The Annals of Probability
1990, Vol. 18, No. 3, 1410–1415

ON DYKSTRA'S ITERATIVE FITTING PROCEDURE

BY WILLIAM E. WINKLER

Energy Information Administration

This paper shows that Dykstra's procedure for finding the I-projection onto the intersection of closed convex sets holds in general. It does this by first showing that each of the I-projections onto individual convex sets defined in Dykstra's iterative procedure exists and that no condition such as imposed by Dykstra is required to prove the convergence of the iterative procedure to a unique I-projection.

1. Introduction. This paper shows that Dykstra's (1985) iterative procedure for finding the I-projection onto the intersection of closed convex sets holds in general. Specifically, given the existence of an I-projection onto the intersection of a finite number of closed convex sets, Dykstra iteratively defines a sequence of I-projections onto each of t closed convex sets. Given the assumption that each of the successive I-projections exists and under an additional restraint, Dykstra proves that the sequence of I projections converges to the desired I-projection on the intersection.

In this paper, we prove that each of the iteratively defined I-projections exists and, thus, Dykstra's procedure is well defined. We show that Dykstra's additional restraint, which is sufficient to show that some subsequence of the I-projections converges to an I-projection on the intersection, is not needed. As our proof closely parallels Dykstra's, we will repeat some of his preliminaries and notation. Following the definition of Dykstra's algorithm, we will begin with generally new results.

Let p and q be probability measures defined on subsets of the finite set X, which without loss of generality we take to be the first K positive integers. We use $p(k)$ to denote the mass that measure p assigns to point k.

The I-divergence of p with respect to q, also called the Kullback–Liebler information number, is given by

$$I(p\|q) = \begin{cases} \sum_k p(k)\ln(p(k)/q(k)), & \text{if } p \ll q, \\ \infty, & \text{otherwise.} \end{cases} \tag{1.1}$$

Following Csiszár [(1975), page 146] we observe the conventions $\ln 0 = -\infty$, $\ln(a/0) = +\infty$ and $0 \cdot (\pm\infty) = 0$. The conventions are consistent with definition (1.1) and will be used in characterizing Dykstra's algorithm.

We let P denote the set of all probability measures on X and use the convention that products (or quotients) are to be interpreted as pointwise

Received September 1987; revised March 1988.
AMS 1980 *subject classifications*. Primary 90C99; secondary 49D99.
Key words and phrases. Iterative fitting procedure, I-projection, Kullback–Liebler information number.

multiplication (or division). For example, $s_{12} = r \cdot p_{11}/s_{11}$ means the measure s_{12} putting mass $r(k) \cdot p_{11}(k)/s_{11}(k)$ at k.

We define the I-projection of the probability distribution r onto a set E of probability distributions as being q in E such that $I(q\|r) < \infty$ and

$$I(q\|r) = \min_{p \in E} I(p\|r). \tag{1.2}$$

Csiszár (1975) provided the following elegant characterization of I-projections.

THEOREM 1.1 (Csiszár). *A probability $q \in E$ with $I(q\|r) < \infty$ is the I-projection of r onto the convex set E of probability distributions iff*

$$I(p\|r) \geq I(p\|q) + I(q\|r) \quad \textit{for all } p \in E. \tag{1.3}$$

The next section contains a description of Dykstra's procedure and the proof that it converges without the restraint imposed by Dykstra. We note that our proof is heavily dependent on the finiteness of X.

2. The procedure. The procedure we describe is precisely Dykstra's. We assume that there exists a q in E such that $I(q\|r) < \infty$ and that we wish to find the I-projection of r onto E, the intersection of closed, convex sets $E_1, E_2, \dots, E_t$ of probability distributions. We denote the I-projection of s onto E_i by $\pi_i(s)$ and the I-projection of s onto E by $\pi(s)$. That $\pi(r)$ and $\pi_i(r)$ exist is an immediate consequence of Theorem 2.1 of Csiszár (1975). We wish to show that, if we can project r onto each E_i individually, then we can define an iterative procedure in which a sequence of successive projections of measures (not necessarily probability distributions) onto all the E_is converges to the projection on E. For this we need

DYKSTRA'S ALGORITHM.
Initialization. Let $s_{1,1} = r$ and $p_{1,1} = \pi_1(s_{1,1})$. Let $s_{1,i} = p_{1,i-1}$ and $p_{1,i} = \pi_i(s_{1,i})$ for $i = 2, \dots, t$.
Induction. For $n \geq 2$, let $s_{n,i} = p_{n,i-1}/(p_{n-1,i}/s_{n-1,i})$, $2 \leq i \leq t$, $s_{n,1} = p_{n-1,t}/(p_{n-1,1}/s_{n-1,1})$ and $p_{n,i} = \pi_i(s_{n,i})$.

Note that if $s_{n,i}(k) = 0$, then $p_{n,i}(k) = 0$. Take $0/0 = 1$.

We observe that if r_0 is an arbitrary distribution and p is a probability, then $I(p\|r_0)$ is still well defined but may not be nonnegative. (1.3) still holds with r_0 replacing r. We now prove that each of the successive projections defined in Dykstra's algorithm exists. Dykstra assumed that the successive projections exist.

LEMMA 2.1. *Let X be a finite state space, $E = \bigcap_{i=1}^{t} E_i$ where $E_1, \dots, E_t$ are closed, convex sets of probability distributions on X and let r be a measure on*

X. Assume there exists $q \in E$ *such that* $I(q\|r) < \infty$. *Then the projections* $\pi_j(s_{n,j})$ *of* $s_{n,j}$ *onto* E_j *defined by Dykstra's algorithm exist for all* n *and* j.

PROOF. By hypothesis, the projection of r onto E and, consequently, each E_j must necessarily exist. Using (1.3) yields that $\pi_j(r)$ only assigns mass 0 to a point k when $r(k) = 0$ or when $p(k) = 0$ for all $p \in E_j$.

Reasoning inductively yields that each $s_{n,j}$ can only assign mass 0 to a point k if $r(k) = 0$ or if there exists j such that $p(k) = 0$ for all $p \in E_j$. Consequently, each E_j contains $p \ll s_{n,j}$ for $n = 1, 2 \dots$ and by Csiszár [(1975), page 154], $I(p\|s_{n,j}) < \infty$. By Csiszár (1975, Theorem 2.1), $\pi_j(s_{n,j})$ exists for $n = 1, 2, \dots$ and $j = 1, 2, \dots$. □

The proof of Lemma 2.1 immediately yields that if a projection in the sequence determined by Dykstra's algorithm assigns mass 0 to a point, then each successive projection must also assign mass 0 to the same point.

That the algorithm can be used to find a sequence of probabilities converging to the desired projection follows from Theorem 2.1. The proof extends Dykstra's proof. To prove his results, Dykstra had to assume [(1985), condition (2.3), page 979] that the sequence of projections $p_{n,j}$ satisfies the following condition:

There exists a convergent subsequence $p_{n_j,i} \to p$ for some i such that

$$\text{(D)} \qquad \liminf_{j \to \infty} \sum_k \left(p_{n_j,i}(k) - p(k)\right) \ln\left(p_{n_j,i}(k)/s_{n_j,i}(k)\right) \geq 0.$$

Condition (D) is sufficient to show that the sequence of projections $p_{n,j}$ converges to a projection. In the remainder of the paper, for brevity, we will use summations of the form $\Sigma_k p$ to denote $\Sigma_k p(k)$.

THEOREM 2.1. *Assume* $E = \bigcap_{j=1}^t E_j$, *where the* E_j *are closed, convex sets of probability distributions. Let* $r \neq 0$ *be a nonnegative vector such that there exists a* $q \in E$ *for which* $I(q\|r) < \infty$. *Let* $p_{n,i}$ *be the sequence of I-projections onto* E_i *defined by Dykstra's algorithm. Then for every* i, $p_{n,i} \to p$ *as* $n \to \infty$ *and* $p = \pi(r)$.

PROOF. For any n and i,

$$(2.1) \quad I(p_{n,i}\|s_{n,i}) = \sum_k p_{n,i} \ln(p_{n,i}/p_{n,i-1}) + \sum_k p_{n,i} \ln(p_{n-1,i}/s_{n-1,i}).$$

Iterating we obtain that, for all n and i and for all $j' < n$,

$$(2.2) \quad \begin{aligned} I(p_{n,i}\|s_{n,i}) = &\sum_{m=j'+1}^{n} \sum_k p_{n,i} \ln(p_{m,i}/p_{m,i-1}) \\ &+ \sum_k (p_{n,i} - p_{j',i}) \ln(p_{j',i}/s_{j',i}) + \sum_k p_{j',i} \ln(p_{j',i}/s_{j',i}). \end{aligned}$$

As the $p_{n,i}$ are bounded, we can choose a subsequence $p_{n_j,i}$, $j = 1, 2, \dots,$ such that $p_{n_j,i} \to p$ for some p as $j \to \infty$. By Dykstra [(1985), page 980] for each i, $I(p_{n,i} \| s_{n,i})$ is bounded and monotonely increases to a limit as $n \to \infty$.

Using (2.2), we obtain that, for any j',

$$\lim_{n_j \to \infty} \sum_{m=j'+1}^{n_j} \sum_k p_{n_j,i} \ln(p_{m,i}/p_{m,i-1})$$

exists and is finite.

For all $j' < j''$,

$$\begin{aligned} \lim_{n_j \to \infty} \sum_{m=j'+1}^{n_j} \sum_k p_{n_j,i} \ln(p_{m,i}/p_{m,i-1}) &= \sum_{m=j'+1}^{j''} \sum_k p_{n_j,i} \ln(p_{m,i}/p_{m,i-1}) \\ &\quad + \lim_{n_j \to \infty} \sum_{m=j''+1}^{n_j} \sum_k p_{n_j,i} \ln(p_{m,i}/p_{m,i-1}). \end{aligned} \tag{2.3}$$

Letting $j'' \to \infty$ along subsequence n_j yields

$$\lim_{n_k \to \infty} \lim_{n_j \to \infty} \sum_{m=n_k+1}^{n_j} \sum_k p_{n_j,i} \ln(p_{m,i}/p_{m,i-1}) = 0, \tag{2.4}$$

where n_k is also used to denote subsequence n_j.

In (2.2), we let $n \to \infty$ along subsequence n_j, $j = 1, 2, \dots,$ to obtain, for all j',

$$\begin{aligned} \lim_{n_j \to \infty} I(p_{n_j,i} \| s_{n_j,i}) &= \lim_{n_j \to \infty} \sum_{m=j'+1}^{n_j} \sum_k p_{n,i} \ln(p_{m,i}/p_{m,i-1}) \\ &\quad + \lim_{n_j \to \infty} \sum_k (p_{n_j,i} - p_{j',i}) \ln(p_{j',i}/s_{j',i}) \\ &\quad + \sum_k p_{j',i} \ln(p_{j',i}/s_{j',i}). \end{aligned} \tag{2.5}$$

In (2.5) let $j' \to \infty$ along subsequence n_j, $j = 1, 2, \dots$. To avoid confusion we again also label subsequence n_j by n_k. We then have that

$$\begin{aligned} \lim_{n_j \to \infty} I(p_{n_j,i} \| s_{n_j,i}) &= \lim_{n_k \to \infty} \lim_{n_j \to \infty} \sum_{m=n_k+1}^{n_j} \sum_k p_{n_j,i} \ln(p_{m,i}/p_{m,i-1}) \\ &\quad + \lim_{n_k \to \infty} \sum_k (p - p_{n_k,i}) \ln(p_{n_k,i}/s_{n_k,i}) \\ &\quad + \lim_{n_k \to \infty} I(p_{n_k,i} \| s_{n_k,i}). \end{aligned} \tag{2.6}$$

As the first limit on the right-hand side of (2.6) is zero, we obtain Dykstra's condition (D) and the theorem follows. □

TABLE 1
Population array used as initial array in Dykstra's GIFP

Variable 2 strata	Variable 1 strata 1	2	3	4	5	Total
1	0	4	7	0	1	12
2	10	8	0	0	16	34
3	2	0	0	47	85	134
4	6	4	3	78	157	248
5	0	38	26	67	451	582
	18	54	36	192	710	1010

TABLE 2
Fitted array using classical IPF, cells (1, 3) *and* (3, 1) *exceed population values*

Variable 2 strata	Variable 1 strata 1	2	3	4	5	Total
1	0.0	3.027	7.942	0.0	0.031	11
2	7.340	6.149	0.0	0.0	0.511	14
3	2.060	0.0	0.0	5.132	3.808	11
4	1.600	1.117	1.256	2.206	1.821	8
5	0.0	3.707	3.802	0.662	1.829	10
	11	14	13	8	8	54

TABLE 3
Fitted array using Dykstra's GIFP with cells (1, 3) *and* (3, 1) *restrained cell* (3, 1) *less than population value*

Variable 2 strata	Variable 1 strata 1	2	3	4	5	Total
1	0.0	3.954	7.000	0.0	0.046	11
2	7.530	5.916	0.0	0.0	0.554	14
3	1.970	0.0	0.0	5.183	3.847	11
4	1.500	0.982	1.529	2.184	1.805	8
5	0.0	3.148	4.471	0.633	1.748	10
	11	14	13	8	8	54

3. Example. The following example illustrates a situation for which ordinary iterative proportional fitting [see, e.g., Bishop, Fienberg and Holland (1975)] cannot be properly used but for which Dykstra's iterative fitting procedure can. The data are similar to data arising in multipurpose sampling [Winkler (1986)].

We have a two-way population matrix that is induced by two univariate stratifications (Table 1). We wish to fit two univariate samples (the margins of Table 2) to a two-way matrix. Using classical iterative proportional fitting with population matrix (Table 1) as the initial matrix yields the matrix given by Table 2. Cells $(1, 3)$ and $(3, 1)$ exceed available population values of 7 and 2, respectively.

Using Dykstra's generalized iterative fitting procedure with cells $(1, 3)$ and $(3, 1)$ constrained to be less than 7 and 2, respectively, yields the matrix given in Table 3. Cell $(1, 3)$ equals its maximum of 7 and cell $(3, 1)$ takes value 1.970, which is less than its maximum of 2.

Acknowledgments. The author would like to thank Dr. Yvonne Bishop for a number of helpful comments, Professor Richard Dykstra for detailed comments on an earlier version of this paper and Professor Peter Wollan for supplying the FORTRAN routines [Dykstra and Wollan, (1987)] used in fitting the example.

REFERENCES

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

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

DYKSTRA, R. and WOLLAN, P. (1987). Finding *I*-projections subject to a finite set of linear inequality constraints. *Appl. Statist.* **36** 377–383.

WINKLER, W. (1986). Multi-purpose survey sampling. *ASA Proc. Survey Res. Methods Sec.* 614–619.

7705 HERITAGE DRIVE
ANNANDALE, VIRGINIA 22003