A Guide to M.D.I. Statistics for Planning and Management Model Building

By: **Fred Young Phillips**

Date: July 1981

Abstract:

This monograph is intended as a practical guide to business applications of the theory of discrimination information statistics as developed by Kullback (1959) and Charnes and Cooper (1975 et seq.). A guide to modeling and computation methods is presented, with references to published applications and a discussion of their implications for business and planning. These implications are developed by means of detailed examples showing MDI to be a practically workable unifying principle for the analysis of demand and market structure. Some applications in other management areas are also noted.

Keywords: MDI statistics; minimum discrimination information statistics; business modeling; statistical information theory

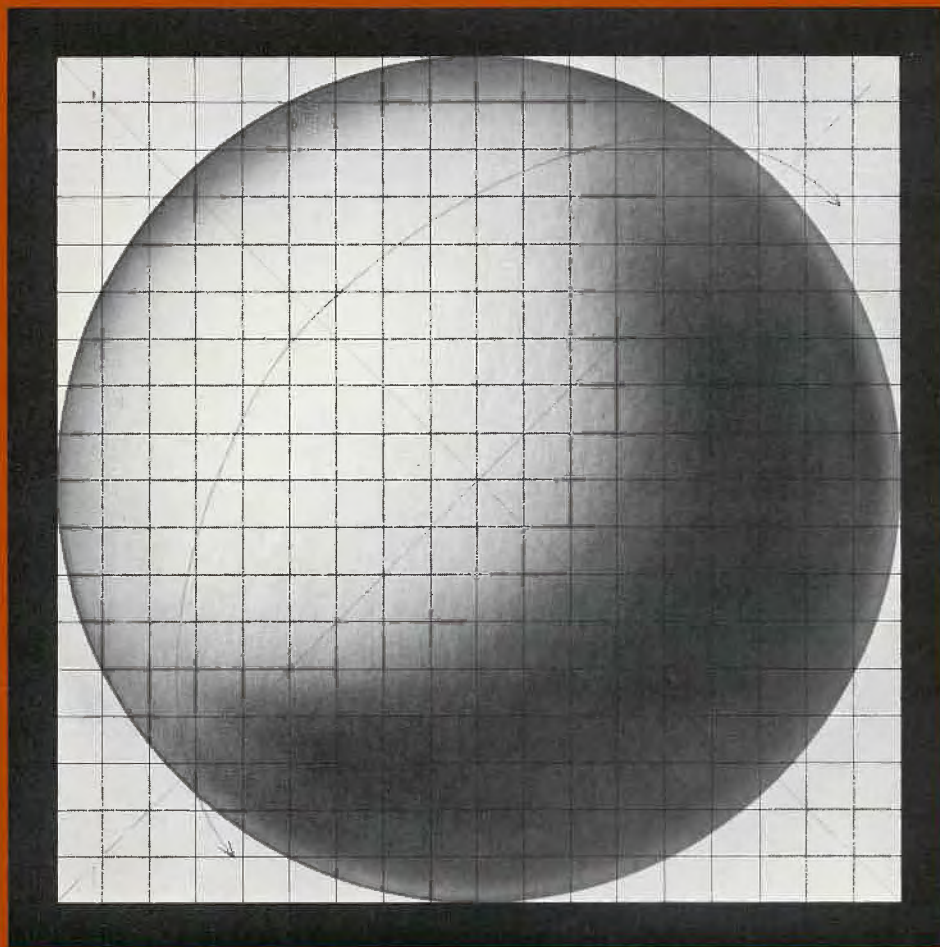**Institute for Constructive Capitalism Technical Series no. 4**

IC² INSTITUTE

THE UNIVERSITY OF TEXAS AT AUSTIN

IC² Institute, The University of Texas at Austin
http://ic2.utexas.edu

FOUR:

A GUIDE TO M.D.I. STATISTICS FOR PLANNING AND MANAGEMENT MODEL BUILDING

By Fred Young Phillips

TECHNICAL SERIES

The Institute for Constructive Capitalism
The University of Texas at Austin

A Guide to M.D.I. Statistics
for Planning and
Management Model Building

Fred Young Phillips
Market Research Corporation of America

July 1981

ABSTRACT

This monograph is intended as a practical guide to business applications of the theory of discrimination information statistics as developed by Kullback (1959) and Charnes and Cooper (1975 et seq.). A guide to modelling and computation methods is presented, with references to published applications and a discussion of their implications for business and planning. These implications are developed by means of detailed examples showing MDI to be a practically workable unifying principle for the analysis of demand and market structure. Some applications in other management areas are also noted.

CONTENTS

I. INTRODUCTION

The purpose of this monograph is to provide a guide to the construction, use, and interpretation of management models using Khintchine-Kullback-Leibler statistics (often called minimum discrimination information (MDI) statistics).

Despite their formidable name, these statistics and their accompanying theory promise to make many aspects of modelling for managerial information and control a much easier task. Several familiar model types are encompassed by the theory (formally known as statistical information theory), and these models fall out of an MDI analysis as special cases. Useful connections have been established between information theoretic statistics and the management tools of linear and nonlinear programming. These linkages provide for straightforward computation and interpretation of MDI models and extend the range of management situations that can be treated by statistical methods.

MDI methods are also useful for developing analytic structures of markets (for public goods or private goods, as the examples will show), for testing whether changes have occurred in these structures over time, and for measuring the divergence of these structures from ideal or targeted states. Examples and references to follow include highway planning problems; economic input-output analysis; market area determination, market segmentation, reach-frequency-volume analysis, and brand switching studies for consumer goods; resource markets with cartels; and income inequality measurement. Additional examples illustrate the application of MDI to production management and planning

1

(estimation of production functions and of manufacturing tolerances), to business forecasting and the managerial uses of forecasts, and to competitive analysis via game theory.

MDI methods provide a unified and simplified treatment of individual choice and aggregate demand analysis, which allows for the analysis of market behavior (Charnes, Cooper, Learner and Phillips (1980)). They also play a unifying role in statistical inference in general, leading to many non-market-oriented business and planning applications, of which some examples are provided below. Further, the methods are of great interest to management scientists since they are applicable both to aggregated and disaggregated data.

In recent years a number of established management and planning models have been shown to be equivalent to MDI formulations. These include models for analyzing brand shifting behavior, for updating inter-industry matrices, and for estimating traffic patterns. Further equivalencies of this type, as well as original applications of MDI to business analysis, are now being published regularly. Theoretical considerations discussed by Learner and Phillips (1980) reinforce this evidence that the properties of MDI statistics are particularly appropriate for management and decision analysis.

The thrust of this paper is methodological. References are given to theoretical and historical sources and to fuller accounts of applications. The discussion is organized into five sections, examining the procedural and interpretive implications of each of the areas that combine to make MDI statistics a powerful modelling tool: Mathematical statistics, mathematical programming, business planning, and computational algorithms. The paper concludes with a summary of MDI modelling procedures.

The most complete published treatment of practical MDI models is that of Gokhale and Kullback (1978). Although it provides some guide to methodology and computer programs, the example analyses are agricultural and medical rather than managerial, and the presentation is opaque. In addition, the computational algorithms given by Gokhale and Kullback have limited capability.

This monograph will provide examples in the business and planning contexts, noting especially the relationship of the statistical models to decision processes and the diversity of managerial models that have been shown to be equivalent to MDI formulations. We will note additional results from the author's work and from other sources[*] that have a bearing on practical methodology, as well as some very basic procedural problems and solutions that have arisen in practice. To a great extent, this monograph is a progress report and a description of statistical, computational and interpretive issues that remain open. For example, Section V describes the most advanced computational procedures currently available for MDI problems.

Akaike (1973), (1977), (1978) has extended the original work of Kullback (1959) by applying information theoretic statistics to problems of decision theory, multivariate analysis, and Bayesian analysis. Although these contributions illustrate the generality of information theoretic

---

[*]In the text, the pronouns "we" and "our" are sometimes used to informally refer to previous and current work on MDI statistics. The research indicated in these instances is that of A. Charnes, W. W. Cooper, D. B. Learner, F. Y. Phillips, and colleagues of the above who have coauthored the cited publications. Where a specific previous work is cited, however, formal acknowledgement is given.

methods, all of the applications discussed in this paper belong to a
special class of discrete MDI models called <u>linearly constrained MDI</u>
<u>estimators</u>. These models lead to an Analysis of Information table (a summary
of the additive effects of the model's constraints); to estimates distributed
as exponential families; to loglinear representations of the estimates;
to a mathematical programming theory allowing for exact optimality condi-
tions and duality states; to efficient solutions via an <u>unconstrained</u> dual
problem (Charnes, Cooper, Seiford (1978)) and to other features attractive
for management decision modeling.

The models to be discussed can be represented generically in the form
of the mathematical programming problem

$$\text{Maximize} \quad -\Sigma_i \ \delta_i \ \ln(\frac{\delta_i}{ec_i})$$

subject to $A^T \delta = b$

$$\delta \geq 0 \tag{I.1}$$

Note that a concave functional is maximized subject to linear equality
constraints (written here in matrix form).

The form (I.1) will be specialized in various ways for the example
applications to follow. In particular the specialized form

$$\text{Maximize} \quad -I \ (p{:}q) \ \underline{=} \ -\Sigma_i \ p_i \ \ln(\frac{p_i}{q_i})$$

subject to $A^T p \ = \ \theta$

$$\Sigma_i \ p_i \ = \ 1$$

$$p_i \ \geq \ 0 \tag{I.2}$$

is of interest for statistical purposes. If $q_i > 0$ and $\Sigma_i \ q_i = 1$, then

4

the function I (p:q) is a measure of the divergence of the probability mass function p from the mass function q -- the "distance" between the two distributions in the information theoretic sense (see Kullback (1959) for a more detailed development of I (p:q)). When the constants q_i or θ_j are observations from a sample of size N, and when an optimal solution p* exists for (I.2), 2NI (p*:q) is asymptotically distributed as chi-square, under the null hypothesis that p = q. Thus solving (I.2) yields (1) an estimate p* which satisfies Ap* = θ and "is as close as possible" to the distribution q in the information theoretic sense; (2) a test of the hypothesis that the p and q distributions are identical; and (3) a loglinear model of the quantities p_i^* :

$$\ln p_i^* = \ln q_i + \sum_j a_{ij} z_j^* \qquad (I.3)$$

where the z_j^* are dual evaluators corresponding to the constraints of (I.2). In the terminology of Gokhale and Kullback (1978) the z_j^* are the scale parameters and the natural parameters of the loglinear model.

Gokhale and Kullback (1978) remark that for applications, the q_i (and, we add, the θ_j) may be observed, theoretical, or estimated quantities. Learner and Phillips (1980) also add that in the managerial context, the q_i may be historical figures, anticipated figures, or desired figures. The use of (I.1) (or (I.2)) as a decision model depends on some of the co-efficients c_i, a_{ij} or b_j being managerially controllable, at least implicitly. Then the δ_i or p_i may be intelligently influenced through the information given by their loglinear representation -- possibly with the intent of affecting the status of the hypothesis H_0 : p = q or other hypotheses.

5

II. STATISTICAL ASPECTS

The concept of information is implicit in all statistical inference, which is, after all, a way of inductive logic. Information was first considered explicitly and quantitatively by R. A. Fisher in the 1920's. A more recent concept of statistical information due to Khintchine (1949), (1953) and to Kullback and Leibler (1951) is, under certain conditions, asymptotically equivalent to Fisher information. Kullback and Leibler's (1951) suggestion that a complete system of statistical inference could be built around this definition of information was fulfilled by Kullback (1959). Additional historical and developmental background is available in Kullback (1959), Guiasu (1977), Phillips (1978), and Haynes, Phillips and Mohrfeld (1980).

Modern statistical information theory provides a unified treatment of estimation, hypothesis testing, and statistical decisions; and provides an objective determination of some previously heuristic procedures in multivariate and Bayesian analysis (see Charnes and Cooper (1975), and Akaike (1973, (1978)). In particular, the linearly constrained MDI estimators introduced in section I generalize univariate and multivariate logit analysis and quantal response analysis (Gokhale and Kullback (1978)) and some classes of probit and individual choice models (Charnes, Cooper, Learner and Phillips (1980)), treating these models as special cases. This is possible because the loglinear (multiplicative) representations and exponential distributions of MDI estimates are a consequence of the optimal solution of of problem (I.2), not a necessary prior assumption (Gokhale and Kullback (1978) and Brockett, Charnes and Cooper (1978).

6

The MDI estimates, under the null hypothesis, are equivalent asymptotically or in probability to maximum likelihood estimates (ML), least squares estimates, and many chi-square type estimates. In some important cases, MDI and ML estimates are identical. MDI estimates are also, in general, best asymptotically normal (Gokhale and Kullback (1978)).

Kullback's (1959) development of the information measure, $I(p:q)$, is summarized by Phillips (1978). Some essential properties of $I(p:q)$ are informally noted below and may be motivated by observing that, since p_i/q_i is a likelihood ratio, $I(p:q)$ is the expected value of the log-likelihood ratio, under the hypothesis that the random variable x is distributed according to the probability law $p(x)$.

If p^* is the optimal solution of (I.2), then $I(p^*:q)$ is the minimum discrimination information in favor of the hypothesis H_1 against H_0 where ("$\sim$" denotes "is distributed according to"):

H_0: $x \sim q(x)$ and

H_1: $x \sim p(x) \neq q(x)$ and

$I(p^*:q)$ may be interpreted as the mean information per observation for discrimination between the distributions p and q.

$I(p:q)$ has properties which are consistent with intuitive notions of "information." The measure is additive in that twice as much quantitative information may be said to result from twice as many statistically independent observations. The information measure has a related Pythagorean property which leads to the Analysis of Information table. $I(p:q)$ is nonnegative, equal to zero if and only if p = q. The information statistic formalizes the notion of a sufficient statistic as one that entails no loss of information about a sample. The discrimination information for

7

statistics of observations is always less than or equal to that for the original observations, with equality only if the statistics are sufficient.

An algebra of conditional information can be applied when events or observations are not independent. However the models discussed here are based on assumptions that (1) an individual's probability of classification into a category i remains constant, and (2) individuals and events are sampled independently.

See Kullback (1959), Phillips (1978) or Gokhale and Kullback (1978) for more detailed development of these properties.

Gokhale and Kullback's book (1978) is concerned with contingency tables and the analysis of categorical data in cross-tabulated form. It develops a complete and general information theoretic treatment of this subject, including tests for conditional and unconditional homogeneity, independence and interaction of factors, and computational algorithms for these tests. It distinguishes two variants of the constrained MDI estimator -- the "internal constraints problem" and the "external constraints problem."

The internal constraints problem (ICP) is concerned with the estimation of a "smoothed" distribution, p_i, such that certain moments of p_i are equal to the corresponding moments of an observed distribution. When the moments can be expressed as linear statistics of the p_i, their observed values are assigned to the θ_j of problem (I.2). The q_i of (I.2) represent the null hypothesis to be tested, for example the theoretical values of p_i under an hypothesis of independence of certain linear combinations of the

p_i. The solution of (I.2) then determines the consistency of this hypothesis with the observed statistics at a desired significance level -- the less the value of the minimum discrimination information, the greater the allowable confidence in the null hypothesis, depending of course, on the sample size.

The MDI statistic $2NI(p^*:q)$ (where N is the sample size) for the ICP is asymptotically distributed as chi-square with degrees of freedom equal to the number of parameters estimated less the number of linearly independent constraints, under the null hypothesis.

Composite hypotheses may be tested via the Pythagorean property of MDI numbers (Kullback (1959)). If a set of internal constraints, $Ap = \theta$, is implied by or contained in a distinct constraint set, $A'p = \theta$ and p^* and $p^{*'}$ are the respective MDI estimates, then

$$2NI(p^*:q) = 2NI(p^*:p'^*) + 2NI(p'^*:q). \qquad (II.1)$$

The term, $2NI(p^*:p'^*)$, is called the "effect term" and measures the impact of the constraints in $A'p = \theta$ which are not present in $Ap = \theta$. Each of the terms of equation (II.1) is itself an MDI statistic, and the respective degrees of freedom are additive. Equation (II.1) constitutes the "Analysis of Information" for the ICP.

The "interzonal transfer" or "spatial interaction" problem is an example of an ICP. In this problem, the daily numbers of interzonal trips originating and terminating in each district are known, as is some friction-of-distance function such as the travel time between pairs of districts. The objective is to estimate the number of trips between each pair of districts, for purposes of road and public transportation planning, commercial zoning, or store location. The origin and destination data form the right-hand-

9

sides of the constraints of the problem:

$$\text{Maximize} \quad -\sum_i \sum_{\substack{j \\ i \neq j}} p_{ij} \ln(p_{ij}/q_{ij})$$

$$\text{subject to} \quad \sum_{\substack{j \\ i \neq j}} p_{ij} = O_i \quad i = 1, \ldots, n$$

$$\sum_{\substack{i \\ i \neq j}} p_{ij} = D_j \quad j = 1, \ldots, n$$

$$p_{ij} \geq 0 \tag{II.2}$$

The q_{ij} are perhaps the inverses of the travel times between zone pairs.
The optimal p^*_{ij} are the estimates of the interzonal transfers conforming
most closely to the inverse travel time distribution. If the observed
marginals (or "moments"), O_i and D_j, are consistent with a number of transfers
that are strictly inversely proportional to travel time, with discrepancies
due only to sampling error, then $2NI(p^*_{ij}:q_{ij})$ will be distributed chi-
square with d.f. $= n^2-n-(2n-1) = n^2-3n+1$. (The travel matrix without
its diagonal "intrazonal" elements has n^2-n elements. The n x n constraint
system (II.2) has $2n-1$ linearly independent rows. Hence $n^2-n-(2n-1)$ degrees
of freedom.)

An external constraints problem (ECP) example from economic plan-
ning is the update of the Leontieff input-output matrix. The compilation
of an inter-industry matrix from hard data sources is an enormous job even
at a high level of aggregation of industrial categories. Sometimes current
data on the marginals (total inputs and outputs of each industrial category)
are available and it is felt that the proportional relationships of a known
I/O matrix from an earlier point in time will still hold (i.e., there have
been no technological innovations significantly changing the input mix of

any industry). The new matrix can then be estimated as those p_{ij} satisfying the current marginals, IN_i and OUT_j, but resembling the old martix q_{ij} most closely:

$$\text{Maximize} \quad - \sum_i \sum_j p_{ij} \ln(p_{ij}/q_{ij})$$

$$\text{subject to} \quad \sum_j p_{ij} = IN_i \qquad i = 1, \ldots, n$$

$$\sum_i p_{ij} = OUT_j \qquad j = 1, \ldots, n$$

$$p_{ij} \geq 0 \qquad \qquad \qquad \qquad (II.3)$$

This problem differs from the interzonal transfer problem in that the diagonals, p_{ii}, are now of interest. Further, the positions of observed and hypothesized quantities have been reversed: q_{ij} now represents the observed I/O table rather than the hypothesized friction-of-distance function, and the right-hand-sides of the constraints now represent hypothetically consistent marginal sums rather than observed origins and destinations. In this ECP, the degrees of freedom under H_o: $p = q$ are $(2n - 1)$--the number of linearly independent constraints. This is also equal to the number of parameters necessary for the loglinear representation of each p_i (I.3). Note that if the impact of some technological change is conjectured, it may be incorporated in an alternative hypotheses, q'_{ij}, and tested as an ICP; or represented as additional "external" linear constraints and tested as an ECP.

The following ECP example is adapted from Gokhale and Kullback (1978). Suppose we have taken two samples, π^1 and π^2, of respective sizes, n_1 and n_2, on the spaces, Ω_1 and Ω_2. We wish to test the equality of the population means:

$$\text{Maximize} \quad -2n_1 \sum_{i \in \Omega_1} p_i^1 \ln\left(\frac{p_i^1}{\pi_i^2}\right) -2n_2 \sum_{j \in \Omega_2} p_j^2 \ln\left(\frac{p_j^2}{\pi_j^2}\right)$$

11

subject to

$$\sum_{i \in \Omega_1} p_i^1 = 1$$

$$\sum_{j \in \Omega_2} p_j^2 = 1$$

$$\sum_{i \in \Omega_1} i p_i^1 - \sum_{j \in \Omega_2} j p_j^2 = 0$$

$$p_i^1, \ p_j^2 \geq 0 \tag{II.4}$$

The last equation represents the null hypothesis, H_o; $\mu_1 = \mu_2$.

For the general ECP,

$$\text{Maximize} \quad -\sum_i p_i \ln(p_i/\pi_i)$$

$$A^T p = \theta$$

$$p_i \geq 0 \tag{II.5}$$

$2NI(p^*:q)$ is asymptotically distributed χ_r^2 when the constraint matrix, A^T, is of rank r, excluding the normalization constraints $\sum p_i = 1$.

In the case of multi-sample problems (k-sample problems) such as example (II.4),

$$2 \sum_k n_k \ I(p^{k^*}:q) \sim \chi_s^2 \ , \tag{II.6}$$

where s is the number of linearly independent constraints not counting the k normalization constraints.

If $A'p = \theta' \Rightarrow Ap = \theta$, then the resulting measures of fit and of effect, and the associated degrees of freedom, are additive, as in the ICP model. To continue example (II.4), suppose we affix a constraint to problem (II.4) requiring that

$$\sum_{i \in \Omega_1} i^2 p_i^1 - \sum_{j \in \Omega_2} j^2 p_j^1 = 0 \tag{II.7}$$

12

The constraints now imply an hypothesis that the population means and variances are equal. The MDI number for the second problem, I^2, has one more degree of freedom than I^1 (the MDI number for the first problem), and $I^2 = I^{12} + I^1$. If it happens that the first hypothesis is not rejected and the second hypothesis is rejected, then I^{12}, the measure of the effect of the added constraint, is said to be significant.

Generally, the MDI estimate under ECP is not equivalent to the maximum likelihood estimate (Gokhale and Kullback (1978)). In the notation of (II.5), the loglinear representation of the k-sample MCI/ECP estimate is

$$\ln\left(\frac{p_i^*}{\pi_i}\right) = \sum_{j=1}^{k} a_{ij} L_j + \sum_{l=1}^{r} a_{i,1+k} \tau_l \qquad (II.8)$$

The p_i^* and the the rows of $[a_{ij}]$ are reindexed lexicographically in (11.8) so that $i = 1, 2, \ldots, \sum_k J_k$; where J_k is the number of categories in the kth sample. Gokhale and Kullback (1978) call the L_i "normalization constants," and the τ_l the "natural parameters" of interest for the model.

A brand switching model used as a research tool at the Market Research Corporation of America will provide an example of an ECP. Given a brand switching matrix $[p_{ij}]$, $(\sum_i \sum_j p_{ij} = 1)$, a brand's "vulnerability ratio" may be defined as a linear function, $R^i(p_{ii})$, of the brand's repeat buying probability, p_{ii}. The segmentation of the brands of a market into groups of brands having vulnerability ratios which are not significantly different can result in marketing insights. To test whether two brands, k and k', can be considered members of the same segment, we begin with the observed switching matrix $[q_{ij}]$ and solve

13

$$\text{Maximize} \quad -\sum_i \sum_j p_{ij} \ln(p_{ij}/q_{ij})$$

$$\text{subject to} \quad \sum_i \sum_j p_{ij} = 1$$

$$R^k(p_{kk}) - R^{k'}(p_{k'k'}) = 0$$

$$p_{ij} \geq 0 \tag{II.9}$$

$2NI(p_{ij}^*:q_{ij})$ is approximately distributed χ_1^2 under the hypothesis. If we accept this hypothesis and go on to conjecture that brand k'' is also a member of the segment, we must solve

$$\text{Maximize} \quad -\sum_i \sum_j p_{ij} \ln(p_{ij}/q_{ij})$$

$$\text{subject to} \quad \sum_i \sum_j p_{ij} = 1$$

$$R^k(p_{kk}) - R^{k'}(p_{k'k'}) = 0$$

$$R^{k'}(p_{k'k'}) - R^{k''}(p_{k''k''}) = 0$$

$$p_{ij} \geq 0 \tag{II.10}$$

yielding an optimal solution $p_{ij}^{*'}$. The results may be summarized in an Analysis of Information table:

Analysis of Information

		2NI	d.f.
1.	fit of k,k'	2NI(p*:q)	1
2.	effect of k"	2NI(p*':p*)	1
3.	fit of k,k',k"	2NI(p*':q)	2

Note that care must be exercised in writing down the sample size. If the switching matrix $[q_{ij}]$ is compiled from observing many consecutive pairs of purchases, we may want to say that N = the number of observed pairs of purchases, not N = the number of observed purchases. Note also that although for a short-term study both the row and column marginal sums of

14

$[q_{ij}]$ are equal to the market shares of the respective brands, this fact is not explicitly utilized in (II.9) or (II.10). Although marginal sum constraints are ubiquitous in the ICP, the entire observed distribution is written down as the q constants in the ECP. If the ECP estimate, p, is accepted as identical to q, then a fortiori the marginals of p will be equal to those of q.

The linear equation representation of a given hypothesis may not be unique. When an external hypothesis can be represented as two distinct but equivalent sets of constraints, the test statistics are not sensitive to the formulation (constraint set) used. However, the number and values of the coefficients in the loglinear representation of the estimates are affected by the formulation, since these are measures of the effects of individual constaints. See Gokhale and Kullback (1978a) for examples.

The ECP examples just given indicate that testing composite hypotheses requires the solution of several MDI problems. Usually these are solved sequentially rather than simultaneously, with the decision for the next stage based on the results of the previous stage. Gokhale and Kullback (1978) stress that common sense is the best determinant for selecting the sequence of problems to be solved.

The different rules for computing the degrees of freedom in the ICP and the ECP seem to be based on the rationale that (1) in the ECP we are usually more interested in the hypothesis test than in the estimates of the p_i, (2) the test depends on the significance of the linear equations comprising the representation of the hypothesis, and (3) therefore the r unconstrained "natural parameters", τ_1, are of greatest interest. Also, the roles of the observed quantities and the hypothesized quantities

15

are essentially reversed in the mathematical formulations of the ICP and the ECP.

Recent experience in applications, however, has led to doubts that these rules for the ICP and the ECP cover the complete range of realistic cases. First, we can conceive of cases in which the π_i of the ECP are not observed, but estimated. Second, as Gokhale and Kullback stress, $I(p^*:q)$ is useful not only as a test statistic but as a general measure of the "distance" from an hypothesis, and it has a known approximate distribution even when the null hypothesis is rejected. Therefore, especially in managerial situations, the estimates of all primal and dual parameters may be of interest for suggesting further hypotheses. This is especially true in light of the mathematical programming theory discussed in the next section, which provides additional interpretations of the dual parameters (i.e., the scale and natural parameters).

Unfortunately, differences of opinion concerning the degrees of freedom of an estimate are not rare, even among professional statisticians (see, for example, Jaynes (1978)). These arguments are usually due to the existence of additional restraints on the parameters which are latent, implicit or hidden. As further research clarifies the nature of these extra restraints, active discussion may resolve the matter of whether it is legitimate to "count" them in the determination of degrees of freedom. Since we are still in the early stages of practical experience with constrained MDI estimators, these uncertainties may soon be resolved.

Meanwhile, if the number of d.f. is uncertain for a given application, a quick and dirty approximation can be achieved by restating the problem in terms of a very large number of free parameters, thus rendering the

16

critical value of the test less sensitive to the degrees of freedom. In this case, we would use a normal approximation of the chi-square distribution (Kendall and Stuart (1958)) under which the null hypothesis implies that $2(NI*)^{\frac{1}{2}}$ has a normal distribution with unit variance and mean equal to $(2k-1)^{\frac{1}{2}}$, where k is the degrees of freedom.

Gokhale and Kullback (1978) specify the covariance matrix of the τ_1 on the basis of the constraint matrix, A^T, and a diagonal matrix, D, defined as $d_{ii} = Np_i^*$: compute $S = ADA^T$, and partition S as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

where, for the case of one "natural" or normalizing constraint, S_{11} is 1x1 and S_{22} is rxr. The estimated asymptotic covariance matrix is

$$\Sigma_\tau = [S_{22} - S_{21} S_{11}^{-1} S_{12}]^{1} \qquad (II.11)$$

Gokhale and Kullback (1978) continue with a procedure for asymptotic simultaneous confidence intervals for the taus and for linear combinations of the taus (note that the log-odds expressions, $\ln(p_i/q_i)$, are linear combinations of the taus).

For the practical MDI applications noted in this section and in section IV (and especially, as Gokhale and Kullback (1978) remark, for problems involving contingency tables), the sample sizes are naturally quite large. We can therefore feel comfortable with tests based on an asymptotic chi-square distribution for 2NI*.

III. Mathematical Programming Aspects

The development of a mathematical programming theory of MDI estimation
was motivated by both applied (Charnes, Raike and Bettinger (1972)) and
theoretical (see, Charnes, Cooper and Seiford (1978) and Phillips (1978))
considerations. Its development has added a precise duality theory and
efficient computational methods to the inferential capability of statistical
information theory.

Of course, mathematical programming techniques are applicable to a
great variety of statistical problems (see, Armstrong, Frome and Kung
(1979) for just one example). But they are especially attractive in the
case of MDI estimation because of the unconstrained dual to problem (I.1)
and because of the special applicability of MDI statistics to business and
management problems.

Consider the dual programs

Primal	Dual

$$\sup_{\delta} v(\delta) = -\sum_{i} \delta_i \ln\left(\frac{\delta_i}{ec_i}\right) \qquad \inf_{z} \xi(z) = c_i e^{{}_i Az} - b^T z$$

subject to $A^T \delta = b$ $\qquad\qquad\qquad$ z unconstrained

$$\delta \geq 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad (III.1)$$

where ${}_i A$ indicates the ith row of A, and A^T is the transpose of A.

Brockett, Charnes and Cooper (1978) set down a theorem developed by
Charnes and Cooper that explains the most specific and complete duality
theory for these problems. In summary, if the primal problem has a nonzero
feasible solution, then (1) $v(\delta)$ has a unique maximum at $\delta * > 0$, and

18

$\xi(z)$ has a unique minimum at z^*; (2) $v(\delta^*) = \xi(z^*)$; and (3) $\delta_i^* = c_i e^{iAz^*}$

Note that this last equation is the basis of the loglinear model (I.3) and that, as a mathematical programming optimality condition <u>it is true regardless of whether $H_0 : p = q$ is accepted or rejected</u>. The dual evaluators, z^*, take the role of regression coefficeints in this equation. The Analysis of Information table tests the significance of these coefficients -- more precisely, the significance for the hypothesis of the constraints corresponding to the coefficients.

Note also that the primal problem (III.1) is not identical to the maximization of $-I(p:q)$ of problem (I.2) in that (1) the quantity, e, appears in the denominator of the argument of the logarithm. This means that v^* will differ from I^* by a constant equal to $\sum_i \delta_i^*$; and (2) the δ_i and c_i are not necessarily required to sum to unity. From a mathematical programming standpoint, these differences do not affect the solution since, if $\Sigma_i \delta_i = \Delta$, then the optimal δ_i^* and p_i^* will be in the same <u>relative proportions</u> for any given Δ whether $v(\delta)$ or $I(p:q)$ is maximized. Thus it is always possible to normalize or otherwise reconcile the optimal solutions of (III.1) and (I.2) <u>after</u> the solution , via $p_i^* = \delta_i^*/\Sigma_i \delta_i^*$.

For example, if we solve both

$$\text{Max. } -I(p:q) = \sum_i p_i \ln\left(\frac{p_i}{q_i}\right) \qquad \text{Min. } -v(\delta) = \sum_i \delta_i \ln\left(\frac{\delta_i}{eq_i}\right)$$

$$\text{and}$$

$$\text{subject to } A^T p = b \qquad\qquad \text{subject to } A^T \delta = b\Delta$$

$$\text{and } \sum_i p_i = 1 \qquad\qquad \sum_i \delta_i = \Delta$$

$$p \geq 0 \qquad\qquad\qquad \delta \geq 0 \qquad \text{(III.2)}$$

19

then $\delta_i^* = p_i^* \Delta$ and $-\sum_i \delta_i^* \ln(\delta_i^*/eq_i) =$

$$-\sum_i p_i^* \Delta \ln(\frac{p_i^*}{q_i}) - \ln(\frac{\Delta}{e}) \sum_i p_i^* =$$

$$-\Delta \sum_i p_i^* \ln(\frac{p_i^*}{q_i}) + \Delta \ln \Delta = -\Delta I(p^*:q) + \Delta \ln \Delta. \qquad (III.3)$$

From a statistical standpoint, however, it is essential to have the optimal value of $I(p^*:q)$ in order to proceed with hypothesis testing (since the test is based on the sampling distribution of $2NI(p^*:q)$).

In this regard we may note that for contingency tables, the functional can be expressed in terms of <u>frequencies</u>, x_i, rather than in terms of the proportions, p_i, where we have $x_i = Np_i$. In that case, e.g., for the ECP,

$$-\sum_i x_i \ln(\frac{x_i}{x_i^{obs}}) = -N \sum_i p \ln(\frac{Np_i}{Np_i^{obs}}) =$$

$$-N\sum_i p_i \ln(\frac{p_i}{p_i^{obs}}) = NI(p:p^{obs}) \qquad (III.4)$$

Thus $-2\sum_i x_i \ln(x_i/x_i^{obs})$ may be used directly as the test statistic for the

ECP and similarly for the ICP.

Depending on the computer algorithm used, the transformation (III.3) may be advantageous for scaling the variables for purposes of numerical stability. Care must be taken to retransform the solution before performing any statistical interpretation.

This transformation is also the key to the fact that MDI models are a

generalization of the so-called "maximum entropy" models (as shown by Phillips, White and Haynes (1976); see also Wilson (1970)). These models involve maximizing the "entropy" function of a distribution p:

$$\text{Maximize} \quad H(p) = \sum_i p_i \ln p_i \qquad (III.5)$$

subject to linear constraints as in (III.2). It is easily seen that $H(p)$ finds its maximum at the same point as does $I(p:\tilde{q})$, if $\tilde{q}_i = 1/n$, for $i = 1, 2, \ldots, n$;

$$I(p:\tilde{q}) = \sum_i p_i \ln(p_i/n^{-1})$$

$$= -\sum_i p_i \ln p_i - \ln n$$

$$= -H(p) - \ln n = -H(p) - (\text{constant}). \qquad (III.6)$$

$H(p)$ and $I(p:\tilde{q})$ are therefore measures of the deviation of p from a discrete uniform distribution. All the remarks in this paper also apply to the use and computation of max-entropy models, since these are a special case of MDI models.

A recurring problem in the measurement and analysis of business processes involves the grouping of values from many categories into fewer categories for purposes of tractability. In the most extreme and general form of the problem, one must approximate by discrete intervals a variable which would otherwise be regarded as continuous. Examples include distributions of income, brand loyalty, corporate size, etc.

The implied problem for MDI modelling is as follows: suppose $p_i (i = 1, \ldots, n)$ represent approximate discrete category proportions, and Ⓗ is a vector of observed moments of the latent continuous distribution (not necessarily simple marginal sums). Then, in practice, it is quite likely

21

that the system, $A^Tp = $ (H), will have no solution. A second problem is that, even when $A^Tp = $ (H) does have a solution, this solution is sensitive to the choice of n and the interval boundaries.

We first attacked these problems by replacing $A^Tp = $ (H) by various sets of inequalities involving the $\hat{r}$, p_i, and θ_i, drawn from probability theory. These inequalities allowed for a feasible solution to the primal problem, but the approach had serious disadvantages. Each interval must be assigned a "value", x_i, so that moments of the ratio-scaled variable may be expressed mathematically (e.g., $E(x) = \Sigma_i x_i p(x_i)$). Usually the mid-point of the interval is chosen for x_i, but this can mean a very unsatis-factory approximation. Also, the statistical theory of the MDI was developed for the case of equality constraints; qualifications may be necessary for a statistical interpretation of the inequality-constrained MDI. Finally, sensitivity analysis has to be performed by trial and error when the inequality constraint approach is used.
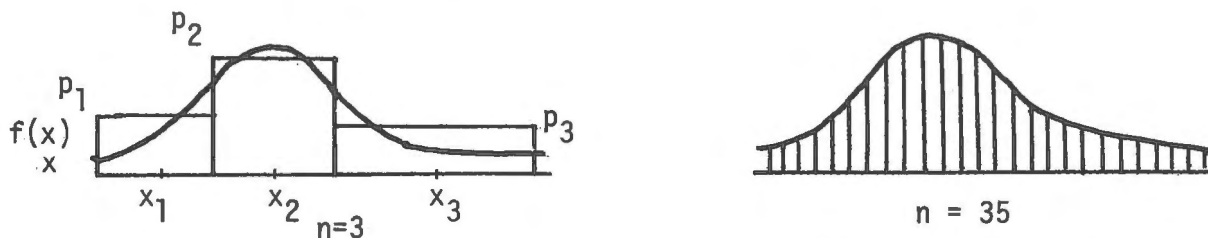
An alternative strategy, suggested by John Rousseau (1979), is to restate the model in terms of a vary large number of free parameters (i.e., primal variables or discrete categories). This will increase the similarity of the discrete distribution to the latent continuous distribution, and "increase the chances" of a feasible solution to $A^Tp = $ (H). This strategy imposes little extra computational burden on the dual side, since the number of dual variables depends only on the number of primal constraints -- and

22

there are no dual constraints. (The optimal variables are translated into
an optimal primal vector via the equation $\delta_i^* = c_i e^{iAz^*}$, which is a condition
of optimality.) This approach may also have some statistical advantages,
as suggested in section II. It will, however, be inapplicable if the number
of categories in the model must conform to the number of categories of
some outside data which the model must address.

An important special case of (III.1) is the maximization of $I(p:q)$
subject to "origin-destination" or "supply-demand" equations. The con-
straints for this problem are the same as those for the network problem
without transshipment, i.e., the distribution problem, of linear program-
ming. Many MDI applications to date have been of this form, including the
I/O matrix example given earlier. The dual programs for this special MDI
model are written below:

<div style="display:flex; justify-content:space-between;">

Primal

Dual
</div>

$$\text{Max. } -\sum_i \sum_j p_{ij} \ln\left(\frac{p_{ij}}{eq_{ij}} \right) \qquad\qquad \text{Min. } \sum_i \sum_j r_i \, q_{ij} \, s_j -$$

$$\text{subject to } \sum_J p_{ij} = O_i' \qquad\qquad\qquad \sum_i r_i \, O_i - \sum_J s_j \, D_j.$$

$$\sum_i p_{ij} = D_j \qquad\qquad\qquad r_i \, s_j \text{ unconstrained}$$

$$p_{ij} \geq 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad (III.7)$$

Note the double-subscripted primal variables. All aspects of problems
(III.7) are strictly parallel to the general problems (III.1).

IV. Business and Managerial Aspects

The main function of this section is to briefly review past and recent business applications of MDI techniques, with references to sources, and to indicate the areas of current research and future potential.

The one methodological point made in this section is simply that the models discussed are data oriented. One must be careful to devise appropriate units of measurement (see Learner and Phillips (1980)). Sources of business data are often secondary sources rather than direct experiments, and figures for different time periods may be drawn from two or more sources whose units of measurement are incompatible. Furthermore, sampling units and sample sizes can be difficult or impossible to infer from library sources of business data. These considerations mean that the statistical interpretations of the formal MDI model must be tempered with judgment in applications.

Business and management applications of MDI to date include the traffic planning (see Phillips (1978)), interindustry matrix,* and product segmentation (Charnes, Cooper, Learner and Phillips (1978)) problems used as examples earlier in this paper. Some others include:

1. An individual choice model (the multiplicative competitive interaction "MCI" model of Nakanishi and Cooper (1974), basing probability of choice on attributes of the choice objects, has been shown to be equivalent to an MDI formulation

---

*Charnes, Phillips, Rousseau and Narasimhan (forthcoming). The problem is discussed, but not related to MDI, in Bachem and Korte (1978) and Fisch and Gordon (undated).

(Charnes, Cooper, Learner and Phillips (1980)). The model has application in studies of brand choice and shopping location choice.

2.   Charnes and Cooper (1974) devised a model of the behavior of a cartel economy which is equivalent to an MDI.

3.   Charnes, Cooper and Learner (1978) demonstrated the equivalence of SANDDABS, an established model of the shifting of consumers' brand preferences, to an MDI estimation.

4.   Theil (1967) uses entropic measures to model the distribution of resources and income across regions and populations. Semple (1970), Semple and Demko (1977) and Batty (1972) also develop information theoretic measures of concentration of markets, incomes, and corporate headquarters sites.

5.   Phillips (1978) relates the use of the ICP approach to monitor the components of a brand's sales in a model called MARK-IT.

6.   Charnes and Cooper (1975) have defined and characterized entropic solutions to n-person games having favorable theoretical, computational and statistical properties.

7.   Jaynes (1978) describes an application of a max-entropy principle to the estimation of manufacturing tolerances.

8.   Learner and Phillips (1980) show how the properties of the MDI measures lead to methods of controllable forecasting that are particularly appropriate in the business/managerial environment.

9.   Charnes, Cooper and Schninnar (1976) derive production functions of the multiplicative or Cobb-Doublas type via MDI estimation.

Current research includes refining the procedures for applying the equivalencies mentioned above and investigating the connections between MDI techniques and chance-constrained programming.

The properties of the Analysis of Information table have another particular potential for business planning -- the simplification of the estimation procedures for some simultaneous equation and econometric models. The Pythagorean property of $I(p:q)$ could be used to measure the incremental effect of each restriction in the model, making it unnecessary to devise separate maximum likelihood estimates of each model parameter.

V. Computational Aspects

This section reviews some solution algorithms published prior to our own work and then describes our recent approaches to computation of the MDI estimates, namely (1) solution as a linear programming problem via a piecewise linear approximation to the information functional (Charnes, Haynes and Phillips (1976)); (2) solution as a nonlinear programming problem on both the primal and dual sides (Charnes, Narasimhan, Phillips and Rousseau (1980)); and, for some special problems (3) analytical solutions obtainable without the use of a computer (Charnes, Cooper and Learner (1978)).

Prior to the establishment of the mathematical programming equivalences, several iterative algorithms existed for solution of MDI-type problems. These included the algorithms of D'Esopo and Lefkowitz (1963) and Wilson (1970) for the spatial interaction model, various implementations of the "RAS" algorithm for the update of the inter-industry matrix (Bacharach (1970), Bachem and Korte (1978) and Fisch and Gordon (undated)), and the MDI algorithms of Gokhale, et.al. (see Gokhale and Kullback (1978)). None of these algorithms were proven to converge to the true minimum value of the functional, and their behavior under ill-conditioned problems was undocumented. Still, these procedures are easy to program and may be conveniently used for small problems or when powerful computational facilities are not available.

Gokhale and Kullback (1978) provide a package of programs (available through the statistics department at George Washington University) for MDI analysis of contingency tables. Most of these programs have specialized features for the analysis of particular interaction hypotheses, etc. Slightly different procedures are used for solving special cases of the ICP and ECP, each with a different stopping rule (condition of fit).

27

The mathematical programming approaches treat all of these problems equally, although on a less-particularized level. A single optimality criterion is applied to all of the problem types.

Our first implemented application of mathematical programming to MDI was a linear programming approximation to the interzonal transfer problem (Charnes, Haynes and Phillips (1976)). In this procedure, the information functional in problem (II.2) is rendered as $-\sum_i \sum_{\substack{J \\ i \neq j}} [t_{ij} \ln t_{ij} - t_{ij} \ln K_{ij}]$.

The portion of the separable functional associated with each t_{ij} has a linear term, $t_{ij}(\ln K_{ij})$, and a nonlinear term, $t_{ij} \ln t_{ij}$, graphed below.

minimum at (1/e, -1/e)
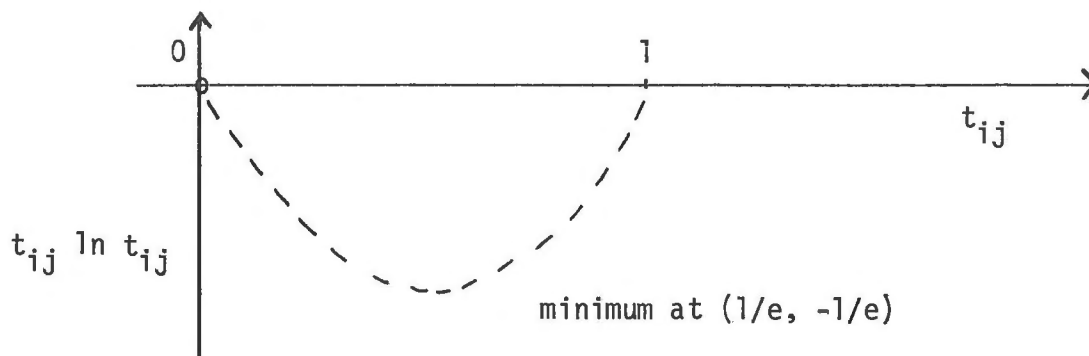
It is assumed that t_{ij} is constrained to be between zero and one. The piecewise linear approximation is achieved by splitting each t_{ij} into two variables, t_{ij}^1 and t_{ij}^2, as suggested by the graph below:

(1/e, -1/e)
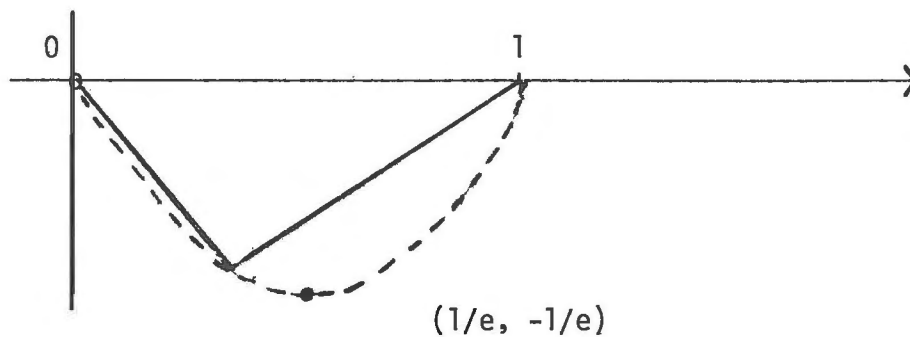
The slope of the line segment on the left is -1; the segment on the right has a slope of $1/(e-1)$. Stipulating that $t_{ij}^1 \leq e^{-1}$, we can therefore approximate $t_{ij} \ln t_{ij}$ by $-t_{ij}^1 + \frac{1}{e-1} t_{ij}^2$. The following steps will then solve the interzonal transfer problem:

1. Substitute $-t_{ij}^1 + \frac{1}{e-1} t_{ij}^2$ for $t_{ij} \ln t_{ij}$ in the functional of problem (II.2).

2. Substitute $t_{ij}^1 + t_{ij}^2$ for t_{ij} for every remaining appearance of t_{ij} in problem (II.2).

3. Add the condition that $t_{ij}^1 \leq e^{-1}$.

4. Solve the resulting capacitated linear distribution problem.

5. Make the backward transformation $t_{ij}^* \leftarrow t_{ij}^{1*} + t_{ij}^{2*}$.

See Charnes and Cooper (1961) for a fuller explanation of the use of piecewise linear functionals.

This approach is disadvantageous in that (1) the duality theory of the nonlinear formulation is inapplicable; (2) the number of nonzero basis elements in the optimal linear programming solution is almost always less than the total number of variables in the problem, whereas an MDI solution is strictly positive; and (3) the rough approximation of the information functional severely limits the potential for hypothesis testing.

Nonetheless, this procedure can be useful and even advantageous in cases where the estimates, t_{ij}^*, are of primary importance and hypothesis testing is of minimal interest, or where the size of the problem exceeds the capacity of available nonlinear programming codes. Both of these

conditions held at the time of the Corpus Christi intraurban transfer study (56 origins and 56 destinations). The linear programming approximation provided an immediate solution by means of an available capacitated network code (in terms of which this was a small problem).

Of course, the quality of the piecewise linear approximation can be improved by using a three-piece, four-piece, or n-piece breakdown instead of the two-piece approximation just illustrated.

Several of the other examples used in this paper have been solved by the Sequential Unconstrained Minimization Techniques (SUMT) of Fiacco and McCormick (see Mylander, Holmes and McCormick (1971)). This set of algorithms deals handily with the nonlinear MDI problems, especially the unconstrained dual form; and SUMT's regularization options seem well suited to mending MDI problems with ill-conditioned constraint matrices. SUMT, which analyzes the Hessian of a transformed functional, solves MDI problems faster (subject to some difficulties of direct comparison) than an alternative numerical descent code that analyzes only first derivatives, and faster than GOKHALE, the most general of the programs available from Gokhale and Kullback.

Section II implied that the test of a composite hypothesis with MDI involves solving several constrained MDI estimates in sequence. It follows that solution times for later problems in such a sequence can be reduced by using the optimal solution of an earlier problem as a starting point or as an indication of a favorable starting point. This may, in fact, be one consideration in determining the order of solving problems in the sequence.

In Section III, emphasis was given to the quest for a feasible solution. Brockett, Charnes and Cooper (1978) provide a linear programming problem

which can be applied when the existence of a feasible solution is questionable. The solution of this LP indicates the feasibility/non-feasibility of the nonlinear-constrained MDI problem.

Finally we mention that some classes of constrained MDI's may be solved manually by means of a simple formula. Charnes, Cooper and Learner (1978) showed that a component of the SANDDABS model corresponds to an MDI with an analytical solution. This result depended on a functional relationship between the q_{ij} and the O_i and D_j of problem (III.9). An exhaustive characterization of the class of MDI problems having analytic solutions remains a subject for future research.

VI. Summary of Modelling Methodology

To summarize the modelling procedures presented in this monograph, let us reiterate the hypothesis testing paradigm within the context of statistical information theory. The distribution, $f_1(x)$, (with parameters θ) closest in the information theoretic sense to another distribution, $f_2(x)$, is that $f_1^*(x)$ satisfying the minimization

$$\text{Minimize} \quad I(1:2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} \, dx$$

subject to $\int T(x) \, f_1(x) \, dx = \theta$ \hspace{2cm} (VI.1)

where $T(x)$ is a measurable (in this example, linear) statistic for θ. If there is more than one unbiased statistic for θ, that one yielding the largest value of $I^*(1:2)$ is used (and the estimate is really a "maximin"). The minimum value $I^*(1:2)$ of $I(1:2)$ is called the minimum discrimination information.

We first decide whether to represent our hypothesis by means of $f_2(x)$ ("ICP") or by means of θ ("ECP"). In either case, the solution of (VI.1) will test the consistency of $f_2(x)$ and θ, where one of these is hypothesized and the other is observed.

Having set up the appropriate constrained MDI problem in the form of (VI.1), we take a sample of N observations on x. If we are working with an ICP, we substitute the value of the sample statistic, $T(x)$, for the population parameter, θ, in (VI.1). If we are in ECP, we substitute the empirical distribution, $q(x)$, for $f_2(x)$ in (VI.1).

The null hypothesis, H_o, that x is a sample from the population

32

characterized by $f_2(x)$ is then tested. The test takes the form:

reject H_0 if $2NI*(1:2) > c$, for a suitable c. (VI.2)

For ICP, the degrees of freedom for $2NI*(1:2)$ is given on page 9 . For ECP, the d.f., is defined on page 11. $2NI*(1:2)$ approaches a chi-square distribution as N grows; for small numbers of degrees of freedom, c should be a chi-square fractile. For large numbers of degrees of freedom, we reject H_0 if $2(NI*(1:2))^{\frac{1}{2}} - (2k-1)^{\frac{1}{2}}$ is greater than c', where k is the degrees of freedom and c' is a standard normal fractile (see page 15). All tests are one-tailed.

It should be noted that if H_0 is rejected under this test, we are nonetheless left with an estimate of the distribution of x which yields minimum information for discrimination against $f_2(x)$ and may yield further insights in the context of the application.

The loglinear models, and their derivative forms, which result from the MDI solution present useful variate relationships based on optimal use of management information. These relationships are simplified and managerially useful representations of aggregated or disaggregated empirical data. The examples given in this monograph illustrate the usefulness of these relationships in a wide range of business and planning situations and the unifying role of MDI principles in the analysis of demand and market structure.

References

Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle." In 2nd International Symposium on Information Theory (B. N. Petrov and F. Csaki, eds.), pp. 267-281, Akademiai Kiado, Budapest (1973).

Akaike, H., "An Extension of the Method of Maximum Likelihood and the Stein's Problem." Ann. Inst. Statist. Math. 29 Part A, (1977), pp. 153-164.

Akaike, H., "A New Look at the Bayes Procedure." Biometrika 65 No. 1 (1978), pp. 53-59.

Armstrong, R. D., E. L. Frome and D. S. Kung, "A Revised Simplex Algorithm for the Absolute Deviation Curve Fitting Problem". Commun. Statist.-Simula. Computa., B8(2), 175-190 (1979).

Bacharach, M., Biproportional Matrices and Input-Output Change. Cambridge: Cambridge University Press (1970).

Bachem, A., and B. Korte, "A Note on the RAS Algorithm." Report #78100-OR Institut fur Okonometrie and Operations Research Rheinische Freidrich-Wilhelms-Universitat, Nassestrasse 2, D-5300 Bonn, West Germany (1973).

Batty, M., "Entropy and Spatial Geometry." Area 4 (1972), P. 230.

Brockett, P. L., A. Charnes and W. W. Cooper, "M.D.I. Estimation via Unconstrained Convex Programming." Center for Cybernetic Studies Report CCS 326 (1978).

Charnes, A. and W. W. Cooper, Management Models and Industrial Applications of Linear Programming. New York: John Wiley & Sons (1961).

Charnes, A., and W. W. Cooper, "An Extremal Principle for Accounting Balance of a Resource Value-Transfer Economy Existence, Uniqueness and Computation," Center for Cybernetic Studies Report CCS 185 and Rendiconti di Accademia Nazionale dei Lincei (April 1974).

Charnes, A., and W. W. Cooper, "Constrained Kullback-Leibler Estimation, Generalized Cobb-Doublas Balance, and Unconstrained Convex Programming." Rendiconti di Accademia Nazionale dei Lincei (1975).

Charnes, A., W. W. Cooper and D. B. Learner, "Constrained Information Theoretic Characterizations in Consumer Purchase Behavior." Journal Opl. Res. Soc. 29, No. 9 (1978).

Charnes, A., W. W. Cooper, D. B. Learner and F. Y. Phillips, in R. Leone, ed., <u>Market Measurement and Analysis</u>, The Institute of Management Sciences, (1980).

Charnes, A., W. W. Cooper, D. B. Learner and F. Y. Phillips, "M.D.I. Methods as a Generalization of Logit, Probit and Hendry Analyses." Presented at ORSA/TIMS Joint National Meeting, Washington, D.C. (1980).

Charnes, A., W. W. Cooper, and L. Seiford, "Extremal Principles and Optimization Dualities for Khinchin-Kullback-Leiber Estimation." Research Report CCS261, Center for Cybernetic Studies, The University of Texas at Austin (April 1976); and <u>Math. Operationsforsch. Statist., Ser. Optimization</u> <u>9</u>, No. 1 (1978), pp. 21-29.

Charnes, A., W. W. Cooper and A. P. Schinnar, "A Theorem on Homogeneous Functions and Extended Cobb-Douglas Forms." <u>Proc. National Academy of Science, U.S.A.</u> <u>73</u>, No. 10 (October 1976).

Charnes, A., K. E. Haynes and F. Y. Phillips, "A 'Generalized Distance' Estimation Procedure for Intra-Urban Interaction." <u>Geographical Analysis</u> (July 1976).

Charnes, A., V. L. Narasimhan, F. Y. Phillips and J. J. Rousseau, "The Computation of Discrete Constrained M.D.I. Distributions." Market Research Corporation of America Discussion Paper PDA-RP80 (forthcoming).

Charnes, A., W. Raike, Co. O. Bettinger, "An Extremal and Information-Theoretic Characterization of Some Interzonal Transfer Models." <u>Socio-Econ. Plan. Sci.</u> <u>6</u> (1972), pp. 531-537.

D'Esopo, D. A. and B. Lefkowtiz, "An Algorithm for Computing Interzonal Transfers Using the Gravity Model." <u>Op. Res.</u> <u>11</u>, (1963), pp. 901-907.

Fisch, O., and S. I. Gordon, "Re-examination of the RAS Algorithm in Minimum Survey Method of Regional Input-Output Tables." Department of City and Regional Planning, Ohio State University, Columbus, Ohio (undated).

Gokhale, D. V., "Analysis of Loglinear Models." <u>Journal of Royal Statist. Soc. Ser. B</u> <u>34</u>, No. 3 (1972), pp. 371-376.

Gokhale, D. V. and S. Kullback, <u>The Information in Contingency Tables</u>. Marcel Dekker New York (c. 1978).

Gokhale, D. V., and S. Kullback, "The Minimum Discrimination Information Approach in Analyzing Categorical Data." <u>Commun. Statist.-Theor. Meth.</u> <u>A7</u>, No. 10 (1978a), pp. 987-1005.

Guiasu, S. Information Theory with Applications. McGraw-Hill (1977).

Haynes, K. E., F. Y. Phillips and J. Mohrfeld, "The Entropies: Some Roots of Ambiguity." Socio-Econ. Plan. Sci. (1980).

Jaynes, E. T., "Where Do We Stand on Maximum Entropy?" In Levine, R. D. and M. Tribus, The Maximum Entropy Formalism. Cambridge: MIT Press (1978).

Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics. London: Chas. Griffin and Co. (1958).

Khintchine, A. I., Mathematical Foundations of Statistical Mechanics. New York: Dover (c. 1949).

Khintchine, A. I., Mathematical Foundations of Information Theory. New York: Dover (c. 1957). (New translation of Khintchine's papers "The Entropy Concept in Probability Theory" and "On the Fundamental Theorems of Information Theory," originally published in Russian in Uspekni Matematicheskikh VII, No. 3 (1953), and XI, No. 1, 1956, respectively.)

Kullback, S., Information Theory and Statistics. New York: John Wiley and Sons (1959)

Kullback, S., and R. A. Leibler, "On Information and Sufficiency." Ann. Math. Statist. 22 (1951), pp. 78-86.

Learner, D. B., and F. Y. Phillips, "Controllable Forecasting in Marketing." Discussion paper, Market Research Corporation of America (1980).

Mylander, W. C., R. L. Holmes and G. P. McCormick, A Guide to SUMT Version 4: The Computer Program Implementing the Sequential Unconstrained Minimization Technique for Nonlinear Programming. Paper RAC-P-63, Advanced Research Department, Research Analysis Corporation, McLean, Virginia 22101 (1971).

Nakanishi, M., and L. G. Cooper, "Parameter Estimation for a Multiplicative Competitive Interaction Model." Journal of Marketing XI (August 1974).

Phillips, R. Y., "Some Information Theoretic Methods for Management Analysis In Marketing and Resources." Ph.D. dissertation, University of Texas at Austin (1978).

Phillips, R. Y., G. M. White, and K. E. Haynes, "Extremal Approaches to Estimating Spatial Interaction." Geographical Analysis (April 1976).

Rousseay, J., Personal communication (1980).

Semple, R. K., "Recent Trends in the Spatial Concentration of Corporate Headquarters". _Economic Geography 46_, 157-180 (1970).

Semple, R. K. and G. Demko, "An Information Theoretic Analysis: An Application to Soviet-COMECON Trade Flows." _Geographic Analysis IX_ (January 1977).

Theil, H., _Economics and Information Theory_. Chicago: Rand McNally and Company (1967).

Wilson, A. G., _Entropy in Urban and Regional Modelling_. London: Pion Ltd. (1970).