The Annals of Probability
1985, Vol. 13, No. 3, 975–984

AN ITERATIVE PROCEDURE FOR OBTAINING *I*-PROJECTIONS ONTO THE INTERSECTION OF CONVEX SETS[1]

BY RICHARD L. DYKSTRA

The University of Iowa

A frequently occurring problem is to find a probability distribution lying within a set $\mathscr{E}$ which minimizes the I-divergence between it and a given distribution R. This is referred to as the I-projection of R onto $\mathscr{E}$. Csiszar (1975) has shown that when $\mathscr{E} = \cap_1^t \mathscr{E}_i$ is a finite intersection of closed, linear sets, a cyclic, iterative procedure which projects onto the individual $\mathscr{E}_i$ must converge to the desired I-projection on $\mathscr{E}$, provided the sample space is finite.

Here we propose an iterative procedure, which requires only that the $\mathscr{E}_i$ be convex (and not necessarily linear), which under general conditions will converge to the desired I-projection of R onto $\cap_1^t \mathscr{E}_i$.

1. Introduction. Suppose p and q are probability measures defined on subsets of the finite set $\mathscr{X}$, which without loss of generality we take to be the first m positive integers. The I-divergence of p with respect to q, also called the Kullback-Liebler information number, cross-entropy between p and q, information for discrimination, entropy of p relative to q, etc., is given by

$$I(p \| q) = \sum_{k=1}^m p(k)\ln(p(k)/q(k)) \tag{1.1}$$

where $p(k)$ is the mass placed by the probability measure p at the point k. We let P denote the set of all probability measures on $\mathscr{X}$, and use the convention that products (or quotients) are to be interpreted as pointwise multiplication (or division). For example, $s_{12} = rp_{11}/s_{11}$ means that the measure s_{12} puts mass $r(k)p_{11}(k)/s_{11}(k)$ at k.

We mention that $I(p \| q)$ is defined analogously for general probability measures on infinite spaces, but for simplicity, we will only consider finite sets. (See Kullback, 1959, or Csiszar, 1975, for the general definition.)

It is well known that $I(p \| q) \geq 0$, and that $I(p \| q) = 0$ iff $p \equiv q$. Thus it is heuristically reasonable to think of $I(p \| q)$ as representing a "distance" between p and q. However, $I(\cdot \| \cdot)$ is not a metric, nor is the symmetrized version

$$J(p \| q) = \frac{I(p \| q) + I(q \| p)}{2}$$

used by Jeffreys (1948). Nevertheless,

$$\sum_{k=1}^m |p(k) - q(k)| \leq [2I(p \| q)]^{1/2}, \tag{1.2}$$

Received November 1983; revised November 1984.

[1] This work was partially supported by ONR Contract N00014-83-K-0249.

AMS 1970 *subject classifications.* Primary 90C99; secondary 49D99.

Key words and phrases. I-divergence, I-projections, convexity, Kullback-Liebler information number, cross-entropy, iterative projections, iterative proportional fitting procedure.

as shown independently by Kullback (1967), Kemperman (1967), and Csiszar (1967), so that we have some idea of what small values of $I(p \| q)$ imply.

If we interpret $I(p \| q)$ as distance, then it seems natural to define the I-projection of the probability distribution r onto a set $\mathscr{E}$ of probability distributions as being a $q \in \mathscr{E}$ such that $I(q \| r) < \infty$ and

$$I(q \| r) = \min_{p \in \mathscr{E}} I(p \| r). \tag{1.3}$$

In some sense, q is a measure in $\mathscr{E}$ that lies as close as possible to r.

Minimization problems of the form (1.3) play a key role in the information-theoretic approach to statistics (e.g., Kullback, 1959; Good, 1963; etc.) and also occur in other areas such as the theory of large deviations (Sanov, 1957) and maximation of entropy (Rao, 1965; and Jaynes, 1957).

However, in statistical circles, I-projections are probably most important for being dual problems to certain log-linear model maximum likelihood estimation (MLE) problems. In particular, it is known that the multinomial MLE problem:

$$\text{Maximize} \quad \prod_{k=1}^{m} p(k)^{n(k)} \quad \text{subject to} \quad p \in P, \quad \ln p \in \mathscr{M} \tag{1.4}$$

(where $\mathscr{M}$ is some subspace of $\bar{R}^m$ containing the constant vectors) has precisely the same solution as the I-projection problem:

$$\text{Minimize} \quad I(p \| u) \quad \text{over } P \quad \text{such that} \quad s - p \in \mathscr{M}^{\perp} \tag{1.5}$$

(where $s(k) = (\sum_1^m n(i))^{-1} n(k)$, and u is the uniform distribution on $\mathscr{X}$). Note that for a subspace spanned by the vectors $a_1, \cdots, a_t$, we have

$$\mathscr{M}^{\perp} = \{a_1, \cdots, a_t\}^{\perp} = \cap_1^t \{a_i\}^{\perp},$$

and hence (1.4) is equivalent to minimizing $I(p \| u)$ for p in the set $\cap_1^t (s - \{a_i\}^{\perp})$. Since the $s - \{a_i\}^{\perp}$ are linear spaces, Csiszar's algorithm of cyclic, iterated I-projections is appropriate here. In this format, it is easy to see the connection between Csiszar's procedure and the IPFP (iterative proportional fitting procedure) which has received so much attention in the general area of categorical data. Meyer (1980) has an extensive discussion and several examples where he relates Csiszar's procedure and the general IPFP.

Suppose now that $\mathscr{M} = K_1 + \cdots + K_t$ is a closed, convex cone expressible as a direct sum of closed, convex cones containing the constant vectors, rather than a direct sum of subspaces. Such a configuration would arise naturally if one were considering order constraints in a log-linear model. Then it is well known that the dual (polar) cone of $\mathscr{M}$, defined as

$$\mathscr{M}^* = \{y; \sum_1^m y(i)x(i) \leq 0 \text{ for all } x \in \mathscr{M}\}$$

is expressible as

$$\mathscr{M}^* = (K_1 + K_2 + \cdots + K_t)^* = K_1^* \cap K_2^* \cap \cdots \cap K_t^*.$$

Lemke and Dykstra (1984) have generalized the (1.4)–(1.5) duality results to the case where $\mathscr{M}$ is a closed, convex cone (with $\mathscr{M}^*$ replacing $\mathscr{M}^{\perp}$ in (1.5)). This means that many MLE problems involving partial orders in log-linear models

are equivalent to I-projection problems of the form:

$$\text{Minimize} \quad I(p \| u) \quad \text{for} \quad p \in \cap_1^t (s - K_i^*).$$

If the K_i^*'s are not subspaces, Csiszar's cyclic, iterated scheme need not work. However, the procedure described in this paper will work since the sets $s - K_i^*$ will be closed, convex sets of probability distributions.

Of course, we would really like to be able to identify structure in these log-linear model situations, which leads to the area of inference for various competing models. While these are important questions, we shall only be concerned with the MLE problem in this paper.

Csiszar (1975) discusses I-projections in great detail, and has a geometric development for I-projections which is quite appealing. (Cencov, 1972, also has a geometric development of I-projections, but with the arguments interchanged.) Csiszar also discusses the existence of I-projections, and shows that if $\mathscr{E}$ is a convex set of probability distributions (PDs) which is variation closed, the unique existence of the I-projection of r onto $\mathscr{E}$ is guaranteed provided there exists a $p \in \mathscr{E}$ such that $I(p \| r) < \infty$. This result is clearly applicable for closed, convex sets of PDs on the finite set $\mathscr{X}$. We shall make repeated use of the elegant characterization of I-projections given in the following theorem.

THEOREM 1.1 (Csiszar). *A probability $q \in \mathscr{E}$ with $I(q \| r)$ finite is the I-projection of r onto the convex set $\mathscr{E}$ of probability distributions iff*

$$I(p \| r) \geq I(p \| q) + I(q \| r) \quad \forall p \in \mathscr{E}. \tag{1.6}$$

Note that in our setting of finite $\mathscr{X}$, it follows from (1.1) *that* (1.6) *is equivalent to*

$$\sum_{k=1}^m (p(k) - q(k)) \ln(q(k)/r(k)) \geq 0 \quad \forall p \in \mathscr{E}. \tag{1.7}$$

If in fact q is an algebraic inner point of $\mathscr{E}$, i.e., for every p in $\mathscr{E} - \{q\}$, there exists $0 < \alpha < 1$ and $p' \in \mathscr{E}$ such that $q = \alpha p + (1 - \alpha)p'$, equality must hold in (1.6) and (1.7).

This situation is roughly akin to projecting onto subspaces in least squares theory. In particular, Csiszar defines an $\mathscr{E}$ to be a linear set if $p, p' \in \mathscr{E}$ implies $\alpha p + (1 - \alpha)p' \in \mathscr{E}$ for *every* α for which it is a probability. If $\mathscr{E}$ is a linear set, then the inequality sign in (1.6) and (1.7) may be replaced by an equality sign as long as $\mathscr{X}$ is finite. Based upon this characterization, Csiszar is able to prove that if $\mathscr{X}$ is finite, $\mathscr{E} = \cap_1^t \mathscr{E}_i$ is a finite intersection of arbitrary linear sets, and there exists a $p \in \mathscr{E}$ such that $I(p \| r) < \infty$, then successive, cyclic, iterated I-projections onto the individual sets must converge to the I-projection of r onto $\mathscr{E}$. Thus if $q_0 = r$, and q_n denotes the I-projection of q_{n-1} onto $\mathscr{E}_n$ (where $\mathscr{E}_n = \mathscr{E}_i$ if $n = \ell t + i$, $1 \leq i \leq t$), then q_n must converge to $q \in \mathscr{E}$ as $n \to \infty$ where $I(q \| r) = \min_{p \in \mathscr{E}} I(p \| r)$.

This result is very much dependent upon the assumption that the $\mathscr{E}_i$ be linear sets (in fact it is not true in general) and the accompanying fact that equality holds in (1.6).

2. The procedure. We now propose a procedure which will enable one to obtain I-projections onto a finite intersection of arbitrary closed, convex sets of

probabilities by iteratively finding I-projections onto the individual sets. We will prove that that under a mild restriction, the procedure must give the correct solution, and then we examine an example.

First let us note that we can also define an I-projection onto $\mathscr{E}$ for any nonzero, finite measure r on $\mathscr{X}$ as being a probability in $\mathscr{E}$ which minimizes

$$\sum_1^m q(k)\ln(q(k)/r(k))$$

over $\mathscr{E}$ and has finite I-divergence with respect to r.

Of course, since $\mathscr{E}$ contains only probabilities, the I-projection of r onto a convex set $\mathscr{E}$ is identical to the I-projection onto $\mathscr{E}$ of the normalized measure r' defined by $r'(k) = r(k)/r(\mathscr{X})$. It is easily shown that the characterization of I-projections given in (1.6) and (1.7) is still valid, even though $I(p \parallel r) \geq 0$ need no longer be true.

Let us now state our algorithm. We assume that we wish to find the I-projection of r onto $\mathscr{E} = \cap_1^t \mathscr{E}_i$, where each $\mathscr{E}_i$ is a closed, convex set of probabilities. We assume that we can project onto each $\mathscr{E}_i$ individually and shall denote the I-projection of s onto $\mathscr{E}_i$ by $\pi_i(s)$ and the I-projection of s onto $\mathscr{E}$ by $\pi(s)$. We also assume there exists a $t \in \mathscr{E}$ such that $I(t \parallel r) < \infty$. (Remember that multiplication and division of vectors will refer to the operations being performed coordinate-wise.)

ALGORITHM.

1. Let $s_{11} = r$, and let $p_{11} = \pi_1(s_{11})$. We then set $s_{12} = p_{11} = r(p_{11}/s_{11})$. (We note that if $s_{11}(k) = 0$, then so is $p_{11}(k)$. We take 0/0 to be 1.)
2. Let $p_{12} = \pi_2(s_{12})$. Set $s_{13} = p_{12} = r(p_{11}/s_{11})(p_{12}/s_{12})$.
3. Continue, until $p_{1t} = \pi_t(s_{1t})$ where $s_{1t} = p_{1,t-1} = r(p_{11}/s_{11}) \cdots (p_{1,t-1}/s_{1,t-1})$. Set $s_{21} = r(p_{12}/s_{12}) \cdots (p_{1t}/s_{1t})$. Note that $s_{21} = p_{1t}/(p_{11}/s_{11})$.
4. Set $p_{21} = \pi_1(s_{21})$, and then set $s_{22} = r(p_{21}/s_{21})(p_{13}/s_{13}) \cdots (p_{1t}/s_{1t})$, or equivalently, $s_{22} = p_{21}/(p_{12}/s_{12})$.
5. Continue. In general, set $s_{ni} = p_{n,i-1}/(p_{n-1,i}/s_{n-1,i})$, $2 \leq i \leq t$ and set $s_{n,1} = p_{n-1,t}/(p_{n-1,1}/s_{n-1,1})$. We then let $p_{ni} = \pi_i(s_{ni})$, and define $s_{n,i+1}$, or $s_{n+1,1}$ if $i = t$, in similar fashion.

Suppose now that the $\mathscr{E}_i$ are actually linear sets so that equality holds in (1.7). Noting that for any $p \in \mathscr{E}_i$

$$\begin{aligned} I(p \parallel s_{n,i}) &= \textstyle\sum_k p(k)\ln[p(k)/s_{ni}(k)] \\ &= \textstyle\sum_k p(k)\ln[p(k)/(p_{n,i-1}(k)/(p_{n-1,i}(k)/s_{n-1,i}(k)))] \qquad (2.1) \\ &= \textstyle\sum_k p(k)\ln[p(k)/p_{n,i-1}(k)] + \sum_k p(k)\ln[p_{n-1,i}(k)/s_{n-1,i}(k)], \qquad (2.2) \end{aligned}$$

we observe that the last term must be equal to

$$\textstyle\sum_k p_{n-1,i}(k)\ln[p_{n-1,i}(k)/s_{n-1,i}(k)]$$

and hence free of p. Thus the $p \in \mathscr{E}_i$ which minimizes (2.1), is also the one which minimizes the first part of (2.2), i.e., the I-projection of $p_{n,i-1}$ onto $\mathscr{E}_i$. It easily follows that our procedure reduces to the cyclic, iterative procedure given by Csiszar when the $\mathscr{E}_i$ are closed, linear sets.

For a simple example to show that Csiszar's procedure does not work for general convex sets, consider the following:

$$\mathscr{E}_1 = \left\{ \begin{pmatrix} p_{11}, & p_{12} \\ p_{21}, & p_{22} \end{pmatrix}; p_{11} \geq p_{12}, p_{21} \geq p_{22}, p_{ij} \geq 0, \sum_1^2 \sum_1^2 p_{ij} = 1 \right\},$$

$$\mathscr{E}_2 = \left\{ \begin{pmatrix} p_{11}, & p_{12} \\ p_{21}, & p_{22} \end{pmatrix}; p_{11} \geq p_{21}, p_{12} \geq p_{22}, p_{ij} \geq 0, \sum_1^2 \sum_1^2 p_{ij} = 1 \right\},$$

and

$$r = \begin{pmatrix} 1/16 & 3/16 \\ 7/16 & 5/16 \end{pmatrix}.$$

Csiszar's procedure yields ($9/32$ $7/32$, $9/32$ $7/32$), whereas ($1/4$ $1/4$, $1/4$ $1/4$) is the correct solution.

Everything hinges on the following theorem, which retains the notation used in describing the algorithm.

THEOREM 2.1. *Assume $\mathscr{E} = \cap_1^t \mathscr{E}_i$, where the $\mathscr{E}_i$ are closed, convex sets of probability distributions and $r \neq 0$ is a nonnegative vector such that there exists a $t \in \mathscr{E}$ where $I(t \| r) < \infty$. If there exists a convergent subsequence $p_{n_j,i} \to p$ for some i such that*

$$\liminf_j \sum_k (p_{n_j,i}(k) - p(k)) \ln\left(\frac{p_{n_j,i}(k)}{s_{n_j,i}(k)}\right) \geq 0 \tag{2.3}$$

for every i, then $p_{n,i} \to p$ as $n \to \infty$ and $p = \pi(r)$.

PROOF. Recall that $p_{ni} = \pi_i(s_{ni})$ where

$$s_{ni} = r \frac{p_{n1}}{s_{n1}} \cdots \frac{p_{n,i-1}}{s_{n,i-1}} \frac{p_{n-1,i+1}}{s_{n-1,i+1}} \cdots \frac{p_{n-1,t}}{s_{n-1,t}}.$$

$$= \begin{cases} p_{n-1,t}\left(\dfrac{p_{n-1,1}}{s_{n-1,1}}\right)^{-1} & \text{if } i = 1 \\ p_{n,i-1}\left(\dfrac{p_{n-1,i}}{s_{n-1,i}}\right)^{-1} & \text{if } 2 \leq i \leq t. \end{cases} \tag{2.4}$$

Thus

$$
\begin{aligned}
&I(p_{ni} \,\|\, s_{ni}) - I(p_{n-1,i} \,\|\, s_{n-1,i}) \\
&\qquad = \sum p_{ni} \ln\left(\frac{p_{ni}}{s_{ni}}\right) - \sum p_{n-1,i} \ln\left(\frac{p_{n-1,i}}{s_{n-1,i}}\right) \\
&\qquad = \sum p_{ni} \ln\left(\frac{p_{ni}}{p_{n,i-1}}\right) + \sum p_{ni} \ln\left(\frac{p_{n-1,i}}{s_{n-1,i}}\right) - \sum p_{n-1,i} \ln\left(\frac{p_{n-1,i}}{s_{n-1,i}}\right) \qquad (2.5) \\
&\qquad = \sum p_{ni} \ln\left(\frac{p_{ni}}{p_{n,i-1}}\right) + \sum (p_{ni} - p_{n-1,i}) \ln\left(\frac{p_{n-1,i}}{s_{n-1,i}}\right) \\
&\qquad \geq I(p_{ni} \,\|\, p_{n,i-1}) \quad \text{if} \quad 2 \leq i \leq t, \quad \text{and similarly if} \quad i = 1,
\end{aligned}
$$

since (by 1.7) the last term must be nonnegative because $p_{ni} \in \mathscr{E}_i$. Noting that $I(p_{ni} \,\|\, p_{n,i-1}) \geq 0$ since p_{ni} and $p_{n,i-1}$ are probabilities, we have that $I(p_{ni} \,\|\, s_{ni})$ is nondecreasing in n for each i. Let us now show that these sequences are bounded above.

For $v \in \bigcap_1^t \mathscr{E}_i$,

$$
\begin{aligned}
I(v \,\|\, p_{ni}) &= \sum v \ln\left(\frac{v}{p_{ni}}\right) \\
&= \sum v \ln v - \sum v \ln r \left[\frac{p_{n1}}{s_{n1}} \cdots \frac{p_{ni}}{s_{ni}} \frac{p_{n-1,\,i+1}}{s_{n-1,i+1}} \cdots \frac{p_{nt}}{s_{nt}}\right] \\
&= \sum v \ln v - \sum v \ln r - \sum_{j=1}^t \left[\sum v \ln \frac{p_{aj}}{s_{aj}} - I(p_{aj} \,\|\, s_{aj})\right] \qquad (2.6) \\
&\quad - \sum_{j=1}^t I(p_{aj} \,\|\, s_{aj}) \quad \text{where} \quad a = \begin{cases} n, & j \leq i \\ n-1, & j > i \end{cases} \\
&\leq I(v \,\|\, r) - \sum_{j=1}^t I(p_{aj} \,\|\, s_{aj})
\end{aligned}
$$

by (1.6) and the fact that v belongs to *every* $\mathscr{E}_i$. Thus, choosing v such that $I(v \,\|\, r) < \infty$, we have a uniform upper bound on $I(p_{ni} \,\|\, s_{ni})$. Hence the $\lim_{n\to\infty} I(p_{ni} \,\|\, s_{ni})$ exists finite for every i, and by (2.5)

$$
I(p_{ni} \,\|\, p_{n,i-1}) \to 0 \quad \text{as} \quad n \to \infty \quad \text{for} \quad 2 \leq i \leq t. \qquad (2.7)
$$

(Similarly $I(p_{n1} \,\|\, p_{n-1,t}) \to 0$ as $n \to \infty$.)

Note that (1.2) and (2.7) imply that if $p_{n_j,i} \to p$ for some i, then $p_{n_j,i} \to p$ for every i. Thus $p \in \bigcap_1^t \mathscr{E}_i$ since $p_{n_j,i} \in \mathscr{E}_i$ and the $\mathscr{E}_i$ are closed.

Using (1.7) and (2.3), for any $v \in \bigcap_1^t \mathscr{E}_i$,

$$
\begin{aligned}
0 &\leq \textstyle\sum_i \liminf_j \sum_k (p_{n_j,i} - p) \ln(p_{n_j,i}/s_{n_j,i}) + \sum_i \sum_k (v - p_{n_j,i}) \ln(p_{n_j,i}/s_{n_j,i}) \\
&\leq \liminf_j \textstyle\sum_i \sum_k (v - p) \ln(p_{n_j,i}/s_{n_j,i}) \\
&= \liminf_j \textstyle\sum_k (v - p) \ln r \prod_i (p_{n_j,i}/s_{n_j,i})/r \\
&= \lim_j \textstyle\sum_k (v - p) \ln(p_{n_j} t/r) = \sum_k (v - p) \ln(p/r).
\end{aligned}
$$

Thus $p = \pi(r)$ by (1.7). Setting $v = p$ in (2.6) and using (2.3), it follows that for the subsequence $\{n_j\}$,

$$\sum_{i=1}^{t} I(p_{n_j,i} \| s_{n_j,i}) \to I(p \| r).$$

However, by the monotonicity in n of $I(p_{ni} \| s_{ni})$, we have that

$$\sum_{i=1}^{t} I(p_{n,i} \| s_{n,i}) \to I(p \| r).$$

Thus $I(p \| p_{nt}) \to 0$, so that by (1.2) and (2.5), $p_{ni} \to p$ as $n \to \infty$ for all i.

We remark that it would take rather surprising behavior of the p_{ni} for condition (2.3) to not hold, and we strongly conjecture that this condition is always true. As we note in the following corollary, if the $p_{n,i} \nrightarrow \pi(r)$ we must have $\sup_{n,i} \sum_k s_{ni}(k) \to \infty$, and we have been unable to construct examples where this happens. We point out that when one uses the algorithm, a step can be put in to check the value of $\sup_{n,i} \sum_k s_{ni}(k)$. If the algorithm is not going to converge correctly, then this value must become excessively large. Otherwise, the algorithm must converge to the correct solution.

COROLLARY 2.1. *If the algorithm should not converge correctly, then* $\sup_{n,i} \sum_k s_{ni}(k) \to \infty$ *as* $n \to \infty$.

PROOF. Suppose $s_{ni}(k)$ is uniformly bounded above, and $p_{n_j,i} \to p$ is a convergent subsequence (which must exist since $0 \le p_{ni}(k) \le 1$). If there exists $0 < m \le s_{n_j,i}(k)$ for all n_j, k, then (2.3) follows by standard continuity properties. If not, there exists at least one k such that $s_{n_j,i}(k)$ takes on arbitrarily small values. Then either $p_{n_j,i}(k) \to 0$ for all such k (which cause no problems), or there exists some k such that $p_{n_j,i}(k) \to p(k) > 0$, while $s_{n_j,i}(k)$ takes on arbitrarily small values. This contradicts the uniform upper bound of $I(p_{ni} \| s_{ni})$.

We recommend that when one uses the algorithm, one should compute the average value over an entire cycle ($(1/t) \sum_{i=1}^{t} p_{n,i}$) rather than a single projection p_{ni} to estimate the I-projection. Convergence is still guaranteed, and this value is much more stable and seems to converge much more quickly to the correct solution.

3. An example. We consider the case where $\mathscr{X}$ is an $n \times n$ table. A probability on $\mathscr{X}$ is represented by an $n \times n$ matrix (m_{ij}) of nonnegative numbers which sum to one. We denote the corresponding marginals by

$$m_{k.} = \sum_{j=1}^{n} m_{kj} \quad \text{and} \quad m_{.j} = \sum_{k=1}^{n} m_{kj}, \quad k, j = 1, \cdots, n.$$

We now consider the problem of finding the I-projection of a fixed array (r_{kj}) subject to the marginal PD's being stochastically ordered, i.e.,

$$\sum_{\ell=1}^{i} m_{\ell.} \ge \sum_{\ell=1}^{i} m_{.\ell}, \quad \text{for all} \quad i.$$

(Kullback, 1971, has given an iterative procedure for I-projections where equality is forced to hold for all i, also known as marginal homogeneity.)

Equivalently, we want to find the I-projection of (r_{kj}) onto

$$(3.1) \quad \mathscr{E} = \cap_1^{n-1} \mathscr{E}_i \quad \text{where} \quad \mathscr{E}_i = \{(m_{kj}); \sum_{\ell=1}^{i} \sum_{h=1}^{n} m_{\ell h} \ge \sum_{h=1}^{i} \sum_{\ell=1}^{n} m_{\ell h}\}.$$

Note that the $\mathscr{E}_i$ are closed, convex sets of probabilities which are *not* linear sets. I-projections of (v_{kj}) onto the $\mathscr{E}_i$ can be found by forcing equality in the constraint if the (v_{kj}) violate the constraint. (See Theorem 2.11 of Barlow et al., 1972, which can be modified to apply to arbitrary, convex functions).

To express $\pi_i(v)$, we let

$$A_i = \{(\ell, m);\ 1 \le \ell \le i,\ i+1 \le m \le n\},$$

$$B_i = \{(\ell, m);\ 1 \le m \le i,\ i+1 \le \ell \le n\},$$

$$C_i = \{(\ell, m);\ \ell, m = 1, \cdots, i\} \cup \{(\ell, m);\ \ell, m = i+1, \cdots, n\}$$

and

$$\hat{v}_i = (\textstyle\sum_{A_i} v_{\ell m} \sum_{B_i} v_{\ell m})^{1/2}.$$

If (v_{kj}) satisfies the constraint of $\mathscr{E}_i$, namely $\sum_{A_i} v_{\ell m} \ge \sum_{B_i} v_{\ell m}$, then $\pi_i(v)$ has (k, j)th component

$$\pi_i(v)_{kj} = v_{kj} / \textstyle\sum_1^n \sum_1^n v_{\ell m} \quad \text{for all} \quad k, j.$$

If (v_{kj}) does not satisfy the constraint imposed by $\mathscr{E}_i$, that is $\sum_{A_i} v_{\ell m} < \sum_{B_i} v_{\ell m}$, then

$$\pi_i(v)_{kj} = \begin{cases} v_{kj}\left(\dfrac{\sum_{B_i} v_{\ell m}}{\sum_{A_i} v_{\ell m}}\right)^{1/2} [2\hat{v}_i + \textstyle\sum_{C_i} v_{\ell m}]^{-1}, & (k, j) \in A_i, \\ v_{kj}\left(\dfrac{\sum_{A_i} v_{\ell m}}{\sum_{B_i} v_{\ell m}}\right)^{1/2} [2\hat{v}_i + \textstyle\sum_{C_i} v_{\ell m}]^{-1}, & (k, j) \in B_i, \\ v_{kj}[2\hat{v}_i + \textstyle\sum_{C_i} v_{\ell m}]^{-1}, & (k, j) \in C_i. \end{cases}$$

The key point is that finding the I-projection onto $\mathscr{E}_i$ is quite easy (and easily programmed), while finding the I-projection onto $\mathscr{E} = \cap^{n-1} \mathscr{E}_i$ is very difficult. However our algorithm enables one to find the latter I-projection using only the ability to handle the I-projections onto the individual $\mathscr{E}_i$.

To illustrate our example with some numbers, we consider some rather famous data from Stuart (1953) concerning grades of unaided distance vision for left and

TABLE 1
Unaided distance vision (From Kendall, 1974)

Right eye	Left eye				Totals
	Highest grade	Second grade	Third grade	Lowest grade	
Highest grade	821	112	85	35	1053
Second grade	116	494	145	27	782
Third grade	72	151	583	87	893
Lowest grade	43	34	106	331	514
Totals	1052	791	919	480	3242

TABLE 2
I-Projection of data in Table 1
(Values in parentheses are Table 1 values normed to sum to unity)

Right eye	Left eye				Totals
	Highest grade	Second grade	Third grade	Lowest grade	
Highest grade	.2534	.0344	.0262	.0120	.3260
	(.2532)	(.0345)	(.0262)	(.0108)	(.3247)
Second grade	.0358	.1525	.0447	.0092	.2422
	(.0358)	(.1524)	(.0447)	(.0083)	(.2412)
Third grade	.0222	.0466	.1799	.0298	.2785
	(.0222)	(.0466)	(.1798)	(.0268)	(.2754)
Lowest grade	.0120	.0095	.0295	.1022	.1532
	(.0133)	(.0105)	(.0327)	(.1021)	(.1586)
Totals	.3234	.2430	.2803	.1532	.9999
	(.3245)	(.2440)	(.2834)	(.1480)	(.9999)

right eyes. If one wished to estimate the probabilities of falling into the various categories, subject to the provision that right eye vision is at least as good as left eye vision, one might find the I-projection of the data in Table 1 onto the $\mathscr{E}$ given in (3.1). Using this algorithm, we have essentially obtained convergence to the true I-projection by 3 cycles. These values are listed in Table 2 (with the unrestricted MLEs given in parentheses).

This estimate might prove useful for constructing a likelihood ratio type test for testing whether right eye vision is better than left eye vision. This data is treated by Plackett (1981) who uses it for tests involving marginal homogeneity and quasi-symmetry. It also appears in Kendall (1974).

4. Acknowledgement. The author would like to thank a referee for helpful, knowledgeable comments and Peter Wollan for doing the calculations in Table 2.

REFERENCES

BARLOW, R. E, BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

CENCOV, N. N. (1972). *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs, American Mathematical Society, Vol. 53, 115–125.

CSISZAR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318.

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

GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34** 911–934.

JAYNES, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.* **106** 620–630.

JEFFREYS, H. (1948). *Theory of Probability.* 2nd ed. Clarendon Press, Oxford.

KEMPERMAN, J. H. B. (1967). On the optimum rate of transmitting information. Probability and Information Theory. *Lecture Notes in Mathematics*, 126–169. Springer-Verlag, New York.

KENDALL, M. G. (1974). *Rank Correlation Methods.* Griffin, London.
KULLBACK, S. (1959). *Information Theory and Statistics.* Wiley, New York.
KULLBACK, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Informa. Theory* **IT-13** 126–127.
KULLBACK, S. (1971). Marginal homogeneity of multidimensional contingency tables. *Ann. Math. Statist.* **42** 594–606.
LEMKE, J. H. and DYKSTRA, R. L. (1984). Multinomial maximum likelihood estimation with multiple cone restrictions (submitted, *Ann. Statist.*).
MEYER, M. M. (1980). Generalizing the iterative proportional fitting procedure. Dept. of Applied Statistics, Tech. Report No. 371, Univ. of Minnesota.
PLACKETT, R. L. (1981). *The Analysis of Categorical Data.* Macmillan, New York.
RAO, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.
SANOV, I. N. (1957). On the probability of large deviations of random variables. *Mat. Sb.* **42** 11–44.
STUART, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika* **40** 105–110.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCES
THE UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242