/\exp r_i)$ *is monotone nonincreasing and bounded below.*

THEOREM 3.2. *If there exists* $\mathbf{p} \in C$ *such that* $\ln \mathbf{p} \in \overline{K} + \mathbf{r}$, *then the sequences* $\mathbf{q}^{(n)}$ *and* $\mathbf{p}^{(n)}$ *converge. Moreover*, $\lim_{n\to\infty} \mathbf{p}^{(n)} = \mathbf{p}^{\infty}$ *is the unique solution to* (2.1).

4. Accuracy. One of the negative aspects of iterative procedures for computing MLEs is the difficulty in ascertaining accuracy of the estimates. Even though the rate of change over subsequent cycles may be small, close proximity to the true MLEs may not be the case. The duality approach, however, often allows additional statements concerning accuracy.

Suppose we define the function

$$\tilde{f}(\mathbf{z}, \mathbf{y}) = \sum_{i=1}^{k} (\hat{p}_i - z_i) \ln\left(\frac{\hat{p}_i - z_i}{(1 + y_i)\exp r_i} \right)$$

and also suppose $\tilde{\mathbf{p}} \in C$ and $\ln \tilde{\mathbf{p}} \in \overline{K} + \mathbf{r}$ [where $\tilde{\mathbf{p}}$ is near $\mathbf{p}^{(n)} = ((\hat{\mathbf{p}} - \mathbf{z}^{(n)})/(1 + \mathbf{y}^{(n)}))$]. If we let

$$\varepsilon = \tilde{f}(\mathbf{z}^{(n)}, \mathbf{y}^{(n)}) - \sum_{i=1}^{k} \hat{p}_i \ln\left(\frac{\tilde{p}_i}{\exp r_i} \right),$$

then

$$\tilde{f}(\mathbf{z}^{(n)}, \mathbf{y}^{(n)}) \geq \inf_{\mathbf{z}\in K^*,\, \mathbf{y}\in K_C^*} \tilde{f}(\mathbf{z}, \mathbf{y}) = \sup_{\mathbf{p}\in C,\, \ln \mathbf{p} \in \overline{K}+\mathbf{r}} \sum_{i=1}^{k} \hat{p}_i \ln\left(\frac{p_i}{\exp r_i} \right),$$

or, equivalently,

$$\begin{aligned} \varepsilon &\geq \sup_{\mathbf{p}\in C,\, \ln \mathbf{p}\in \overline{K}+\mathbf{r}} \sum_{i=1}^{k} \hat{p}_i \ln p_i - \sum_{i=1}^{k} \hat{p}_i \ln \tilde{p}_i \\ &= \sum_{i=1}^{k} \hat{p}_i \ln p_i^* - \sum_{i=1}^{k} \hat{p}_i \ln \tilde{p}_i, \end{aligned}$$

where $\mathbf{p}^*$ denotes the solution to (2.1). Expanding the second part in a Taylor series about $\mathbf{p}^*$ gives

$$\varepsilon \geq -\sum_{i=1}^{k} \frac{\hat{p}_i(\tilde{p}_i - p_i^*)}{p_i^*} + \frac{1}{2} \sum_{i=1}^{k} \hat{p}_i \frac{(\tilde{p}_i - p_i^*)^2}{\beta_i^2}, \tag{4.1}$$

where β_i lies between $\tilde{p}_i$ and p_i^*. Although the constraint region for (2.1) need not always be a convex set, it often is (e.g., if K is an isotonic cone). In this case, $(1-\alpha)\mathbf{p}^* + \alpha\tilde{\mathbf{p}}$ lies within the constraint region and hence

$$\frac{d}{d\alpha}\sum_{i=1}^{k}\hat{p}_i \ln\big(p_i^* + \alpha(\tilde{p}_i - p_i^*)\big)\bigg|_{\alpha=0} = \sum_{i=1}^{k}\frac{\hat{p}_i}{p_i^*}(\tilde{p}_i - p_i^*) \le 0, \tag{4.2}$$

so that

$$2\varepsilon \ge \sum_{i=1}^{k}\hat{p}_i\left(\frac{\tilde{p}_i \wedge p_i^*}{\tilde{p}_i \vee p_i^*} - 1\right)^2$$

(where "$\vee$" denotes sup and "$\wedge$" denotes inf). Thus, the region

$$S = \left\{\mathbf{p}\in P;\ \left\|\frac{\tilde{\mathbf{p}}\wedge\mathbf{p}}{\tilde{\mathbf{p}}\vee\mathbf{p}} - 1\right\|_{\hat{\mathbf{p}}}^2 \le 2\varepsilon\right\} \tag{4.3}$$

must contain the solution to (2.1) where $\|\cdot\|_{\hat{\mathbf{p}}}$ denotes the L_2 norm with weights $\hat{\mathbf{p}}$.

Even when the preceding convexity assumption does not hold, the value of (4.2) is typically small and S will generally contain the true solution. Moreover, since the β_i in (4.1) are generally very close to $\tilde{p}_i$, S is generally very similar to the region

$$S^* = \left\{\mathbf{p}\in P;\ \|\mathbf{p}-\tilde{\mathbf{p}}\|_{\hat{\mathbf{p}}/\tilde{\mathbf{p}}^2}^2 \le 2\varepsilon\right\}. \tag{4.4}$$

The rate of convergence for the algorithm described here will vary substantially depending on the nature of the constraint regions. However, this algorithm is a cyclic descent algorithm similar in many aspects to the least squares projection method known as Von Neumann's algorithm [also referred to as alternating conditional expectations (ACE)] as well as Boyle and Dykstra's (1985) generalization to arbitrary, convex sets. There is a long history of interesting applications of this type of algorithm [see Deutsch (1983, 1992)].

Recent elegant work by Deutsch (1995) and Deutsch and Hundal (1994) has established geometric convergence rates for Boyle and Dykstra's procedure in a wide variety of settings. That is, $\|\mathbf{x}_n - \lim_{n\to\infty}\mathbf{x}_n\| \le \rho c^n$, for all n, for fixed constant ρ and $c < 1$ where $\mathbf{x}_n$ denotes the estimate at the nth step of the algorithm and $\|\cdot\|$ is a least squares norm. The constants ρ and c depend only on the constraint sets and not on the initial starting points.

Since the algorithm in this paper is similar in terms of its basic approach, we would conjecture that its convergence properties are similar to those derived by Deutsch and Hundal in the least squares situation.

5. Intersection of convex regions. The duality correspondence is especially useful when the constraint regions can be expressed as the intersection of a finite number of "nice" convex regions. To be specific, suppose we want to minimize (1.1) subject to $\mathbf{p} \in \bigcap_{j=1}^{s} C_j$ and $\ln \mathbf{p} \in \bigcap_{j=1}^{t} \overline{K}_j + \mathbf{r}$, where C_j is a closed, convex subset of P, $\overline{K}_j$ is a closed, convex cone containing the constant vectors for all j and $\mathbf{r}$ is a fixed, known vector. Define K_C as in Section 2 to be the cone corresponding to $C = \bigcap_{j=1}^{s} C_j$. We assume that $\bigcap_{j=1}^{s} \mathrm{ri}(K_{C_j})$ and $\bigcap_{j=1}^{t} \mathrm{ri}(K_j)$ are nonempty (where ri denotes the "relative interior") to avoid problems with closure [see Rockafellar (1970), page 146]. We have the following theorem.

THEOREM 5.1. *If* $(\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_s^*) \in K_{C_1}^* \times K_{C_2}^* \times \cdots \times K_{C_s}^*$ *and* $(\mathbf{z}_1^*, \mathbf{z}_2^*, \dots, \mathbf{z}_t^*) \in K_1^* \times K_2^* \times \cdots \times K_t^*$ *solve*

$$\min \sum_{i=1}^{k} (\hat{p}_i - z_{1i} - \cdots - z_{ti}) \ln\left(\frac{\hat{p}_i - z_{1i} - \cdots - z_{ti}}{(1 + y_{1i} + \cdots + y_{si}) \exp r_i} \right) \tag{5.1}$$

subject to $\mathbf{y}_j \in K_{C_j}^*$, $j = 1, 2, \dots, s$, $\mathbf{z}_j \in K_j^*$, $j = 1, 2, \dots, t$, *and* $1 + \mathbf{y}_1 + \mathbf{y}_2 + \cdots + \mathbf{y}_s > \mathbf{0}$, *then the vector* $\mathbf{p}^*$ *whose ith component is*

$$p_i^* = \frac{\hat{p}_i - z_{1i}^* - \cdots - z_{ti}^*}{1 + y_{1i}^* + y_{2i}^* + \cdots + y_{si}^*}$$

minimizes (1.1) *subject to* $\mathbf{p} \in \bigcap_{j=1}^{s} C_j$ *and* $\ln \mathbf{p} \in \bigcap_{j=1}^{t} \overline{K}_j + r$.

PROOF. This result follows from Theorem 2.1 and the fact that

$$\left(\bigcap_{j=1}^{s} K_{C_j} \right)^* = K_{C_1}^* + \cdots + K_{C_s}^*$$

and

$$\left(\bigcap_{j=1}^{t} K_j \right)^* = K_1^* + \cdots + K_t^*.$$

As one would expect, it is possible to extend the algorithm in Section 3 to cyclically optimize over one $\mathbf{z}_j$ (or $\mathbf{y}_j$) at a time and still establish correct, monotone convergence. However, we forbear from imposing the messy details on the reader.

6. Examples.

EXAMPLE 1 (Nonnegative local log odds ratios and marginal homogeneity). Odds ratios are commonly used to measure association between ordinal categorical variables. It is well known that independence is equivalent to all local odds ratios being equal to 1. When all the local odds ratios are at least 1 (at most 1), the variables are said to be positively (negatively) associated. We consider the estimation of the cell probabilities of a $k \times k$ contingency table when, in addition to being positively associated, the variables are equal in their marginal distributions (i.e., marginal homogeneity holds).

If $K_{ij} = \{\mathbf{x};\ x_{ij} + x_{i+1,j+1} - x_{i+1,j} - x_{i,j+1} \geq 0\}$ for $i, j = 1, \dots, k-1$ and $L_m = \{\mathbf{x};\ \sum_{j=1}^k x_{mj} - \sum_{j=1}^k x_{jm} = x_{m+} - x_{+m} = 0\}$ for $m = 1, \dots, k-1$, then the problem we wish to solve can be expressed as follows:

$$\max \prod_{i=1}^{k} \prod_{j=1}^{k} p_{ij}^{n_{ij}}$$

subject to

$$\mathbf{p} \in \bigcap_{m=1}^{k-1} L_m,$$

$$\ln \mathbf{p} \in \bigcap_{j=1}^{k-1} \bigcap_{i=1}^{k-1} K_{ij}$$

and $\mathbf{p} \in P$. Note that

$$\left(\bigcap_{m=1}^{k-1} L_m \right)^* = L_1^\perp + \cdots + L_{k-1}^\perp,$$

where $L_m^\perp$ is the orthogonal complement (subspace) of L_m and is given by

$$L_m^\perp = \Big\{ \alpha \mathbf{y}^{(m)};\ y_{mj}^{(m)} = -y_{jm}^{(m)} = 1, \qquad j = 1, \dots, k, \\ y_{ij}^{(m)} = 0, \text{ otherwise}, \ \alpha \in R \Big\}.$$

Dykstra and Lemke (1988) showed that if

$$A_{ij} = \{(l, m);\ 1 \leq l \leq i,\ 1 \leq m \leq j\}$$

and

$$H_{ij} = \begin{cases} \left\{ \mathbf{x};\ \sum_{A_{ij}} x_{lm} \leq 0 \right\}, & \text{if } 1 \leq i \leq k-1,\ 1 \leq j \leq k-1, \\ \left\{ \mathbf{x};\ \sum_{A_{ij}} x_{lm} = 0 \right\}, & \text{if } i = k \text{ or } j = k, \end{cases}$$

then

$$\left(\bigcap_{j=1}^{k-1} \bigcap_{i=1}^{k-1} K_{ij} \right)^* = \bigcap_{j=1}^{k} \bigcap_{i=1}^{k} H_{ij}.$$

The dual problem can then be expressed as

$$\min \sum_{i=1}^{k} \sum_{j=1}^{k} (\hat{p}_{ij} - z_{ij}) \ln \left(\frac{\hat{p}_{ij} - z_{ij}}{1 + \sum_{m=1}^{k-1} \alpha_l y_{ij}^{(m)}} \right)$$

subject to $\alpha_l \in \mathbb{R}$ for $l = 1, \dots, k-1$ and

$$\mathbf{z} \in \bigcap_{j=1}^{k} \bigcap_{i=1}^{k} H_{ij}.$$

Therefore, the algorithm developed in El Barmi and Dykstra (1994) can be used in the second step of the algorithm to compute the α's while Dykstra's (1985) iterative algorithm for I-projections can be used in the first step to compute the optimal $\mathbf{z}$. As noted in Robertson, Wright and Dykstra (1988), the individual projections needed for using Dykstra's algorithm can be easily programmed since the solution for an I-projection problem of the form

$$\inf_{\substack{p \in P \\ \sum_{j \in A} p_j \le b}} \sum_{i=1}^{k} p_i \ln\left(\frac{p_i}{r_i}\right)$$

is given by

$$p_i^* = \begin{cases} \dfrac{r_i}{\sum_j r_j}, & \text{if } \dfrac{\sum_{j \in A} r_j}{\sum_j r_j} \le b, \\ \dfrac{r_i b}{\sum_{j \in A} r_j}, & \text{if } i \in A \text{ and } \dfrac{\sum_{j \in A} r_j}{\sum_j r_j} > b, \\ \dfrac{r_i(1-b)}{\sum_{j \in A^c} r_j} & \text{if } i \in A^c \text{ and } \dfrac{\sum_{j \in A} r_j}{\sum_j r_j} > b. \end{cases}$$

EXAMPLE 2. Next we consider problem (2.1) when K is an isotonic cone. At the first step of the nth iteration of the proposed algorithm, we need to solve the following problem:

$$\inf_{\mathbf{z} \in K^*} \sum_{i=1}^{k} (\hat{p}_i - z_i) \ln\left(\frac{\hat{p}_i - z_i}{(1 + y_i^{(n-1)}) \exp r_i}\right).$$

If we let $\mathbf{v} = \hat{\mathbf{p}} - \mathbf{z}$, this problem is equivalent to minimizing

$$\sum_{i=1}^{k} v_i \ln\left(\frac{v_i}{(1 + y_i^{(n-1)}) \exp r_i}\right)$$

subject to

$$\frac{\hat{\mathbf{p}}}{(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}} - \frac{\mathbf{v}}{(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}} \in K^{*(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}},$$

where $K^{*(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}}$ is the dual cone of K when the inner product is $(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{k} u_i v_i (1 + y_i^{(n-1)}) \exp r_i$. By Corollary A of Theorem 1.7.4 in Robertson, Wright and Dykstra (1988), the solution $\mathbf{v}^*$ is given by

$$\mathbf{v}^* = (1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}} E_{(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}}\left(\frac{\hat{\mathbf{p}}}{(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}} | K\right),$$

where $E_{(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}}(\hat{\mathbf{p}}/(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}} | K)$ is the least squares projection of $\hat{\mathbf{p}}/(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}$ onto K with weights $(1 + \mathbf{y}^{(n-1)}) e^{\mathbf{r}}$. This least squares projection can be computed using one of the algorithms described in Robertson, Wright and Dykstra (1988) such as the pool adjacent violators algorithm (PAVA). The value of $\mathbf{z}^{(n)}$ is then given by $\hat{\mathbf{p}} - \mathbf{v}^*$. In the second step of the nth iteration, we need to find $\mathbf{y}^{(n)}$ which solves

$$\sup_{\mathbf{y} \in K_C^*} \sum_{i=1}^{k} \left(\hat{p}_i - z_i^{(n)} \right) \ln(1 + y_i).$$

The nth cycle estimate of $\mathbf{p}$ is then the vector $\mathbf{p}^{(n)}$ whose ith component is given by

$$p_i^{(n)} = \frac{\left(1 + y_i^{(n-1)}\right) \exp r_i E_{(1+\mathbf{y}^{(n-1)}) \exp \mathbf{r}} \left(\hat{\mathbf{p}} / (1 + \mathbf{y}^{(n-1)}) \exp \mathbf{r} | K \right)_i}{1 + y_i^{(n)}}.$$

It then follows from the continuity of the projection operator that the solution $\mathbf{p}^* = \lim_{n \to \infty} \mathbf{p}^{(n)}$ is given by the vector whose ith component is

$$p_i^* = \exp r_i E_{(1+\mathbf{y}^*) \exp \mathbf{r}} \left(\frac{\hat{\mathbf{p}}}{(1 + \mathbf{y}^*) \exp \mathbf{r}} | K \right)_i,$$

$i = 1, 2, \ldots, k$, where $\mathbf{y}^* = \lim_{n \to \infty} \mathbf{y}^{(n)}$. Finally, we note that since order constraints on the components of $\mathbf{p}$ are preserved under the log transformation, one can improve the computational efficiency of the algorithm by incorporating these types of constraints as log-convex constraints in the algorithm.

EXAMPLE 3. Stochastic orderings of distributions are an important concept in statistics and applied probability. Many different types of stochastic orderings have been defined in the literature and for an extensive discussion of this and other related issues, see Shaked and Shanthikumar (1994). If F and G are two distribution functions which possess densities (or probability mass functions) f and g respectively, then F is greater than G in the likelihood ratio ordering sense if $f(x)/g(x)$ is nondecreasing in x. Dykstra, Kochar and Robertson (1995) obtain maximum likelihood estimates of the probability vectors corresponding to two likelihood ratio ordered multinomials. For more than two multinomials which are likelihood ratio ordered, maximum likelihood estimates do not exist in a closed form. However, the algorithm described in this paper can be applied to solve this problem.

APPENDIX

PROOF OF LEMMA 3.1. Since K contains constant vectors and $\hat{\mathbf{p}} - \mathbf{q}^{(n)} \in K^*$, we know

$$\sum_{i=1}^{k} \left(\hat{p}_i - q_i^{(n)} \right) = 0,$$

and hence

$$\sum_{i=1}^{k} q_i^{(n)} = \sum_{i=1}^{k} \hat{p}_i = 1.$$

Since $\mathbf{y}^{(n)}$ solves (2.8) (with $\mathbf{z}^*$ replaced by $\mathbf{z}^{(n)}$), we know

$$\sum_{i=1}^{k} \frac{q_i^{(n)} y_i^{(n)}}{1 + y_i^{(n)}} = 0,$$

and hence

$$\sum_{i=1}^{k} p_i^{(n)} = \sum_{i=1}^{k} \frac{q_i^{(n)}}{1 + y_i^{(n)}} = \sum_{i=1}^{k} \frac{q_i^{(n)}\left(1 + y_i^{(n)}\right)}{1 + y_i^{(n)}} = \sum_{i=1}^{k} q_i^{(n)} = 1. \qquad \square$$

PROOF OF LEMMA 3.2. We have

$$\begin{aligned}
\sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{p_i^{(n)}}{\exp r_i}\right) &= \sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{q_i^{(n)}}{\exp r_i}\right) - \sum_{i=1}^{k} q_i^{(n)} \ln\left(1 + y_i^{(n)}\right) \\
&= \sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{q_i^{(n)}}{\left(1 + y_i^{(n-1)}\right)\exp r_i}\right) \\
&\qquad - \sum_{i=1}^{k} q_i^{(n)} \ln\left(1 + y_i^{(n)}\right) + \sum_{i=1}^{k} q_i^{(n)} \ln\left(1 + y_i^{(n-1)}\right) \\
&\le \sum_{i=1}^{k} q_i^{(n-1)} \ln\left(\frac{q_i^{(n-1)}}{\left(1 + y_i^{(n-1)}\right)\exp r_i)}\right) \\
&= \sum_{i=1}^{k} q_i^{(n-1)} \ln\left(\frac{p_i^{(n-1)}}{\exp r_i}\right),
\end{aligned}$$

since

$$\sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{q_i^{(n)}}{1 + y_i^{(n-1)}}\right) \le \sum_{i=1}^{k} q_i^{(n-1)} \ln\left(\frac{q_i^{(n-1)}}{1 + y_i^{(n-1)}}\right)$$

and

$$\sum_{i=1}^{k} q_i^{(n)} \ln\left(1 + y_i^{(n)}\right) \ge \sum_{i=1}^{k} q_i^{(n)} \ln\left(1 + y_i^{(n-1)}\right).$$

Therefore, we have the first conclusion of the lemma. We now note [El Barmi and Dykstra (1994)] that $\mathbf{p}^{(n)}$ solves

$$\max_{p \in C} \sum_{i=1}^{k} q_i^{(n)} \ln p_i.$$

Therefore, if $\tilde{\mathbf{p}} \in C$ and $\ln \tilde{\mathbf{p}} \in \overline{K} + \mathbf{r}$, we have

$$\sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{p_i^{(n)}}{\exp r_i}\right) \ge \sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{\tilde{p}_i}{\exp r_i}\right).$$

In addition, since $\hat{\mathbf{p}} - \mathbf{q}^{(n)} \in K^*$ and $\ln \tilde{\mathbf{p}} \in \overline{K} + \mathbf{r}$ we know

$$\sum_{i=1}^{k} \left(\hat{p}_i - q_i^{(n)} \right) \ln\left(\frac{\tilde{p}_i}{\exp r_i} \right) \le 0,$$

and hence

$$\text{(A.1)} \qquad \sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{p_i^{(n)}}{\exp r_i} \right) \ge \sum_{i=1}^{k} q_i^{(n)} \ln\left(\frac{\tilde{p}_i}{\exp r_i} \right) \ge \sum_{i=1}^{k} \hat{p}_i \ln\left(\frac{\tilde{p}_i}{\exp r_i} \right),$$

which is the desired conclusion. □

Proof of Theorem 3.2. Let $\{n_j\}_{j=1}^{\infty}$ be a subsequence of the positive integers. Then there must exist a sub-subsequence $\{n'_j\}_{j=1}^{\infty}$ such that $\lim_{j\to\infty} \mathbf{q}^{(n'_j)} = \mathbf{q}^{\infty}$ and $\lim_{j\to\infty} \mathbf{p}^{(n'_j)} = \mathbf{p}^{\infty}$ exist. From (A.1),

$$\begin{aligned}
\sup_{\mathbf{p}\in C,\, \ln \mathbf{p} \in \overline{K}+\mathbf{r}} \sum_{i=1} \hat{p}_i \ln\left(\frac{p_i}{\exp r_i} \right) &\le \lim_{j\to\infty} \sum_{i=1}^{k} q_i^{(n'_j)} \ln\left(\frac{p_i^{(n'_j)}}{\exp r_i} \right) \\
&= \lim_{j\to\infty} \sum_{i=1}^{k} \left(\hat{p}_i - z_i^{(n'_j)} \right) \ln\left(\frac{p_i^{(n'_j)}}{\exp r_i} \right) \\
&= \sum_{i=1}^{k} \hat{p}_i \ln\left(\frac{p_i^{\infty}}{\exp r_i} \right).
\end{aligned}$$

Moreover, from arguments similar to those used in deriving (2.9) and (2.11),

$$\sum_{i=1}^{k} y_i p_i^{\infty} \le 0 \quad \forall\, \mathbf{y} \in K_C^*$$

and

$$\sum_{i=1}^{k} z_i \ln\left(\frac{p_i^{\infty}}{\exp r_i} \right) \le 0 \quad \forall\, \mathbf{z} \in K^*,$$

which implies that $\mathbf{p}^{\infty} \in K_C$ and $\ln \mathbf{p}^{\infty} \in \overline{K} + \mathbf{r}$. However, $\sum_{i=1}^{k} p_i^{\infty} = 1$, so that $\mathbf{p}^{\infty} \in C$, and since the solution to (2.1) is unique, the entire sequence $\{\mathbf{p}^{(n)}\}_{n=1}^{\infty}$ must converge to $\mathbf{p}^{\infty}$. □

Acknowledgments. We are grateful to the Editor, Associate Editor and the referees for their thoughtful comments and suggestions which have greatly improved the presentation of this paper.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.

Bishop, Y. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.

Boyle, J. P. and Dykstra, R. L. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. In *Advances in Order Restricted Statistical Inference*. *Lecture Notes in Statist*. **37** 28–47. Springer, New York.

CSISZAR, I. (1975). *I*-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* **3** 146–159.

DARROCH, J. N. and RATCLIFF, D. (1972). Generalized iterative scaling for loglinear models. *Ann. Math. Statist.* **43** 1470–1480.

DEUTSCH, F. (1983). Applications of Von Neumann's alternating projections algorithm. In *Mathematical Methods in Operations Research* (P. Kenderov, ed.) 44–51. Sofia, Bulgaria.

DEUTSCH, F. (1992). Dykstra's cyclic projections algorithm: the rate of convergence. In *Recent Development in Approximation Theory, Wavelets and Applications. NATO ASI Conference* (S. P. Singh, ed.) 87–94.

DEUTSCH, F. (1995). The method of alternating orthogonal projections. In *Approximation Theory, Spline Functions and Applications* (S. P. Singh, ed.) 44–51. Sofia, Bulgaria.

DEUTSCH, F. and HUNDAL, H. (1994). The rate of convergence of Dykstra's cyclic projections algorithm: the polyhedral case. *Numer. Anal. Optim.* **15** 537–565.

DYKSTRA, R. L. (1985). An iterative procedure for obtaining *I*-projections onto the intersection of convex sets. *Ann. Probab.* **13** 975–984.

DYKSTRA, R. L., KOCHAR, S. and ROBERTSON, T. (1995). Inference for likelihood ratio ordering in the two-sample problem. *J. Amer. Statist. Assoc.* **90** 1039–1040.

DYKSTRA, R. L. and LEMKE, J. H. (1988). Duality of *I*-projections and maximum-likelihood estimation for loglinear models under cone constraints. *J. Amer. Statist. Assoc.* **83** 546–554.

EL BARMI, H. and DYKSTRA, R. L. (1994). Restricted multinomial MLE's based upon Fenchel duality. *Statist. Probab. Lett.* **21** 121–130.

HABER, M. and BROWN, M. B. (1986). Maximum likelihood methods for loglinear models when expected frequencies are subject to linear constraints. *J. Amer. Statist. Assoc.* **81** 477–482.

HABERMAN, S. (1974). *The Analysis of Frequency Data*. Univ. Chicago Press.

HABERMAN, S. (1978). *Analysis of Qualitative Data*. Academic Press, New York.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.

SHAKED, M. and SHANTHIKUMAR, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, San Diego.

DEPARTMENT OF STATISTICS
KANSAS STATE UNIVERSITY
MANHATTAN, KANSAS 66506
E-MAIL: barmi@stat.ksu.edu

DEPARTMENT OF STATISTICS AND
ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242-1419
E-MAIL: dykstra@stat.uiowa.edu