

**Information Geometry of Probability Distributions and Minimization Problems**



I. Csiszar

*The Annals of Probability*, Vol. 3, No. 1 (Feb., 1975), 146-158.

Stable URL:

<http://links.jstor.org/sici?sici=0091-1798%28197502%293%3A1%3C146%3AGOPDAM%3E2.0.CO%3B2-R>

*The Annals of Probability* is currently published by Institute of Mathematical Statistics.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ims.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## I-DIVERGENCE GEOMETRY OF PROBABILITY DISTRIBUTIONS AND MINIMIZATION PROBLEMS

BY I. CSISZÁR

*Mathematical Institute of the Hungarian Academy of Sciences*

Some geometric properties of PD's are established, Kullback's *I*-divergence playing the role of squared Euclidean distance. The minimum discrimination information problem is viewed as that of projecting a PD onto a convex set of PD's and useful existence theorems for and characterizations of the minimizing PD are arrived at. A natural generalization of known iterative algorithms converging to the minimizing PD in special situations is given; even for those special cases, our convergence proof is more generally valid than those previously published. As corollaries of independent interest, generalizations of known results on the existence of PD's or nonnegative matrices of a certain form are obtained. The Lagrange multiplier technique is not used.

**1. Introduction.** Capital  $P, Q, R$  will denote PD's (probability distributions) on a measurable space  $(X, \mathcal{L})$  which will not be mentioned in the sequel. If  $P \ll Q$  (or  $Q \ll R$ , etc.) the corresponding density (Radon-Nikodym derivative) will be denoted by  $p_Q(x)$  (or  $q_R(x)$ , etc.); the argument  $x$  will be omitted if this does not cause ambiguity.

The *I*-divergence or Kullback-Leibler information number  $I(P \parallel Q)$ —also called information for discrimination, information gain or entropy of  $P$  relative to  $Q$ —is defined as

$$(1.1) \quad \begin{aligned} I(P \parallel Q) &= \int \log p_Q dP = \int p_Q \log p_Q dQ && \text{if } P \ll Q \\ &= +\infty && \text{if } P \not\ll Q. \end{aligned}$$

If  $R$  is any PD with  $P \ll R, Q \ll R$  (1.1) may be equivalently written as

$$(1.2) \quad I(P \parallel Q) = \int p_R \log \frac{p_R}{q_R} dR.$$

Here and in the sequel we understand

$$(1.3) \quad \log 0 = -\infty, \quad \log \frac{a}{0} = +\infty, \quad 0 \cdot (\pm\infty) = 0.$$

$I(P \parallel Q)$  is always nonnegative and vanishes only for  $P = Q$ .

We shall not be concerned with the information theoretic significance of *I*-divergence; rather, we look at it simply as a quantity measuring how much  $P$  differs from  $Q$ . Given a PD  $R$ , the set of PD's

$$(1.4) \quad S(R, \rho) = \{P : I(P \parallel R) < \rho\} \quad (0 < \rho \leq \infty)$$

---

Received February 20, 1973; revised December 27, 1973.

AMS 1970 subject classifications. Primary 60-00-E05, 62-B10; Secondary 49-F22, 15-A48.

Key words and phrases. Geometry of probability distributions, minimum discrimination information, distributions with given marginals, contingency tables, iterative proportional fitting procedure.

will be called an *I-sphere* with center  $R$  and radius  $\rho$ . If  $\mathcal{E}$  is a convex set of PD's intersecting  $S(R, \infty)$ , a PD  $Q \in \mathcal{E}$  satisfying

$$(1.5) \quad I(Q \parallel R) = \min_{P \in \mathcal{E}} I(P \parallel R)$$

will be called the *I-projection* of  $R$  on  $\mathcal{E}$ . If such  $Q$  exists, the convexity of  $\mathcal{E}$  guarantees its uniqueness since  $I(P \parallel R)$  is strictly convex in  $P$ , as one immediately sees from (1.1).

As demonstrated by Kullback [14], minimization problems of type (1.5) play a basic role in the information-theoretic approach to statistics (see also [7], [9], [13], [17] etc.); they frequently occur also elsewhere, e.g., in the theory of large deviations, cf. Sanov [20], and in statistical physics, as maximization of entropy, cf. Jaynes [10]. In physics, the measure  $R$  is often not a PD;  $R(X)$  may even be infinite. This does not make much difference in most respects, except that in the latter case the integral (1.1) may be negative, even  $-\infty$  (which corresponds to infinite entropy), or undefined.

Let us emphasize that *I-divergence* is not a metric and in general the *I-spheres*  $S(R, \rho)$  do not even define a topology (as a base of the neighborhood system of  $R$ ). This negative statement remains true if  $I(P \parallel Q)$  is replaced by the symmetric divergence  $I(P \parallel Q) + I(Q \parallel P)$ —used already by Jeffreys [11]—or by any reasonable function of  $I(P \parallel Q)$  and  $I(Q \parallel P)$ , see Csiszár [3]. In spite of these discouraging facts, it will be shown that certain analogies exist between properties of PD's and Euclidean geometry, where *I-divergence* plays the role of squared Euclidean distance. In particular, a “geometric” approach will be helpful in the study of *I-projections*, i.e., of the extremum problem (1.5).

In Section 2, using an analogue of the parallelogram identity, we first prove that the *I-projection* always exists if the convex set  $\mathcal{E}$  is closed in the topology of the variation distance

$$(1.6) \quad |P - Q| = \int |p_R - q_R| dR$$

(where  $R$  is any PD with  $P \ll R, Q \ll R$ ). Next we prove a lemma having the geometric interpretation that the PD's with  $\int \log q_R dP = \rho$  form the “tangent hyperplane” of the *I-sphere*  $S(R, \rho)$  at  $Q$ , where  $\rho = I(Q \parallel R) < \infty$ ; for such  $P$ 's

$$(1.7) \quad I(P \parallel R) = I(P \parallel Q) + I(Q \parallel R),$$

which is an analogue of Pythagoras' theorem.

The resulting characterization of *I-projection* will be used in Section 3 to establish a necessary and a sufficient condition on the form of *I-projection* on a set  $\mathcal{E}$  defined by linear constraints of a general type. In case of a finite number of integral constraints or marginal constraints, we obtain a necessary and sufficient characterization of *I-projection*. These results complete the known sufficient conditions following from the minimum discrimination information theorem of Kullback [14] and Kullback and Khairat [18]. As corollaries of independent interest, we arrive at generalizations of known results on existence of bivariate

distributions or nonnegative matrices of a certain product form and with given marginals, see Hobby and Pyke [8] and, e.g., Sinkhorn [21].

Another “geometric” result of Section 2 (asserting the transitivity of  $I$ -projection) is used in Section 3 to prove the convergence of an iterative algorithm for finding the  $I$ -projection, which generalizes the familiar iterative proportional fitting procedure (IPFP) for adjusting a contingency table to given marginal distributions. Though the proof works only for finite  $X$ , it is of more general validity than the known convergence proofs for the IPFP, even if attention is restricted to that case.

Our last result is an existence proof for a case not covered in Section 2.

After having submitted the first version of this paper, the author became aware of related work of Čencov [2]; he has developed a geometry of  $I$ -divergence, looking at it with the reversed order of  $P$  and  $Q$ . Apparently, there is no intersection between his results and ours, except for Theorem 3.3, see the discussion there.

## 2. General “geometric” results on $I$ -projections.

**THEOREM 2.1.** *If the convex set  $\mathcal{E}$  of PD's is variation-closed then each  $R$  with  $S(R, \infty) \cap \mathcal{E} \neq \emptyset$  has an  $I$ -projection on  $\mathcal{E}$ .*

**PROOF.** The idea is similar to the proof of existence of projection in Hilbert space. Pick a sequence  $P_n \in \mathcal{E}$  with  $I(P_n \| R) < \infty$  (in particular,  $P_n \ll R$ ) such that

$$(2.1) \quad I(P_n \| R) \rightarrow \inf_{P \in \mathcal{E}} I(P \| R).$$

Since

$$(2.2) \quad \begin{aligned} I(P_m \| R) + I(P_n \| R) \\ = 2I\left(\frac{P_m + P_n}{2} \| R\right) + I\left(P_m \| \frac{P_m + P_n}{2}\right) + I\left(P_n \| \frac{P_m + P_n}{2}\right) \end{aligned}$$

(this analogue of the parallelogram identity is readily checked by writing all terms as integrals with respect to  $R$ , using (1.2)), where  $(P_m + P_n)/2 \in \mathcal{E}$  by convexity, the last two terms of (2.2) must converge to 0 as  $m, n \rightarrow \infty$ .

Using the inequality

$$(2.3) \quad |P - Q| \leq (2I(P \| Q))^{\frac{1}{2}}$$

proved independently in [4], [12] and [15], one concludes that

$$|P_m - P_n| \leq \left| P_m - \frac{P_m + P_n}{2} \right| + \left| P_n - \frac{P_m + P_n}{2} \right|$$

converges to 0 as  $m, n \rightarrow \infty$  and, consequently,  $P_n$  converges in variation to some PD  $Q$ :

$$(2.4) \quad |P_n - Q| = \int |p_{nr} - q_r| dR \rightarrow 0 \quad (n \rightarrow \infty)$$

(the convergence in variation of the PD's  $P_n \ll R$  to  $Q$  clearly implies  $Q \ll R$ ).

In view of (1.1), from (2.4) follows by Fatou's lemma<sup>1</sup>

$$(2.5) \quad I(Q \parallel R) \leq \liminf_{n \rightarrow \infty} I(P_n \parallel R).$$

As  $\mathcal{E}$  is variation-closed, we have  $Q \in \mathcal{E}$ . On account of (2.1) and (2.5), it follows that  $Q$  is the  $I$ -projection of  $R$  on  $\mathcal{E}$ .

REMARK. The only role of the hypothesis that  $\mathcal{E}$  is variation-closed has been to ensure that the PD  $Q$  with the properties (2.4) and (2.5) belongs to  $\mathcal{E}$ . If this is ensured in some other way, the assertion still holds, see Theorem 3.3.

For any three PD's with  $Q \ll R$  and either of  $I(P \parallel Q) < \infty$  and  $I(P \parallel R) < \infty$  (thus  $P \ll R$ , too), (1.1) and (1.2)—using (1.3) if necessary—give rise to the identity

$$(2.6) \quad \begin{aligned} I(P \parallel R) - I(P \parallel Q) &= \int \left( p_R \log p_R - p_R \log \frac{p_R}{q_R} \right) dR \\ &= \int p_R \log q_R dR = \int \log q_R dP. \end{aligned}$$

Our further results will be based on

LEMMA 2.1. *If  $I(P \parallel Q)$  and  $I(Q \parallel R)$  are finite, the "segment joining  $P$  and  $Q$ " does not intersect the  $I$ -sphere  $S(R, \rho)$  with radius  $\rho = I(Q \parallel R)$ , i.e.,  $I(P_\alpha \parallel R) \geq I(Q \parallel R)$  for each PD*

$$(2.7) \quad P_\alpha = \alpha P + (1 - \alpha)Q, \quad 0 \leq \alpha \leq 1,$$

iff

$$(2.8) \quad \int \log q_R dP \geq I(Q \parallel R).$$

If

$$(2.9) \quad Q = \alpha P + (1 - \alpha)P', \quad 0 < \alpha < 1,$$

then  $I(Q \parallel R) < \infty$  implies  $I(P \parallel R) < \infty$ , and the segment joining  $P$  and  $P'$  does not intersect  $S(R, \rho)$  (with  $\rho = I(Q \parallel R)$ ) iff

$$(2.10) \quad \int \log q_R dP = I(Q \parallel R).$$

PROOF. The hypotheses imply  $P \ll R$ ,  $Q \ll R$ . Let  $p_\alpha = \alpha p_R + (1 - \alpha)q_R$  denote the  $R$ -density of  $P_\alpha$  defined by (2.7) (in particular,  $p_0 = q_R$ ,  $p_1 = p_R$ ). Since  $p_\alpha$  is linear in  $\alpha$  and  $t \log t$  is convex,  $p_\alpha \log p_\alpha$  is a convex function of  $\alpha$  and its difference quotient

$$(2.11) \quad f_\alpha = \frac{1}{\alpha} (p_\alpha \log p_\alpha - q_R \log q_R)$$

converges non-increasingly (as  $\alpha \downarrow 0$ ) to

$$(2.12) \quad \lim_{\alpha \downarrow 0} f_\alpha = \frac{\partial}{\partial \alpha} p_\alpha \log p_\alpha \Big|_{\alpha=0} = (p_R - q_R)(\log q_R + 1).$$

<sup>1</sup> (2.5) is a particular case of a more general lower semicontinuity property of  $I$ -divergence, see Pinsker [19], Section 2.4, Assertion 5.

$f_1 = p_R \log p_R - q_R \log q_R$  is  $R$ -integrable by assumption, thus (1.1), (2.11) and (2.12) give, by the monotone convergence theorem,

$$(2.13) \quad \frac{d}{d\alpha} I(P_\alpha \| R) \Big|_{\alpha=0} = \lim_{\alpha \downarrow 0} \int f_\alpha dR = \int (p_R - q_R)(\log q_R + 1) dR \\ = \int \log q_R dP - I(Q \| R).$$

This proves that if (2.8) does not hold then

$$I(P_\alpha \| R) < I(P_0 \| R) = I(Q \| R) \quad \text{for some } \alpha > 0.$$

The converse is trivial: (2.8) implies  $I(P \| R) \geq I(Q \| R)$  by (2.6), and  $P_\alpha$  also satisfies (2.8) if  $P$  does.

(2.9) with  $I(Q \| R) < \infty$  implies  $P \ll Q \ll R$ ,  $p_R \leq \alpha^{-1} q_R$ , thus by (1.1)  $I(P \| R) < \infty$ , too, and similarly  $I(P' \| Q) < \infty$ . The last assertion of Lemma 2.1 follows from the first one, because (2.8) for both  $P$  and  $P'$  with strict inequality in either case would contradict to (2.9).

Lemma 2.1 means, intuitively, that the "tangent hyperplane" of  $S(R, \rho)$  at  $Q$  consists of the PD's satisfying (2.10); according to (2.6), this is equivalent to (1.7), thus we have an analogue of Pythagoras' theorem. This geometric interpretation is limited, however, to  $P \in S(R, \infty) \cup S(Q, \infty)$ ; if both  $I(P \| R)$  and  $I(Q \| R)$  are infinite, the integral  $\int \log q_R dP$  may or may not be defined and if it is, its value may be arbitrary (the case of  $P \in S(Q, \infty) \setminus S(R, \infty)$ , i.e.,  $I(P \| Q) < I(P \| R) = \infty$  is not contained in Lemma 2.1, either; but then (2.6) applies and shows that (2.8) is trivially valid).

Lemma 2.1 and the identity (2.6) immediately give rise to

**THEOREM 2.2.** *A PD  $Q \in \mathcal{E} \cap S(R, \infty)$  is the  $I$ -projection of  $R$  on the convex set  $\mathcal{E}$  of PD's iff every  $P \in \mathcal{E} \cap S(R, \infty)$  satisfies (2.8) or, equivalently, iff*

$$(2.14) \quad I(P \| R) \geq I(P \| Q) + I(Q \| R) \quad \text{for every } P \in \mathcal{E}.$$

*If the  $I$ -projection  $Q$  is an algebraic inner point of  $\mathcal{E}$  then  $\mathcal{E} \subset S(R, \infty)$  and (2.8) and (2.14) hold with the equality.*

A  $Q \in \mathcal{E}$  is called an algebraic inner point of  $\mathcal{E}$  if for every  $P \in \mathcal{E}$  there exist  $\alpha$  and  $P' \in \mathcal{E}$  satisfying (2.9).

**REMARK.** (2.14) shows, in particular, that if the  $I$ -projection  $Q$  of  $R$  on  $\mathcal{E}$  exists then  $P \ll Q$  for every  $P \in \mathcal{E} \cap S(R, \infty)$ . Thus, if some  $P \in \mathcal{E}$  with  $I(P \| Q) < \infty$  is measure-theoretically equivalent to  $R$ , then so is  $Q$ , as well.

Intuition suggests that if  $\mathcal{E}$  is a linear set of PD's—i.e., if with  $P$  and  $P'$  also  $\alpha P + (1 - \alpha)P'$  belongs to  $\mathcal{E}$  for every real  $\alpha$  for which it is a PD—then  $\mathcal{E}$  always lies in the tangent hyperplane of  $S(R, \rho)$  at  $Q$ , the  $I$ -projection of  $R$  on  $\mathcal{E}$  (with  $\rho = I(Q \| R)$ ), i.e., that the identity (1.7) is valid for every  $P \in \mathcal{E}$ . It will be shown in the next section that this conjecture is not generally true but in the most important cases—in particular, for finite  $X$ —it is. This additivity relation and its consequence, the following transitivity property of  $I$ -projection,

proved for various particular cases by Kullback [14], [17], Ku and Kullback [13], etc., is very essential for informational statistical analysis.

**THEOREM 2.3.** *Let  $\mathcal{E}$  and  $\mathcal{E}_1 \subset \mathcal{E}$  be convex sets of PD's, let  $R$  have  $I$ -projection  $Q$  on  $\mathcal{E}$  and  $I$ -projection  $Q_1$  on  $\mathcal{E}_1$ , and suppose that the identity (1.7) holds for every  $P \in \mathcal{E}$ . Then  $Q_1$  is the  $I$ -projection of  $Q$  on  $\mathcal{E}_1$ .*

**PROOF.** Applying (2.14) with  $Q_1$  in the role of  $Q$  and (1.7) with  $Q_1$  in the role of  $P$ , we have for  $P \in \mathcal{E}_1$

$$(2.15) \quad I(P \parallel R) \geq I(P \parallel Q_1) + I(Q_1 \parallel R) = I(P \parallel Q_1) + I(Q_1 \parallel Q) + I(Q \parallel R).$$

Comparing (2.15) with (1.7),  $I(Q \parallel R)$  cancels out, yielding

$$(2.16) \quad I(P \parallel Q) \geq I(P \parallel Q_1) + I(Q_1 \parallel Q) \quad \text{for every } P \in \mathcal{E}_1.$$

Theorem 2.3 completes the geometric results on  $I$ -divergence needed for our purposes. Of course, intuition should be used with caution. For example, if  $R$  has  $I$ -projection  $Q$  on a convex set  $\mathcal{E}$  of PD's, it does not follow that the elements of the "joining segment" of  $Q$  and  $R$  have the same  $I$ -projection on  $\mathcal{E}$ .

**3. Minimizing  $I$ -divergence under linear constraints.** A general formulation of a useful result known as minimum discrimination information theorem (Kullback [14], Kullback and Khairat [18]) is the following: For any (not necessarily convex) set  $\mathcal{E}$  of PD's, if there exists a  $Q \in \mathcal{E}$  with  $R$ -density  $c \exp g(x)$  where  $\int g dP_1 = \int g dP_2 < \infty$  for any  $P_1, P_2 \in \mathcal{E}$ , then  $I(Q \parallel R) = \min_{P \in \mathcal{E}} I(P \parallel R)$ ; more exactly, in this case

$$(3.1) \quad I(P \parallel R) = I(P \parallel Q) + I(Q \parallel R) \quad \text{for all } P \in \mathcal{E}.$$

Observe that this immediately follows from the identity (2.6).

Two particular cases deserve main attention:

(A)  $\mathcal{E}$  is defined by constraints of form  $\int f_i dP = a_i, i = 1, \dots, k$ . Then, if a  $Q \in \mathcal{E}$  with

$$(3.2) \quad q_R(x) = c \exp \sum_{i=1}^k t_i f_i(x)$$

exists, it is the  $I$ -projection of  $R$  on  $\mathcal{E}$  and (3.1) holds.

(B)  $(X, \mathcal{X}) = (X_1, \mathcal{X}_1) \times (X_2, \mathcal{X}_2)$  and  $\mathcal{E}$  consists of the PD's  $P$  with given marginals  $P_i$  on  $(X_i, \mathcal{X}_i), i = 1, 2$ . Then, if a  $Q \in \mathcal{E}$  with

$$(3.3) \quad q_R(x_1, x_2) = a(x_1)b(x_2), \quad \log a \in L_1(P_1), \quad \log b \in L_1(P_2)$$

exists, it is the  $I$ -projection of  $R$  on  $\mathcal{E}$  and, again, (3.1) holds.

Our next aim is to complete the mentioned results for cases (A) and (B). We shall not explicitly consider the equally important case of PD's on a multiple product space with given marginals of certain (arbitrary) types, since the extension of our results from case (B) to that case is trivial. For example, if  $(X, \mathcal{X}) = \prod_{i=1}^4 (X_i, \mathcal{X}_i)$  and  $\mathcal{E}$  consists of the PD's with given marginals (of types shown by the indices)  $P_{123}, P_{124}$  and  $P_{34}$ , say, then the extension of Corollary

3.1 below is that a  $Q \in \mathcal{E}$  is the  $I$ -projection of  $R$  on  $\mathcal{E}$  iff

$$q_R(x_1, x_2, x_3, x_4) = a(x_1, x_2, x_3)b(x_1, x_2, x_4)c(x_3, x_4)$$

with  $\log a \in L_1(P_{123})$ ,  $\log b \in L_1(P_{124})$ ,  $\log c \in L_1(P_{34})$  except, possibly, for a set  $N$  where  $q_R$  vanishes and  $P(N) = 0$  whenever  $P \in \mathcal{E}$ ,  $I(P \parallel R) < \infty$ ; then (3.1) holds, too.

The following theorem concerns sets of PD's defined by linear constraints of a general type. Since no existence assertions will be made, we need not formally exclude even  $\mathcal{E} = \emptyset$ , i.e., contradicting constraints.

**THEOREM 3.1.** *Let  $\{f_\gamma\}_{\gamma \in \Gamma}$  be an arbitrary set of real-valued  $\mathcal{X}$ -measurable functions on  $X$  and  $\{a_\gamma\}_{\gamma \in \Gamma}$  be real constants. Let  $\mathcal{E}$  be the set of all those PD's  $P$  on  $(X, \mathcal{X})$  for which the integrals  $\int f_\gamma dP$  exist and equal  $a_\gamma$  ( $\gamma \in \Gamma$ ). Then, if a PD  $R$  has  $I$ -projection  $Q$  on  $\mathcal{E}$ , its  $R$ -density is of form*

$$(3.4) \quad \begin{aligned} q_R(x) &= c \exp g(x) && \text{if } x \notin N \\ &= 0 && \text{if } x \in N \end{aligned}$$

where  $N$  has  $P(N) = 0$  for every  $P \in \mathcal{E} \cap S(R, \infty)$  and  $g$  belongs to the closed subspace of  $L_1(Q)$  spanned by the  $f_\gamma$ 's. On the other hand, if a  $Q \in \mathcal{E}$  has  $R$ -density of form (3.4) where  $g$  belongs to the linear space spanned by the  $f_\gamma$ 's (without closure) then  $Q$  is the  $I$ -projection of  $R$  on  $\mathcal{E}$  and (3.1) holds.

**COROLLARY 3.1.** *In case (A) or (B) above, a  $Q \in \mathcal{E}$  is the  $I$ -projection of  $R$  on  $\mathcal{E}$  iff  $q_R$  is of form (3.2) or (3.3), respectively, except possibly for a set  $N$  where  $q_R$  vanishes and  $P(N) = 0$  for every  $P \in \mathcal{E} \cap S(R, \infty)$ ; in both cases, the identity (3.1) holds. If, in particular, some  $P \in \mathcal{E}$  with  $I(P \parallel R) < \infty$  is measure-theoretically equivalent to  $R$  then (3.2) or (3.3) is necessary and sufficient for  $Q$  to be the  $I$ -projection of  $R$  on  $\mathcal{E}$ .*

Before giving the proof, let us show by an example that for the  $I$ -projection on a set  $\mathcal{E}$  defined by linear constraints the identity (3.1) is not generally true (contrary to geometric intuition) and neither the necessary nor the sufficient condition of Theorem 3.1 is both necessary and sufficient, in general.

**EXAMPLE.** Let  $X$  be the unit interval,  $\mathcal{X}$  the Borel  $\sigma$ -algebra and  $Q$  the Lebesgue measure. Let  $\mathcal{E}$  be the set of PD's satisfying  $\int f_n dP = \frac{1}{4}$ ,  $n = 1, 2, \dots$ , where

$$(3.5) \quad \begin{aligned} f_n(x) &= 1 + \frac{n^{\frac{1}{2}}}{4} && \text{if } 0 < x < \frac{1}{4n} \\ &= 1 && \text{if } \frac{1}{4n} \leq x < \frac{1}{4} \\ &= -\frac{1}{4n^{\frac{1}{2}}} && \text{if } \frac{1}{4} \leq x < \frac{1}{2} \\ &= 0 && \text{if } \frac{1}{2} \leq x < 1. \end{aligned}$$

Let the PD  $R$  be determined by the condition  $q_R(x) = c \exp g(x)$  where

$$(3.6) \quad \begin{aligned} g(x) &= -\lim_{n \rightarrow \infty} f_n(x) = -1 && \text{if } 0 < x < \frac{1}{4} \\ &= 0 && \text{if } \frac{1}{4} \leq x < 1. \end{aligned}$$

Then  $Q \in \mathcal{E}$ , and on account of Fatou's lemma

$$(3.7) \quad \begin{aligned} \int \log q_R dP &= \log c - \int \lim_{n \rightarrow \infty} f_n dP \\ &\geq \log c - \frac{1}{4} = I(Q \| R) \end{aligned}$$

for all  $P \in \mathcal{E}$ . This means, by Theorem 2.2, that  $Q$  is the  $I$ -projection of  $R$  on  $\mathcal{E}$ . It is easy to find  $P \in \mathcal{E}$  for which in (3.7) the strict inequality holds, e.g. the PD with  $Q$ -density

$$(3.8) \quad \begin{aligned} p_Q(x) &= \frac{1}{5x^{\frac{1}{2}}} && \text{if } 0 < x < \frac{1}{4} \\ &= 0 && \text{if } \frac{1}{4} \leq x < \frac{3}{5} \\ &= 2 && \text{if } \frac{3}{5} \leq x < 1. \end{aligned}$$

Thus (3.1) is false in this case; in particular,  $Q$  cannot meet the sufficient condition of Theorem 3.1. If  $g(x)$  is given the opposite sign and  $R$  is defined accordingly, we obtain  $\int \log q_R dP < I(Q \| R)$  for the  $P$  defined by (3.8); this means that  $Q$  cannot be the  $I$ -projection of  $R$  on  $\mathcal{E}$ , showing that the necessary condition of Theorem 3.1 is not sufficient.

**PROOF OF THEOREM 3.1.** If  $Q$  is the  $I$ -projection of  $R$  on  $\mathcal{E}$  then for  $N = \{x : q_R(x) = 0\}$  necessarily  $P(N) = 0$  for each  $P \in \mathcal{E} \cap S(R, \infty)$ ; see the remark to Theorem 2.2.

Let  $\mathcal{E}' \subset \mathcal{E}$  be the set of PD's  $P \in \mathcal{E}$  with  $p_Q(x) \leq 2$ . If  $P \in \mathcal{E}'$ , there is a  $P' \in \mathcal{E}'$  with  $p_{Q'}(x) = 2 - p_Q(x)$ , and with it  $Q = (P + P')/2$ ; thus  $Q$  is an algebraic inner point of  $\mathcal{E}'$ . Applying Theorem 2.2 to  $\mathcal{E}'$  instead of  $\mathcal{E}$  we obtain  $\int \log q_R dP = I(Q \| R)$ , i.e.,

$$(3.9) \quad \int \log q_R (p_Q - 1) dQ = 0 \quad \text{for all } P \in \mathcal{E}' .$$

But for any  $\mathcal{E}$ -measurable function  $h$  with  $|h(x)| \leq 1$  such that

$$(3.10) \quad \int h dQ = 0 \quad \text{and} \quad \int f_\gamma h dQ = 0 \quad \text{for each } \gamma \in \Gamma ,$$

there exists a  $P \in \mathcal{E}'$  with  $p_Q = 1 + h$ . Thus (3.9) gives

$$(3.11) \quad \int \log q_R h dQ = 0$$

for all such  $h$  and therefore also for all  $h \in L_\infty(Q)$  satisfying (3.10).

Hence follows that  $\log q_R$  belongs to the (closed) subspace of  $L_1(Q)$  spanned by 1 and the  $f_\gamma$ 's. In fact, were this not the case, in view of the Hahn-Banach theorem ([22] page 106) there would exist a bounded linear functional on  $L_1(Q)$  vanishing on the mentioned subspace but not at  $\log q_R$ ; since the dual of  $L_1(Q)$  is  $L_\infty(Q)$  ([22] page 115), this is a contradiction. This proves the first assertion of Theorem 3.1.

The second part is much easier. Suppose that  $q_R$  is of the stated form. Since  $g$  is a finite linear combination of  $f_\gamma$ 's,  $\int g dP$  is constant for  $P \in \mathcal{E}$  and

$$(3.12) \quad \int \log q_R dP = \log c + \int g dP = \text{const} = I(Q \| P)$$

for  $P \in \mathcal{E}$ ,  $P \ll Q$ . But for  $P \in \mathcal{E}$  both  $I(P \| R) < \infty$  (by hypothesis) and  $I(P \| Q) < \infty$  (by definition) imply  $P \ll Q$ ; thus (3.1) follows from (2.6) and (3.12).

To prove the corollary, observe that case (B) does fit into the considered model, taking for  $f_\gamma$ 's the  $P_i$ -integrable functions  $f(x_i)$ ,  $i = 1, 2$  (looking at them as functions of  $(x_1, x_2)$ ). Theorem 3.1 clearly gives a necessary and sufficient condition on  $q_R$  and guarantees the validity of (3.1) for the  $I$ -projection  $Q$  if the linear space spanned by the  $f_\gamma$ 's is closed in  $L_1(P)$  for each  $P \in \mathcal{E}$ . But the latter hypothesis is fulfilled in both cases (A) and (B), completing the proof.

Theorem 3.1 and its corollary leaves the question of existence of  $I$ -projection open. If  $\mathcal{E}$  is variation-closed, as in the case of bounded  $f_\gamma$ 's or in case (B), Theorem 2.1 guarantees the existence provided that  $\mathcal{E} \neq \emptyset$  and  $I(P \| R) < \infty$  for some  $P \in \mathcal{E}$ . For case (A) with not bounded  $f_i$ 's, see Theorem 3.3 below.

As a consequence of Corollary 3.1 and Theorem 2.1 we obtain

**COROLLARY 3.2.** *To given PD's  $P_i$  on  $(X_i, \mathcal{X}_i)$ ,  $i = 1, 2$  and  $R$  on  $(X_1 \times X_2, \mathcal{X}_1 \times \mathcal{X}_2)$ , there exists a PD  $Q$  on the product space with marginals  $P_1$  and  $P_2$  and with  $R$ -density of form  $a(x_1)b(x_2)$ ,  $\log a \in L_1(P_1)$ ,  $\log b \in L_1(P_2)$  iff there exists any  $P$  measure-theoretically equivalent to  $R$  which has the prescribed marginals and satisfies  $I(P \| R) < \infty$ .*

Considering  $R \ll P_1 \times P_2$  with density  $f(x_1, x_2)$  and using  $P = P_1 \times P_2$  in Corollary 3.2, we obtain for the existence of a PD with marginals  $P_1$  and  $P_2$  and having  $P_1 \times P_2$ -density of form  $a(x_1)b(x_2)f(x_1, x_2)$  the sufficient condition  $\log f \in L_1(P_1 \times P_2)$ . It is interesting to compare this with a result of Hobby and Pyke [8]; their theorem, when specialized to our problem, gives the sufficient condition  $0 < a \leq f(x_1, x_2) \leq K$ .

Specializing Corollary 3.2 to finite  $X_1$  and  $X_2$ , we obtain

**COROLLARY 3.3.** *Let  $A$  be an  $m \times n$  matrix with nonnegative elements. For the existence of positive diagonal matrices  $D_1$  and  $D_2$  such that the row and column sums of  $D_1 A D_2$  be given positive numbers, it is necessary and sufficient that some  $B$  with nonnegative elements and with the given row and column sums have the same zero entries as  $A$  (if any).*

**PROOF.** Without any loss of generality, the elements of  $A$  and both the given row and column sums may be assumed to add up to one. Then  $A$  defines a PD  $R$  on  $X_1 \times X_2$  and  $D_1 A D_2$  defines a PD having  $R$ -density of form (3.3). Since for PD's on finite sets  $P \ll R$  implies  $I(P \| R) < \infty$ , the assertion follows from Corollary 3.2.

Corollary 3.3 solves a matrix-theoretic problem, partial solutions of which have been given by many authors. Sinkhorn [21] has shown that the positivity

of  $A$  is a sufficient condition by proving the convergence of the iterative proportional fitting procedure (IPFP) dating back to Deming and Stephan [5]. This IPFP and its extensions are widely used in the analysis of contingency tables. Of the extensive literature of the subject we mention here only Ireland and Kullback [9], Ku and Kullback [13] and Fienberg [6]; further references may be found there.<sup>2</sup>

The IPFP for adjusting a PD  $R$  given on a finite product space to  $k$  marginal constraints, i.e., to given marginal distributions of arbitrary types, consists in the successive calculation of PD's  $Q_n$  on the product space starting from  $Q_0 = R$ : to obtain  $Q_n$ , the probabilities of  $Q_{n-1}$  are multiplied by the ratios of the corresponding marginal probabilities of the  $n$ th constraint and of  $Q_{n-1}$ . Here the constraints are looked at cyclically repeated. As shown by Ireland and Kullback [9],  $Q_n$  is just the  $I$ -projection of  $Q_{n-1}$  on  $\mathcal{E}_n$ , where  $\mathcal{E}_i$  is the set of PD's satisfying the  $i$ th constraint, and  $Q = \lim_{n \rightarrow \infty} Q_n$  (if it exists) is the  $I$ -projection of  $R$  on  $\mathcal{E} = \bigcap_{i=1}^k \mathcal{E}_i$ , the set of PD's satisfying all  $k$  marginal constraints. Kullback [16] generalized the method for the non-discrete case, too.

Motivated by the approach of Ireland and Kullback [9], we are going to formulate the procedure in a general setup and prove convergence to the required  $I$ -projection, provided that  $X$  is a finite set. Unlike previous convergence proofs for the IPFP (see Fienberg [6] and the references there), we shall not need any assumption on the positivity of the probabilities of  $R$ . It should be noted that the convergence proof in [9] is incomplete since formula (4.38) does not imply (4.39); in [16] there is a similar flaw.

**THEOREM 3.2.** *Let  $\mathcal{E}_1, \dots, \mathcal{E}_k$  be arbitrary linear sets of PD's on a finite set  $X$  with  $\mathcal{E} = \bigcap_{i=1}^k \mathcal{E}_i \neq \emptyset$ , let  $R$  be any PD to which there exists  $P \in \mathcal{E}$  with  $P \ll R$ , and define  $Q_1, Q_2, \dots$  recursively by letting  $Q_n$  be the  $I$ -projection of  $Q_{n-1}$  on  $\mathcal{E}_n$ ,  $n = 1, 2, \dots$  where  $Q_0 = R$  and*

$$(3.13) \quad \mathcal{E}_n = \mathcal{E}_i \quad \text{if } n = mk + i, \quad 1 \leq i \leq k.$$

*Then  $Q_n$  converges (pointwise or, equivalently, in variation) to the  $I$ -projection  $Q$  of  $R$  on  $\mathcal{E}$ .*

**PROOF.** Any linear set  $\mathcal{E}$  of PD's on a finite set  $X$  of size  $r$ , say, can be looked at as the intersection of a linear subset of  $E^r$  with the simplex representing the PD's on  $X$ . Hence  $\mathcal{E}$  is closed and can be defined by a finite number of linear constraints. In view of Theorem 2.1 and Corollary 3.1, the  $I$ -projection of  $R$  on such an  $\mathcal{E}$  always exists if  $I(P \parallel R) < \infty$ —now equivalent to  $P \ll R$ —for some  $P \in \mathcal{E}$ , and then (3.1) holds, as well.

Under the hypotheses of Theorem 3.2, it follows that the  $I$ -projections  $Q_1, Q_2, \dots$  and  $Q$  exist and  $I(P \parallel Q_{n-1}) = I(P \parallel Q_n) + I(Q_n \parallel Q_{n-1})$  for any  $P \in \mathcal{E}_n$ ,

---

<sup>2</sup> The iterative algorithms suggested in [17] apparently do not belong to the class of generalizations of the IPFP considered below. But also for the problems considered there, it is straightforward to give convergent iterations within the framework of Theorem 3.2.

$n = 1, 2, \dots$ . Setting  $P = Q$ , in particular, we obtain by induction

$$(3.14) \quad I(Q \| R) = I(Q \| Q_n) + \sum_{i=1}^n I(Q_i \| Q_{i-1}), \quad n = 1, 2, \dots$$

Since  $X$  is finite, each subsequence of  $Q_n$  contains a convergent subsequence; it suffices to show that  $Q_{n_l} \rightarrow Q'$  implies  $Q' = Q$ . First verify  $Q' \in \mathcal{E}$ . We have from (3.14)

$$\sum_{i=1}^{\infty} I(Q_i \| Q_{i-1}) \leq I(Q \| R) < \infty$$

thus  $I(Q_n \| Q_{n-1}) \rightarrow 0$ , implying  $|Q_n - Q_{n-1}| \rightarrow 0$  by (2.3). Thus  $Q_{n_l+1}, \dots, Q_{n_l+k}$  also converge to  $Q'$  as  $l \rightarrow \infty$ . Since these PD's belong to (a cyclic permutation of) the closed sets  $\mathcal{E}_1, \dots, \mathcal{E}_k$  respectively, see (3.13), we conclude  $Q' \in \bigcap_{i=1}^k \mathcal{E}_i = \mathcal{E}$ .

Repeated application of Theorem 2.3 shows that  $Q$  is the  $I$ -projection on  $\mathcal{E}$  of  $Q_1, \dots, Q_n, \dots$ , as well, thus

$$(3.15) \quad I(P \| Q_n) = I(P \| Q) + I(Q \| Q_n) \quad \text{for all } P \in \mathcal{E},$$

$n = 1, 2, \dots$ . Applying this to  $P = Q'$ , we obtain  $I(Q' \| Q) = 0$ , i.e.,  $Q' = Q$ , because for finite  $X$   $Q_{n_l} \rightarrow Q'$  implies  $I(Q' \| Q_{n_l}) \rightarrow 0$ . The proof is complete.

Finally, let us return to the problem of existence of  $I$ -projection in case (A), if the  $f_i$ 's are not necessarily bounded. One possible approach is to prove in a direct way that there exists a  $Q \in \mathcal{E}$  with  $R$ -density (3.2). This is not easy but has been done under fairly general conditions by Čencov [2], Theorem 23.1. Here we show how the method of Theorem 2.1 applies to this case. Our hypothesis will be that

$$(3.16) \quad T_R = \{(t_1, \dots, t_n) : \exp \sum_{i=1}^k t_i f_i(x) \text{ is } R\text{-integrable}\}$$

is an open set in  $E^k$ ; this clearly implies that  $f_i(x) \exp \sum_{i=1}^k t_i f_i(x)$  is  $R$ -integrable for every  $(t_1, \dots, t_k) \in T_R$ ,  $i = 1, \dots, k$ . (In the first version of this paper,  $T_R = E^k$  was assumed. The strengthening has been inspired by Čencov's result, *loc. cit.*, which implies the existence of  $I$ -projection even under a slightly weaker hypothesis.)

**THEOREM 3.3.** *Let  $\mathcal{E}(a_1, \dots, a_k)$  be the set of PD's satisfying  $\int f_i dP = a_i$ ,  $i = 1, \dots, k$  and let  $A_R$  be the set of points  $(a_1, \dots, a_k) \in E^k$  for which  $\mathcal{E}(a_1, \dots, a_k)$  contains some  $P$  with  $I(P \| R) < \infty$ . Then, supposing that  $T_R$  is open, the  $I$ -projection of  $R$  on  $\mathcal{E}(a_1, \dots, a_k)$  exists for each inner point  $(a_1, \dots, a_k)$  of  $A_R$ , and its  $R$ -density is of form (3.2).*

**REMARKS.** It can be shown that the interior of  $A_R$  coincides with that of the convex hull of the support of  $R'$ , the image of  $R$  in  $E^k$  at the mapping  $x \rightarrow (f_1(x), \dots, f_k(x))$ . Thus, assuming that the  $f_i$ 's are linearly independent mod  $R$ , the interior of  $A_R$  is nonvoid. If  $(a_1, \dots, a_k) \in A_R$  is on the boundary of  $A_R$ , typically there still exists the  $I$ -projection of  $R$  on  $\mathcal{E}(a_1, \dots, a_k)$  but its  $R$ -density vanishes on a set  $N$  of  $R(N) > 0$ . These problems will not be entered here.

We shall need the following lemma of some independent interest.

LEMMA 3.1. For any (measurable) function  $f(x)$  for which  $e^{tf(x)}$  is  $Q$ -integrable if  $|t|$  is sufficiently small,  $I(P_n \| Q) \rightarrow 0$  implies  $\int f dP_n \rightarrow \int f dQ$ .

PROOF. Let  $p_n$  denote the  $Q$ -density of  $P_n$ ; it surely exists if  $I(P_n \| Q) < \infty$ . In view of (2.3),  $I(P_n \| Q) \rightarrow 0$  implies  $\int |p_n - Q| = \int |p_n - 1| dQ \rightarrow 0$ , hence on  $A_k = \{x : |f(x)| \leq K\}$  we have  $\int_{A_k} f dP_n \rightarrow \int_{A_k} f dQ$ . Thus it suffices to show that to any  $\varepsilon > 0$  there exists  $K$  such that

$$(3.17) \quad \limsup_{n \rightarrow \infty} \int_{X \setminus A_k} |f| dP_n = \limsup_{n \rightarrow \infty} \int_{X \setminus A_k} |f| p_n dQ < \varepsilon.$$

But  $I(P_n \| Q) = \int p_n \log p_n dQ \rightarrow 0$  implies  $\lim_{n \rightarrow \infty} \int_A p_n \log p_n dQ = 0$  for every  $A \in \mathcal{L}$  (apply Fatou's lemma to both  $A$  and  $X \setminus A$ ). Choosing  $t > 0$  and  $K$  to satisfy  $\int_{X \setminus A_k} e^{t|f|} dQ < \varepsilon t$ , (3.17) follows from the inequality  $ab < a \log a + e^b$  (see [1] Section 15), substituting  $a = p_n(x)$ ,  $b = t|f(x)|$ .

PROOF OF THEOREM 3.3. On account of the convexity of  $I(P \| R)$  in  $P$ ,  $A_R$  is a convex set and

$$(3.18) \quad F(a_1, \dots, a_k) = \inf_{P \in \mathcal{S}(a_1, \dots, a_k)} I(P \| Q)$$

is a finite valued convex function on  $A_R$ . Hence, if  $(a_1, \dots, a_k)$  is an inner point of  $A_R$ , there exists  $(t_1, \dots, t_k)$  such that

$$(3.19) \quad F(b_1, \dots, b_k) \geq F(a_1, \dots, a_k) + \sum_{i=1}^k t_i (b_i - a_i) \quad \text{for all } (b_1, \dots, b_k) \in A_R.$$

First we show that  $(t_1, \dots, t_k) \in T_R$ , see (3.16).

Let  $P_n \in \mathcal{S}(a_1, \dots, a_k)$ ,  $I(P_n \| R) \rightarrow F(a_1, \dots, a_k)$ ; then  $P_n$  converges in variation to some  $Q$  by the proof of Theorem 2.1. Let  $f_i^{(n)}(x) = f_i(x)$  if  $t_i f_i(x) \leq K_n$  and  $f_i^{(n)}(x) = 0$  else, where  $K_n \uparrow \infty$ , and let  $Q_n$  be the PD with  $R$ -density

$$(3.20) \quad q_{nR}(x) = c_n \exp \sum_{i=1}^k t_i f_i^{(n)}(x).$$

From (3.20) and (1.1) follows

$$(3.21) \quad I(Q_n \| R) = \int \log q_{nR} dP_n + \sum_{i=1}^k t_i (\int f_i^{(n)} dQ_n - \int f_i^{(n)} dP_n).$$

Since  $(0_1, \dots, 0) \in T_R$  and  $T_R$  is open, the  $f_i$ 's are  $R$ -integrable and thus  $Q_n$ -integrable, too; it follows that for large  $n$   $\int f_i^{(n)} dQ_n$  is arbitrarily close to  $\int f_i dQ_n = b_i^n$ , say (note, that the sequence  $c_n$  is non-increasing). Choosing the  $K_n$ 's properly, also  $\int f_i^{(n)} dP_n$  will be close to  $\int f_i dP_n = a_i$  if  $n$  is large, and then the identities (3.21) and (2.6) compared with the inequality (3.19) (with  $b_i^n$  in the role of  $b_i$ ) give rise to  $I(P_n \| Q_n) \rightarrow 0$ .

On account of (2.3), it follows that the  $Q_n$ 's with  $R$ -density (3.20) also converge in variation to  $Q$ , hence the latter has  $R$ -density (3.2); in particular,  $(t_1, \dots, t_k) \in T_R$ .

Setting  $b_i = \int f_i dQ$ , similarly to (3.21) we have

$$(3.22) \quad I(Q \| R) = \int \log q_R dP_n + \sum_{i=1}^k t_i (b_i - a_i),$$

whence—again by (2.6) and (3.19)—we obtain that  $I(P_n \| Q) \rightarrow 0$ . Using the assumption that  $T_R$  is an open set, Lemma 3.1 gives  $\int f_i dQ = \lim_{n \rightarrow \infty} \int f_i dP_n = a_i$ ,  $i = 1, \dots, k$ . The proof is complete.

REMARK. It follows that  $F(a_1, \dots, a_k)$ —see (3.18)—is differentiable at every inner point of  $A_R$  and  $\text{grad } F(a_1, \dots, a_k) = (t_1, \dots, t_k)$  is just the parameter vector in (3.2) for the  $I$ -projection  $Q$  of  $R$  on  $\mathcal{E}(a_1, \dots, a_k)$ .

## REFERENCES

- [1] BECKENBACH, E. F. and BELLMAN, R. (1961). *Inequalities*. Springer-Verlag, Berlin.
- [2] ČENCOV, N. N. (1972). *Statistical Decision Rules and Optimal Decisions* (in Russian). Nauka, Moskow.
- [3] CSISZÁR, I. (1964). Über topologische und metrische Eigenschaften der relativen Information der Ordnung  $\alpha$ . *Trans. Third Prague Conference of Information Theory etc.* Publishing House of the Czechoslovak Academy of Sciences, Prague, 63–73.
- [4] CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318.
- [5] DEMING, W. E. and STEPHAN, F. F. (1940). On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11** 427–444.
- [6] FIENBERG, S. (1970). An iterative procedure for estimation in contingency tables. *Ann. Math. Statist.* **41** 907–917.
- [7] GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34** 911–934.
- [8] HOBBY, C. and PYKE, R. (1965). Doubly stochastic operators obtained from positive operators. *Pacific J. Math.* **15** 153–157.
- [9] IRELAND, C. T. and KULLBACK, S. (1968). Contingency tables with given marginals. *Biometrika* **55** 179–188.
- [10] JAYNES, E. T. (1957). Information theory and statistical mechanics. *Phys. Rev.* **106** 620–630.
- [11] JEFFREYS, H. (1948). *Theory of Probability*. 2nd ed. Clarendon Press, Oxford.
- [12] KEMPERMAN, J. H. B. (1967). On the optimum rate of transmitting information. *Probability and Information Theory*. Lecture Notes in Mathematics, Springer-Verlag, 126–169.
- [13] KU, H. H. and KULLBACK, S. (1968). Interaction in multidimensional contingency tables: an information theoretic approach. *J. Res. Nat. Bur. Standards* **72** 159–199.
- [14] KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York.
- [15] KULLBACK, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*. **IT-13** 126–127.
- [16] KULLBACK, S. (1968). Probability densities with given marginals. *Ann. Math. Statist.* **39** 1236–1243.
- [17] KULLBACK, S. (1971). Marginal homogeneity of multidimensional contingency tables. *Ann. Math. Statist.* **42** 594–606.
- [18] KULLBACK, S. and KHAIRAT, M. A. (1966). A note on minimum discrimination information. *Ann. Math. Statist.* **37** 279–280.
- [19] PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco.
- [20] SANOV, I. N. (1957). On the probability of large deviations of random variables (in Russian). *Mat. Sbornik* **42** 11–44.
- [21] SINKHORN, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Monthly* **74** 402–405.
- [22] YOSIDA, K. (1965). *Functional Analysis*. Springer-Verlag, Berlin.

MATHEMATICAL INSTITUTE  
 HUNGARIAN ACADEMY OF SCIENCES  
 1053 BUDAPEST, REÁLTANODA-U. 13–15  
 HUNGARY