Qüestiió

Quaderns d'Estadística i Investigació Operativa

Any 1994, volum 18, núm. 2
Segona època

Entitats patrocinadores:
Universitat de Barcelona
Universitat Politècnica de Catalunya
Institut d'Estadística de Catalunya

Generalitat de Catalunya
Institut d'Estadística de Catalunya

Any 1994, volum 18, núm. 2

Sumari

PAPER ECOLÒGIC

EDITORIAL

Els dos números anteriors han estat dedicats a publicar algunes de les comunicacions presentades a la SEVENTH INTERNATIONAL CONFERENCE ON MULTIVARIATE ANALYSIS (Barcelona, 21 al 24 de setembre de 1992).

Amb aquest número **Qüestiió** torna a incorporar la secció docent, amb un article sobre com explicar regressió ortogonal mitjançant components principals, i quatre nous problemes, juntament amb les solucions dels problemes que varen ser proposats al Volum 17, número 2.

La secció Estadística General i Investigació Operativa, conté dos articles d'estadística sobre estimació en models de superpoblació i sobre decisió multiobjectiu, i dos articles d'investigació operativa, sobre optimalitat en programació semi-infinita lineal i sobre sensibilitat de sistemes sota perturbació, respectivament.

La part d'Estadística Oficial conté tres articles que aprofundeixen alguns aspectes d'aquest tema, tot seguint les orientacions establertes per **Qüestiió**.

En l'article "Eurostat, la oficina estadística de la Comisión Europea", el senyor Fernando de Esteban, Director de Difusió i de Relacions Públiques d'Eurostat, descriu les missions d'aquest organisme, l'organització del sistema estadística comunitari, les realitzacions d'Eurostat i els nous reptes. Es tracta d'un tema que és fonamental per situar el sistema estadístic de Catalunya en relació amb els altres. Per altra banda les accions i els programes de recerca d'Eurostat afecten cada cop més el món acadèmic i estadístic de la comunitat.

En l'article "El programa anual d'actuació estadística" la senyora Montserrat Biosca, de l'Institut d'Estadística de Catalunya, presenta aquest instrument. La Llei del pla estadístic, que fou comentada en el número anterior, preveu la preparació i elaboració dels plans anuals. El present article desenvolupa aquest segon nivell, n'explica el funcionament general i informa específicament del pla de 1992 i del de 1994.

Finalment en l'article "Els convenis de col·laboració amb l'Institut d'Estadística de Catalunya" els senyors Jordi Bacaria, Joaquim Capellades, Alexandre Costa i Manuel Falguera expliquen el fonament i la forma jurídica d'aquests convenis. També ens informen dels convenis específics establerts amb l'Administració de l'Estat, amb l'Administració local i amb les universitats catalanes.

Estadística General i Investigació Operativa

QÜESTIIÓ, Vol 18, 2 pp. 153-172, 1994

A METHOD OF MULTIOBJECTIVE DECISION MAKING USING A VECTOR VALUE FUNCTION

S. RÍOS-INSUA, J.G. PACHÓN, A. MATEOS*

Universidad Politécnica de Madrid

A decision situation with partial information on preferences by means of a vector value function is assumed. The concept of minimum value dispersion solution as a reference point joined with a pseudodistance function from such a point and a dispersion level ϵ, lead to the notion of $\epsilon-$dispersion set. The dispersion level represents the amount of "value" that the decision maker can be indifferent to, therefore he should choose his most preferred solution in this set. Convergence properties, as well as an interactive method based on the reduction of $\epsilon-$dispersion sets by means of parametric variation of ϵ, to aid decision making in discrete problems is considered. Detailed numerical examples are included.

Key words: Multiobjective decision making, vector value function, efficient set, partial information on preferences.

* S. Ríos-Insua, J.G. Pachón, A. Mateos. Departamento de Inteligencia Artificial. Facultad de Informática. Universidad Politécnica de Madrid. Campus de Montegancedo. 28660 Madrid, SPAIN.

–Article rebut el novembre de 1993.
–Acceptat l'abril de 1994.

1. INTRODUCTION

In this paper we propose an interactive method to aid decision making in solving deterministic multiobjective problems. It tries to help a decision maker (DM) to come up with a decision in the value efficient set, combining optimization and satisficing points of view. The method we present, assumes a vector value function and makes use of an evaluation function obtained as the scalar product of the vector value function and a generic weights vector (scaling), which leads to assume an imprecise weighted additive evaluation function (see Debreu, 1960 or Krantz *et al.*, 1971). A minimum dispersion solution, which is the one with smaller value difference over an information set about component value weights (optimization), is obtained. Such solution, as a reference point to choose the DM, is theoretically provided. Next, a pseudodistance function from a minimum dispersion solution and a dispersion level, which is considered as the amount of value that the DM can be indifferent to, leads to the dispersion set (satisficing). The procedure uses a parametric variation of the dispersion level which produces an interactive reduction of the dispersion set, whose convergence is proved. Different parametric variations are possible and, in practice, the method stops when the dispersion set has been reduced enough for the DM to choose his most preferred solution.

Consider the problem of multiple objective (or vector) optimization: A set $X \subset \mathbb{R}^N$ of alternatives or decisions $\mathbf{x}$ called decision or attribute space and a set of objective functions $\mathbf{z} = (z_1, \ldots, z_n)$, which the DM wishes to maximize. In this way, a function $\mathbf{z}$ defined on Z which take values on the objective or solution space $\mathbb{R}^n$ is defined, with

$$\mathbf{z} = \mathbf{z}(\mathbf{x}) \in Z \quad \text{and} \quad Z = \mathbf{z}(X) \subseteq \mathbb{R}^n$$

where Z is the feasible region in the solution space.

Three different situations depending on the available information on preferences for the DM are considered.

1. The only one information in Z is "more is better" for each objective $z_i (i = 1, \ldots, n)$, which leads to the *null information problem* and we formulate it as

   $$\max_{\mathbf{x} \in X} \mathbf{z}(\mathbf{x})$$

   In this problem, which has been widely studied, arises in a natural way the concept of efficient (Pareto optimal or nondominated) point. Preference optimal points must be efficients and the set of efficient points, defined as

   $$E(X, \mathbf{z}) = \{\mathbf{x} \in \mathbf{X} : \nexists\ \mathbf{x}' \in X \text{ with } \mathbf{z}(\mathbf{x}') \geq \mathbf{z}(\mathbf{x})\}$$

where $[\mathbf{z}(\mathbf{x}') \geq \mathbf{z}(\mathbf{x}) \Leftrightarrow z_i(\mathbf{x}') \geq z_i(\mathbf{x}) \forall i$ and $\mathbf{z}(\mathbf{x}') \neq \mathbf{z}(\mathbf{x})]$, will include the most preferred solution. There are several methods available to generate, under some conditions, all the elements of $E(X, \mathbf{z})$(see among others the books of Goicoechea *et al.*,1982; Chankong *et al.*, 1983; Steuer, 1986).

2. There is complete information on preferences in the sense that we may assess a value function (Debreu, 1954; Fishburn, 1964; Keeney *et al.*, 1993) denoted by v, which is a real-valued function defined on Z and represents the DM preferences by means of a strict weak order $\succ$ (asymmetric and negatively transitive), such that

$$\mathbf{z} \succ \mathbf{z}' \Leftrightarrow v(\mathbf{z}) \geq v(\mathbf{z}')$$

   We call it *complete information problem* and we formulate it as

$$\max_{\mathbf{z} \in Z} v(\mathbf{z})$$

   which it is a classical optimization problem and the optimal solution will be a point of the set

$$\mathrm{Opt}(Z, v) = \{\mathbf{z}' \in Z \colon v(\mathbf{z}') = \max v(\mathbf{z}), \mathbf{z} \in \mathbf{Z}\}$$

3. The DM provides information not enough to assess a value function, but to assess a vector value function $\mathbf{v} : Z \rightarrow \mathbb{R}^p$, which represents (Roberts, 1979; Rietveld, 1980; Ríos-Insua, 1980) a strict partial order $\succ$ (irreflexive and transitive) intersection of p strict weak orders, where

$$\mathbf{z} \succ \mathbf{z}' \Leftrightarrow \mathbf{v}(\mathbf{z}) \geq \mathbf{v}(\mathbf{z}')$$

   We call this situation, which may be considered intermediate between 1 and 2, *partial information problem.* Note that, if there were no more information on preferences in the feasible region $\mathbf{v}(Z) \subseteq \mathbb{R}^p$, we would be in case 1. On other hand, if all the components v_i in $\mathbf{v}$ were equals, would be in case 2.

From a general point of view, the terminology, partial or incomplete information, is also used in the case under uncertainty for utility functions, probability distributions and the evaluation function. Several methodologies have been proposed to deal with this problem on partial information (Chankong *et al.*, 1983; Ríos *et al.*, 1989), but in in our context, we shall refer to the certainty case with a vector value function, that may be seen as a way for lack of precision of the true DM's (scalar) value function and is suitable for hierarchical structures which often exhibits the multiple objective decision making problems. The problem formulation is formally analogous to 1, but in the solution space. We have

$$\max_{\mathbf{z} \in Z} \mathbf{v}(\mathbf{z})$$

and leads us to the value efficient set

$$E(Z, \mathbf{v}) = \{\mathbf{z} \in \mathbf{Z} : \nexists \ \mathbf{z}' \in Z \text{ with } \mathbf{v}(\mathbf{z}') \geq \mathbf{v}(\mathbf{z})\}$$

where

$$E(X, \mathbf{z}) \supseteq \mathbf{z}^{-1}(E(Z, \mathbf{v})) = \{\mathbf{x} \in \mathbf{X} : \mathbf{z}(\mathbf{x}) \in E(Z, \mathbf{v})\}$$

assuming that $\mathbf{z}$ is an increasing function, that is,

$$\mathbf{z}(\mathbf{x} + \Delta\mathbf{x}) > \mathbf{z}(\mathbf{x}) \text{ for } \mathbf{x} \in \mathbf{X}, \Delta\mathbf{x} \geq \mathbf{0} \text{ and } \mathbf{x} + \Delta\mathbf{x} \in \mathbf{X}$$

where $[\mathbf{z}(\mathbf{x} + \Delta\mathbf{x}) > \mathbf{z}(\mathbf{x}) \Leftrightarrow z_i(\mathbf{x} + \Delta\mathbf{x}) > z_i(\mathbf{x}) \forall i]$

Several important useful concepts have been developed by researchers for the case of partial information on value (or utility) functions as Fishburn (1964,1965), Sarin (1977), Hannan (1981), Kirkwood *et al.* (1985), Korhonen *et al.* (1984), White *et al.* (1984), Weber (1985), Malakooti (1989), Ríos-Insua (1990) and Kirkwood (1992) among others, and an interesting overview within a general framework is found in Weber (1987). Therefore, we shall not consider it here. On other hand, there are also a number of helpful computer programs which aids to assess value and utility functions (i.e., Keeney *et al*, 1976; Kirkwood *et al.*, 1986,1987; Logical Decision, 1992; Decision Pad, 1993).

The paper consist of five sections. In the first section the problem is formulated with a brief overview of the context. In the second and third sections, solution concepts, as well as, monotonicity and convergence properties are provided. The fourth section synthesize the method into an algorithm for solving discrete problems. Some numerical examples are given in the last section.

2. SOLUTION CONCEPTS

Given the partial information problem, it leads us to determine the value efficient set $E(Z, \mathbf{v})$ and so $\mathbf{z}^{-1}(E(Z, \mathbf{v}))$, and considering that such set, reduced from $E(Z, \mathbf{z})$, could be still too extensive for the DM to choose an alternative, we propose a method to aid him.

Let us consider a vector value function $\mathbf{v}: Z \longrightarrow \mathbb{R}^p$ and the set $K^\circ \equiv \mathbb{R}^p_+$, which is a convex cone.

Definition 1

Let be $K \supseteq K^{\circ}$ a constant, convex, closed and acute cone which we call *information cone*, and K^p its positive polar. The set

$$K_* = K^p \cap S_p$$

is called *information set* associated to K, where S_p is the unit sphere on R^p.

Let $E(Z, K)$ be the efficient set with respect to $K(\mathbf{z} \in E(Z, K)$ if there is no $\mathbf{z}' \in Z$ such that $\mathbf{z}' \in \mathbf{z} + K)$. It may be seen that, if K^p is a polyhedral cone it will be possible to determine its set of generators $\{\mathbf{k}^1, \ldots, \mathbf{k}^r\}$(Tamura, 1976), which normalized on S_p are denoted $\{\mathbf{k}_*^1, \ldots, \mathbf{k}_*^r\}$.

Definition 2

For each $\mathbf{z} \in \mathbf{Z}$, the numbers

$$v^*(\mathbf{z}) = \max_{\mathbf{k} \in K_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) \qquad v_*(\mathbf{z}) = \min_{\mathbf{k} \in K_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z})$$

are called *upper and lower indexes*, respectively.

Let be $E_K \equiv \{\mathbf{z} \in E(Z, \mathbf{v}) \colon \mathbf{v}(\mathbf{z}) \in E(\mathbf{v}(Z), K)\}$ and $E = E_{K^{\circ}}$ in what follows. Assuming that $v^*(\mathbf{z})$ y $v_*(\mathbf{z})$ represents for each $\mathbf{z}$, the best and the worst values under the information cone K, a criterion to determine a solution in E_K consists of considering those solutions whereby the difference between both indexes will be as small as possible. This setting is based on the idea that carries to choose those points of the efficient set in which the information cone approximates the information given by a value function (scalar) in which both indexes obviously are equal. We introduce a function which measures such difference on E_K.

Definition 3

The function $d_K : E_K \to \mathbb{R}^+$, defined as

$$d_K(\mathbf{z}) = v^*(\mathbf{z}) - v_*(\mathbf{z})$$

is called *dispersion function* or *dispersion* associated to $\mathbf{z}$ under the information cone K.

Note that $v^*(\mathbf{z}) \geq v_*(\mathbf{z})$ for each $\mathbf{z} \in E_K$, therefore, it is always $d_K(\mathbf{z}) \geq 0$. On other hand, we can introduce the ordering $\geq_{(K)}$ on $E_K \times E_K$ defined by

$$\mathbf{z} \geq_{(K)} \mathbf{z}' \Leftrightarrow d_K(\mathbf{z}) \leq d_K(\mathbf{z}')$$

which is a strict linear order (asymmetric, transitive and complete) and leads to a total ordering on E_K. As the most preferred solution, we shall consider that with smaller dispersion.

Definition 4

A solution $\mathbf{z}' \in E_K$ such that

$$d_K(\mathbf{z}') = \min_{\mathbf{z} \in E_K} d_K(\mathbf{z})$$

is called a *minimum dispersion solution* under K.

The set

$$D(Z, K) = \{\mathbf{z}' \in E_K : d_K(\mathbf{z}') = \min d_K(\mathbf{z}), z \in E_K\}$$

is called *minimum dispersion set.*

Once $\mathbf{v}$ is assessed, if there is no more information on preferences on $\mathbf{v}(Z) \subseteq \mathbb{R}^p$, the information cone is $K^\circ \subseteq \mathbb{R}^p$ and the ordering will be $\geq_{(K^\circ)}$. However, the DM may provide more information on his preferences by means of a cone $K \supseteq K^\circ$, what leads to a smaller dispersion.

Proposition 1

Let be K' and K information cones such that $K' \supseteq K$ and $\mathbf{z} \in E_K$, then $d_{K'}(\mathbf{z}) \leq d_K(\mathbf{z})$ and

$$\min_{\mathbf{z} \in E_{K'}} d_{K'}(\mathbf{z}) \leq \min_{\mathbf{z} \in E_K} d_K(\mathbf{z})$$

Proof

We observe that if $K' \supseteq K$ it is $K'_* \subseteq K_*$, $E_{K'} \subseteq E_K$ and

$$\max_{\mathbf{k} \in K'_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) \leq \max_{\mathbf{k} \in K_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) \text{ and } \min_{\mathbf{k} \in K'_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) \geq \min_{\mathbf{k} \in K_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z})$$

hence $d_{K'}(\mathbf{z}) \leq d_K(\mathbf{z})$ for each $\mathbf{z} \in E_K$, and thus

$$\min_{\mathbf{z} \in E_K} (\max_{\mathbf{k} \in K'_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) - \min_{\mathbf{k} \in K'_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z})) \leq \min_{\mathbf{z} \in E_K} (\max_{\mathbf{k} \in K_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) - \min_{\mathbf{k} \in K_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}))$$

which shows the second inequality.

■

Let us now consider a result which states that if a solution $\mathbf{z}$ has minimum dispersion for an increasing sequence of information cones, whose information sets converge to a vector $\mathbf{k}^+$, such solution maximizes the value function $(\mathbf{k}^+ \cdot \mathbf{v})$.

Proposition 2

Let be $\mathbf{z} \in E$ and $\{K^n\}_{n=0}^{\infty}$ an increasing sequence of information cones such that $K_*^n \downarrow \mathbf{k}^+$. If $\mathbf{z} \in D(Z, K^n)$ for each $n = 0, 1, \ldots$, then $\mathbf{z} \in \mathrm{Opt}(Z, (\mathbf{k}^+ \cdot \mathbf{v}))$.

Proof

If $\mathbf{z} \in D(Z, K^n)$ for each n, then $\mathbf{v}(\mathbf{z}) \in E(\mathbf{v}(Z), K^n)$ for each n. Because of the nesting property of the efficient sets with respect to the sequence $\{K^n\}$, it will be $\mathbf{v}(\mathbf{z}) \in E(\mathbf{v}(Z), K^+)$, where K^+ has as polar positive the vector $\mathbf{k}^+$ and thus $\mathbf{z} \in \mathrm{Opt}(Z, (\mathbf{k}^+ \cdot \mathbf{v}))$.

■

We note that $(\mathbf{k}^+ \cdot \mathbf{v})$ represents the value function when all the uncertainty on preferences has been removed and, consequently, we would have a complete information problem. Observe in this last case that the dispersion of all solutions would be zero, we have no objection to this, because we have a value function that provides a total ordering.

We shall not consider here effective ways to asses information sets (cones) based on preferences, but some methods can be found in Malakooti (1989).

3. PSEUDODISTANCE ON DISPERSION

The above criterion, based on minimum dispersion, may be too strict in some cases, so we are going to introduce a less restrictive criterion in the idea of the satisficing approach, that combined with the former, leads to an aid decision making method.

In this satisficing criterion an amount $\epsilon > 0$ is considered, which means the maximum "amount of value" that the DM can ignore or be indifferent to, and the solution(s) to the problem will be those whose distance from a minimum dispersion solution will be smaller than ϵ.

Definition 5

We call *pseudodistance on dispersion* between $\mathbf{z}, \mathbf{z}' \in E_K$, to the mapping $\rho_K : E_K \times E_K \longrightarrow \mathbb{R}^+$ where

$$\rho_K(\mathbf{z}, \mathbf{z}') = |d_K(\mathbf{z}) - d_K(\mathbf{z}')|\,.$$

Proposition 3

The mapping ρ_K is a pseudodistance on E_K.

Proof

a) As a consequence of the definition is immediate that $\rho_K(\mathbf{z}, \mathbf{z}') \geq 0$.

b) For the triangular inequality, note that

$$\rho_K(\mathbf{z}, \mathbf{z}'') = |d_K(\mathbf{z}) - d_K(\mathbf{z}'')| =$$

adding and subtracting $d_K(\mathbf{z}')$

$$= |d_K(\mathbf{z}) - d_K(\mathbf{z}') + d_K(\mathbf{z}') - d_K(\mathbf{z}'')| \leq |d_K(\mathbf{z}) - d_K(\mathbf{z}')| + |d_K(\mathbf{z}') - d_K(\mathbf{z}'')|$$
$$= \rho_K(\mathbf{z}, \mathbf{z}') + \rho_K(\mathbf{z}', \mathbf{z}'')$$

with the last inequality, because of definition 5, proves the triangular inequality for all $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in E_K$.

c) It is trivial from definition 5 that $\rho_K(\mathbf{z}, \mathbf{z}') = \rho_K(\mathbf{z}', \mathbf{z})$.

d) We shall see with an example that can exist two solutions $\mathbf{z}, \mathbf{z}'$ with $\mathbf{z} \neq \mathbf{z}'$, but $\rho_K(\mathbf{z}, \mathbf{z}') = 0$ and so ρ_K will be a pseudodistance.

Let us consider a problem where $\mathbf{v}(\mathbf{z})$ is defined on $\mathbb{R}^2$ and a polyhedral information cone K whose information set K_*, is given by the generators $\{(1/\sqrt{5}, 2/\sqrt{5}), (2/\sqrt{5}, 1/\sqrt{5})\}$. Let be $\mathbf{z}, \mathbf{z}'$ with $\mathbf{z} \neq \mathbf{z}'$ and $\mathbf{v}$ an undefined vector value function such that

$$\mathbf{v}(\mathbf{z}) = (1,\ 0) \text{ and } \mathbf{v}(\mathbf{z}') = (0, 1)$$

From previous definitions we have

$$v^*(\mathbf{z}) = \max_{\mathbf{k} \in K_*} (k_1 \cdot 1 + k_2 \cdot 0) = \max_{\mathbf{k} \in K_*} k_1 = \frac{2}{\sqrt{5}}$$

and analogously

$$v_*(\mathbf{z}) = \min_{\mathbf{k} \in K_*} k_1 = \frac{1}{\sqrt{5}}$$

Hence

$$d_K(\mathbf{z}) = \frac{2}{\sqrt{5}} - \frac{1}{\sqrt{5}} = \frac{1}{\sqrt{5}}$$

In a similar way, we obtain for $\mathbf{z}'$

$$v^*(\mathbf{z}') = \max_{\mathbf{k}\in K_*} k_2 = \frac{2}{\sqrt{5}} \quad \text{and} \quad v_*(\mathbf{z}') = \min_{\mathbf{k}\in K_*} k_2 = \frac{1}{\sqrt{5}}$$

and $d_K(\mathbf{z}') = 1/\sqrt{5}$. Then, $\rho_K(\mathbf{z}, \mathbf{z}') = 0$ but $\mathbf{z} \neq \mathbf{z}'$.

∎

Definition 6

Given $\mathbf{z}^+ \in D(Z, K)$ and a real number $\epsilon > 0$, a solution $\mathbf{z} \in E_K$ such that $\rho_K(\mathbf{z}, \mathbf{z}^+) \leq \epsilon$, is called *$\epsilon-$dispersion solution.*

The number ϵ is called *dispersion level* and the set

$$D_\epsilon(Z, K, \mathbf{z}^+) = \{\mathbf{z} \in E_K : \rho_K(\mathbf{z}, \mathbf{z}^+) \leq \epsilon\}$$

$\epsilon-$dispersion set (with respect to $\mathbf{z}^+$).

The $\epsilon-$dispersion concept is more general than the one of minimum dispersion

Proposition 4

Given $\mathbf{z}^+ \in D(Z, K)$ and a dispersion level ϵ, then

$$D(Z, K) \subseteq D_\epsilon(Z, K, \mathbf{z}^+)$$

Proof

If $\mathbf{z} \in D(Z, K)$ then

$$d_K(\mathbf{z}) = d_K(\mathbf{z}^+) = \min_{\mathbf{z}\in E_K} d_K(\mathbf{z})$$

hence $\rho_K(\mathbf{z}, \mathbf{z}^+) \leq \epsilon$ for all $\epsilon > 0$.

∎

Now, we consider approximation and convergence principles for the ϵ-dispersion sets.

Theorem 1

Given dispersion levels ϵ_1, ϵ_2, where $\epsilon_1 < \epsilon_2$, then for all $\mathbf{z}^+ \in D(Z, K)$ is

$$D_{\epsilon_1}(Z, K, \mathbf{z}^+) \subseteq D_{\epsilon_2}(Z, K, \mathbf{z}^+).$$

Proof

It is immediate.

■

Theorem 2

Let be $\mathbf{z}^+ \in D(Z, K)$ and $\{\epsilon_n\}_{n\in\mathbb{N}}$ a sequence of dispersion levels such that $\epsilon_n \downarrow 0$, then

$$D_{\epsilon_n}(Z, K, \mathbf{z}^+) \downarrow D(Z, K)$$

(in the sense that $\bigcap_{n\in\mathbb{N}} D_{\epsilon_n}(Z, K, \mathbf{z}^+) = D(Z, K)$).

Proof

Let us call $D_{\epsilon_n}(Z, K, \mathbf{z}^+) = D_n$ and $D(Z, K)$=D.

Because $D \subset D_n$ for all n, then $D \subset \bigcap_n D_n$.

To show the other content, note that if $\mathbf{z}' \in \bigcap_n D_n$ then

$$\rho_K(\mathbf{z}', \mathbf{z}^+) \leq \epsilon_n \text{ for all } n$$

so, taking limits and because ρ_K is non negative, we have

$$\rho_K(\mathbf{z}', \mathbf{z}^+) = 0$$

thus, we obtain that

$$d_K(\mathbf{z}') = d_K(\mathbf{z}^+) = \min_{\mathbf{z}\in E_K} d_K(\mathbf{z})$$

and $\mathbf{z}' \in D$.

■

4. ALGORITHM FOR DISCRETE PROBLEMS

Although the above developments are theoretically valid for continuous problems, due to the difficulty to determine E_K, specially in the nonlinear case, a method for multicriteria evaluation for discrete problems with its corresponding algorithm is now considered.

Firstly, the algorithm computes the minimum dispersion set $D \equiv D(Z,K)$ as well as the associated dispersion value. Then, for a solution $\mathbf{z}^+ \in D$ and an initial dispersion level ϵ, the ϵ-dispersion set $D_\epsilon \equiv D_\epsilon(Z,K,\mathbf{z}^+)$ is computed and then presented to the DM to make a choice if possible or otherwise, take a smaller dispersion level and repeat the process.

Let $E_K = \{\mathbf{z}^1, \ldots, \mathbf{z}^q\}$, i the iteration index, M a large positive number, j the index for the dispersion level sequence $\epsilon_j (= \epsilon j)$, $q(j)=$ card$(D_{\epsilon j})$ and F a subset of E_K. We assume that the indexes of the solutions in $D_{\epsilon j}$ are appropriately renumbered from 1 to $q(j)$ in each iteration ($q(1) = q$). The algorithm is as follows

Step 1. Set $i = 0, d = M$ and $D= \emptyset$.

Step 2. Set $i = i + 1$.

Step 3. Choose $\mathbf{z}^i \in E_K$ and compute $d^i = d_K(\mathbf{z}^i)$. If $d^i > d$, go to step 4. If $d^i = d$, set $D = D \cup \{\mathbf{z}^i\}$ and go step 4. Otherwise, set $d = d^i, D = \{\mathbf{z}^i\}$ and go to step 4.

Step 4. If $i < q$, go to step 2. Otherwise, D and d are identified and let be $\mathbf{z}^+$ a solution in D.

Step 5. Set $j = 1$ and $F= \emptyset$.

Step 6. Set $i = 0, D_{\epsilon j} = E_K \backslash F, \epsilon_j = \epsilon / j$ and $q = q(j)$.

Step 7. Set $i = i + 1$.

Step 8. Choose $\mathbf{z}^i \in D_{\epsilon j}$ and compute $\rho_K(\mathbf{z}^i, \mathbf{z}^+)$. If $\rho_K(\mathbf{z}^i, \mathbf{z}^+) \leq \epsilon_j$ go to step 9. Otherwise set $D_{\epsilon j} = D_{\epsilon j} \backslash \{\mathbf{z}^i\}$ and go to step 9.

Step 9. If $i < q$, go to step 7. Otherwise $D_{\epsilon j}$ has been identified.

Step 10. If $D_{\epsilon j}$ is satisfactory for the DM to choose one solution, stop. Otherwise, set $F = D^c_{\epsilon j} (= E_K \backslash D_{\epsilon j})$.

Step 11. Set $j = j + 1$ and go to step 6.

Let us now consider some computational aspects of the method.

$\boxed{1.}$ In the first part of the algorithm it is necessary to compute the indexes $v^*(\mathbf{z})$ and $v_*(\mathbf{z})$ to determine the solution's dispersion. In the case of polyhedral cones, what is usually considered, we shall propose a method based in the Lagrange multipliers, easy to implement.

Let K be a cone and K_* its positive polar (on the unit sphere) with generators $G = \{\mathbf{k}^1, \ldots, \mathbf{k}^r\}$ and $G_* = \{\mathbf{k}_*^1, \ldots, \mathbf{k}_*^r\}$, respectively. Let be $E_K = \{\mathbf{z}^1, \ldots . \mathbf{z}^q\}$ as before. The scheme is as follows: Let be

$$E_K^1 = \{\mathbf{z}^i \in E_K : \mathbf{H} \cdot (\mathbf{v}(\mathbf{z}^i))^T \geq \mathbf{0}^T\}$$

where $\mathbf{H}$ is a $r \times p$ matrix with $\mathbf{k}^j$ being the jth row of $\mathbf{H}$, and

$$E_K^2 = E_K \setminus E_K^1$$

In E_K^1 and E_K^2 we shall compute the lower index

$$v_*(\mathbf{z}^i) = \min_{\mathbf{k} \in G_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}^i)$$

and the upper index as

$$v^*(\mathbf{z}^i) = (\mathbf{k}^u \cdot \mathbf{v})(\mathbf{z}^i)$$

with $\mathbf{k}^u = \mathbf{v}(\mathbf{z}^i)$ in E_K^1, and in E_K^2 as

$$v^*(\mathbf{z}^i) = \max_{\mathbf{k} \in G_*} (\mathbf{k} \cdot \mathbf{v})(\mathbf{z}^i)$$

To explain why the maximum is reached on such a point, and the minimum on G_*, we use the Lagrange method. Let us consider the problem

$$\max(\mathbf{k} \cdot \mathbf{v})(\mathbf{z}) = k_1 v_1(\mathbf{z}) + \cdots + k_p v_p(\mathbf{z})$$

subject to

$$k_1^2 + \cdots + k_p^2 = 1$$

(that is, $(k_1, \ldots, k_p) \in S_p$) and consider the Lagrange function

$$L(k_1, \ldots, k_p) = k_1 v_1(\mathbf{z}) + \cdots + k_p v_p(\mathbf{z}) + \lambda(k_1^2 + \cdots + k_p^2 - 1)$$

Given the partial derivatives of L, we obtain the system

$$(1) \qquad \left\{ \begin{array}{rcl} v_i(\mathbf{z}) + 2\lambda k_i & = & 0, \quad i = 1, \ldots, p \\ k_1^2 + \cdots + k_p^2 & = & 1 \end{array} \right.$$

From any of the above equations where $v_i(\mathbf{z}) \neq 0$, we obtain

$$\lambda = -\frac{v_i(\mathbf{z})}{2k_i} \quad \text{and} \quad k_j = \frac{k_i v_j(\mathbf{z})}{v_i(\mathbf{z})}$$

for any $j \neq i$. If we substitute each k_j in (1), we have

$$k_i^2 \left(\frac{v_1^2(\mathbf{z})}{v_i^2(\mathbf{z})} + \cdots + \frac{v_i^2(\mathbf{z})}{v_i^2(\mathbf{z})} + \cdots + \frac{v_p^2(\mathbf{z})}{v_i^2(\mathbf{z})} \right) = 1$$

so

$$k_i = \frac{\pm v_i(\mathbf{z})}{\sqrt{v_i^2(\mathbf{z}) + \cdots + v_p^2(\mathbf{z})}}$$

and the two critical points will be

$$\left(\frac{v_1(\mathbf{z})}{\sqrt{v_1^2(\mathbf{z}) + \cdots + v_p^2(\mathbf{z})}}, \ldots, \frac{v_p(\mathbf{z})}{\sqrt{v_1^2(\mathbf{z}) + \cdots + v_p^2(\mathbf{z})}} \right)$$

and the opposite.

$\boxed{2.}$ To start the method it is necessary an initial value ϵ to iterate. A way to have an idea about this value may be to put

$$\epsilon = \max_{\mathbf{z}} \rho_K(\mathbf{z}, \mathbf{z}^+)$$

Note that with this initial value, the first D_ϵ set would be E_K.

$\boxed{3.}$ In the algorithm, the considered parameterization is $\epsilon_j = \epsilon/j$, however, others parameterizations could be proposed depending on the characteristics of the problem under consideration (i.e., number of alternatives, speed convergence).

5. EXAMPLES

A) Let the structure of the *DM*'s vector value function be linear

$$\mathbf{v}(\mathbf{z}) = (v_1(\mathbf{z}), v_2(\mathbf{z})) = (4z_1 + z_2, z_1 + 6z_2)$$

that must be maximized and let be $Z \subseteq \mathbb{R}^2$ the set of solutions whose efficient set $E(Z, K^\circ)$ for the Pareto order is shown in table 1.

Table 1

	$\mathbf{z}^1$	$\mathbf{z}^2$	$\mathbf{z}^3$	$\mathbf{z}^4$	$\mathbf{z}^5$	$\mathbf{z}^6$	$\mathbf{z}^7$	$\mathbf{z}^8$	$\mathbf{z}^9$
z_1^i	3	4	8	8.2	9	11	12	15	15.2
z_2^i	12	8	7	6.1	6	5	3	2	1

Table 2

	$\mathbf{z}^1$	$\mathbf{z}^2$	$\mathbf{z}^3$	$\mathbf{z}^4$	$\mathbf{z}^5$	$\mathbf{z}^6$	$\mathbf{z}^7$	$\mathbf{z}^8$	$\mathbf{z}^9$
$v_1(\mathbf{z}^i)$	24	24	39	38.9	42	49	51	62	61.8
$v_2(\mathbf{z}^i)$	75	52	50	44.8	45	41	30	27	21.2

We determine for each point $\mathbf{z}^i$ their values v_1 and v_2 as shown in table 2 and figure 1a), and thus $E = \{\mathbf{z}^1, \mathbf{z}^3, \mathbf{z}^5, \mathbf{z}^6, \mathbf{z}^7, \mathbf{z}^8\}$, because $\mathbf{z}^1$ dominates $\mathbf{z}^2, \mathbf{z}^3$ dominates $\mathbf{z}^4$ and $\mathbf{z}^8$ dominates $\mathbf{z}^9$.

We consider three cases corresponding each one to a different information set (figures 1b), c) and d)).

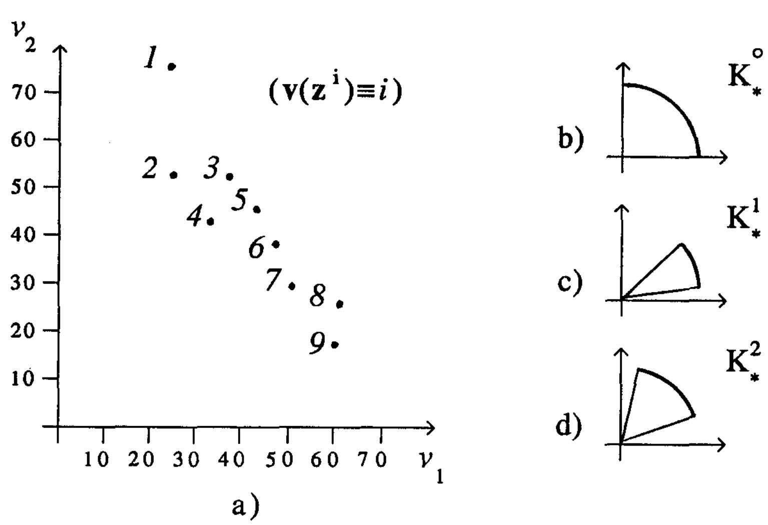

Figure 1.

a) Graphical illustration of the solutions on the $v_1 v_2$ space. b), c) and d), three possible information sets.

A1. Assume that K_*° is the (null) information set (figure 1*b*), with generators $\{(1,0),(0,1)\}$ and let us compute D and d. Initially, let $M = 10^8$. Choosing $\mathbf{z}^1 \in E$, its dispersion value is

$$d^1 = d_{K^\circ}(\mathbf{z}^1) = v^*(\mathbf{z}^1) - v_*(\mathbf{z}^1) = 78.7 - 24 = 54.7$$

Because $d^1 = 54.7 < 10^8 = M$, we set $d = 54.7$ and continue to proceed finally in iteration six of the first part of the algorithm (steps 1 to 4), to obtain $\mathbf{z}^+ = \mathbf{z}^5$ and hence

$$D = \{\mathbf{z}^5\} \text{ and } d = d^5 = 19.6$$

The value indexes, as well as the, dispersion values of all solutions in E are shown in table 3, from which we see that the preference ordering will be

$$\mathbf{z}^5 \succ \mathbf{z}^6 \succ \mathbf{z}^3 \succ \mathbf{z}^7 \succ \mathbf{z}^8 \succ \mathbf{z}^1$$

Table 3

	$\mathbf{z}^1$	$\mathbf{z}^3$	$\mathbf{z}^5$	$\mathbf{z}^6$	$\mathbf{z}^7$	$\mathbf{z}^8$
$v^*(\mathbf{z}^i)$	78.7	63.4	61.6	63.9	59.2	67.6
$v_*(\mathbf{z}^i)$	24	39	42	41	30	27
d_{K°	54.7	24.4	19.6	22.9	29.2	40.6
$\rho_K(\mathbf{z}^i, \mathbf{z}^+)$	35.1	4.8	0	3.3	9.6	21

In the second part of the algorithm (steps 5 to 11), because $\max \rho_K(\mathbf{z}^i, \mathbf{z}^+) = 35.1$, we initially take $\epsilon = 36$ and choose the parametric variation $\epsilon_j = \epsilon/2^{j-1}$. In the first iteration $D_{36} = E$ and, in the fifth one, the algorithm stops because we obtain the unitary set $D_{2.25} = \{\mathbf{z}^5\}$. The iterations, with their respective ϵ-dispersion sets, are shown in table 4, where we see the monotonicity property.

Table 4

Iteration	ϵ	D_ϵ
1	36	E
2	18	$\mathbf{z}^3, \mathbf{z}^5, \mathbf{z}^6, \mathbf{z}^7$
3	9	$\mathbf{z}^3, \mathbf{z}^5, \mathbf{z}^6$
4	4.5	$\mathbf{z}^5, \mathbf{z}^6$
5	2.25	$\mathbf{z}^5$

If the DM states that 4.5 is a satisficing dispersion level he must choose the most preferred solution in $D_{4.5} = \{\mathbf{z}^5, \mathbf{z}^6\}$.

A2. Now, if the DM states that the information set is given by the vectors $\mathbf{k}^1 = (0.8, 0.2)$ and $\mathbf{k}^2 = (0.5, 0.5)$, which normalized on S_2 gives us the extreme vectors for K_*^1(figure 1*c*)

$$\mathbf{k}_*^1 = (0.97, 0.24) \text{ and } \mathbf{k}_*^2 = (0.71, 0.71)$$

we shall have $E_{K^1} = \{\mathbf{z}^1, \mathbf{z}^6, \mathbf{z}^8\}$ because under the corresponding information, $\mathbf{z}^6$ dominates $\mathbf{z}^3, \mathbf{z}^5, \mathbf{z}^7$. The results are shown in table 5 in which we see that

$$D = \{\mathbf{z}^8\} \text{ and } d = 4.4$$

and the preference ordering will be

$$\mathbf{z}^8 \succ \mathbf{z}^6 \succ \mathbf{z}^1$$

Table 5

	$\mathbf{z}^1$	$\mathbf{z}^6$	$\mathbf{z}^8$
$v^*(\mathbf{z}^i)$	70.3	63.9	67.6
$v_*(\mathbf{z}^i)$	41.3	57.4	63.2
d_K	29.0	6.5	4.4
$\rho_K(\mathbf{z}^i, \mathbf{z}^+)$	24.6	2.1	0

Again, we can continue the process, starting with a dispersion level ϵ =25(max ρ_K $(\mathbf{z}^i, \mathbf{z}^+)$=24.6) and for the parametric variation $\epsilon_j = \epsilon/5^{j-1}$, the ϵ-dispersion sets are shown in table 6.

Table 6

Iteration	ϵ	D_ϵ
1	25	E_{K^1}
2	5	$\mathbf{z}^6, \mathbf{z}^8$
3	1	$\mathbf{z}^8$

A3. Finally, if the DM states that the information set is given by the vectors $\mathbf{k}^1 = (0.7, 0.3)$ and $\mathbf{k}^2 = (0.1, 0.9)$ which normalized in S_2 give the extreme vectors for K_*^2(figure 1*c*)

$$\mathbf{k}_*^1 = (0.92, 0.39) \text{ and } \mathbf{k}_*^2 = (0.11, 0.99)$$

lead to $E = \{\mathbf{z}^1, \mathbf{z}^3, \mathbf{z}^5, \mathbf{z}^6, \mathbf{z}^8\}$, because $\mathbf{z}^6$ dominates $\mathbf{z}^7$. In this case, we obtain $D = \{\mathbf{z}^3\}$ and $d = 9.6$ and the preference ordering will be

$$\mathbf{z}^3 \succ \mathbf{z}^5 \succ \mathbf{z}^6 \succ \mathbf{z}^1 \succ \mathbf{z}^8$$

If $\epsilon = 7$, it will be easy to show that $D_7 = \{\mathbf{z}^3, \mathbf{z}^5\}$.

B) In the case of a nonlinear vector value function, as it will be the case of quadratic components, let us see that it is possible to apply in the same way the method. Let us consider quadratic components for $\mathbf{v}$ such that

$$v_1(\mathbf{z}) = -(z_1^2 + z_2^2 - 72z_1 - 20z_2) \text{ and } v_2(\mathbf{z}) = -(z_1^2 + z_2^2 - 16z_1 - 60z_2)$$

and let the information set be defined by the vectors $\mathbf{k}^1 = (0.8, 0.2)$ and $\mathbf{k}^2 = (0.4, 0.6)$, which normalized in S_2 give us the extreme vectors

$$\mathbf{k}_*^1 = (0.97, 0.24) \text{ and } \mathbf{k}_*^2 = (0.55, 0.83)$$

In this case we obtain $E = \{\mathbf{z}^3, \mathbf{z}^6, \mathbf{z}^8\}$ under the corresponding information cone, because $\mathbf{z}^3$ dominates $\mathbf{z}^1, \mathbf{z}^2, \mathbf{z}^4$ and $\mathbf{z}^6$ dominates $\mathbf{z}^5, \mathbf{z}^7$. Also, note that $\mathbf{z}^9$ is not a value efficient solution. The results are shown in table 7, from which we obtain the preference ordering

$$\mathbf{z}^3 \succ \mathbf{z}^6 \succ \mathbf{z}^8$$

with $D = \{\mathbf{z}^3\}$ and $d = 54.2$.

Table 7

	$\mathbf{z}^3$	$\mathbf{z}^6$	$\mathbf{z}^8$
$v^*(\mathbf{z}^i)$	743.5	815.7	895.7
$v_*(\mathbf{z}^i)$	689.3	684.2	598.8
d_K	54.2	131.5	296.9
$\rho_K(\mathbf{z}^i, \mathbf{z}^+)$	0	77.3	242.7

The last row of table 7 is the pseudodistance on dispersion and we can see that for a dispersion level $\epsilon = 80$, it is $D_{80} = \{\mathbf{z}^3, \mathbf{z}^6\}$.

6. CONCLUSIONS

In this paper we have considered multiobjective decision making under partial information on preferences what may leads to determine the DM's vector value function instead of a scalar one. From this vector function, the value efficient set is obtained and considering such set too extensive to choice a solution, a method based on one hand on a minimum dispersion solution concept over a preference information set and on the other, on a dispersion function with respect to a minimum dispersion solution, which it is a pseudodistance, are considered. The idea of dispersion level leads to its associated set which fulfils monotonicity and convergence properties. This set will contain all the indifferent "in value" solutions for a given dispersion level and de DM will choose his solution in this set or will ask for a smaller one. An algorithm for discrete problems with different computational issues supports the method which is proved with some numerical illustrations.

ACKNOWLEDGMENTS

This work has been supported by DGICYT of the Ministry of Education under Grant PB91-0172.

7. REFERENCES

[1] **Chankong V.** and **Haimes Y.Y.** (1983). *Multiobjective Decision Making: Theory and Methodology.* North Holland, Amsterdam.

[2] **Debreu, G.** (1954). "Representation of a Preference Ordering by a Numerical Function". In R. M. Thrall, C. H. Coombs, R. L. Davies (Eds.)., *Decision Processes.* Wiley, New York.

[3] **Debreu, G.** (1960). "Topological Methods in Cardinal Utility Theory". In K.J. Arrow, S. Karlin and P. Suppes (Eds.)., *Mathematical Methods in Social Sciences.* Stanford University Press, Stanford.

[4] **Decision Pad** (1993). APIAN Software, *PO* Box 1224, Menlo Park, CA.

[5] **Fishburn, P.C.** (1964). *Decision and Value Theory.* Wiley, New York.

[6] **Fishburn, P.C.** (1965). "Analysis of Decisions with Incomplete Knowledge Probabilities", *Opn. Res.*, **1, 13**, 217–237.

[7] **Goicoechea, A., D.R. Hansen** and **L. Duckstein** (1982). *Multiobjective Decision Analysis with Engineering and Business Applications.* Wiley, New York.

[8] **Hannan, E.L.** (1981). "Obtaining Nondominated Priority Vectors for Multiple Objective Decisionmaking Problems with Different Combinations of Cardinal and Ordinal Information", *IEEE Trans. Syst. Man Cybernet.* **SMC-11**, 538–543.

[9] **Keeney, R.L.** and **H. Raiffa** (1993). *Decision with Multiple Objectives.* Wiley, New York.

[10] **Keeney, R. L.** and **A. Sicherman** (1976). "Assessing and Analyzing Preferences Concerning Multiple Objectives: An Interactive Computer Program", *Behavioral Sci.*, **21**, 173–182.

[11] **Kirkwood, C.W.** and **R.K. Sarin** (1985). "Ranking with Partial Information: A Method and an Application", *Opns. Res.*, **23**, 38–48.

[12] **Kirkwood, C.W.** and **L.C. van der Feltz** (1986). "Personal Computer Programs for Decision Analysis", Tech. Rep. DIS-86/87-4, *Dept. of Decision and Information Systems*, College of Business, Arizona State University.

[13] **Kirkwood, C.W.** and **L.C. van der Feltz** (1987). "Personal Computer Programs for Multiobjective Decision Analysis Using Power-Additive Utility Functions", Tech. Rep. DIS-87/88-1, *Dept. of Decision and Information Systems*, College of Business, Arizona State University.

[14] **Kirkwood, C.W.** (1992). "An Overview of Methods for Applied Decision Analysis", *Interfaces,* **22, 6**, 28–39.

[15] **Korhonen, P.J. Wallenius** and **S. Zionts** (1984). "Solving the Discrete Multiple Criteria Problem using Convex Cones", *Mgmt. Sci.*, **30**, 1336–1345.

[16] **Krantz D.H., R.D. Luce, P. Suppes** and **A. Tversky A.** (1971). *Foundations of Measurement*, Vol. 1. Academic Press, New York.

[17] **Logical Decision** (1992). *Multi-Measure Decision Analysis Software*, 164 E. Scenic. Ave., Point Richmond, CA, 94801.

[18] **Malakooti, B.** (1989). "Ranking Multiple Criteria Alternatives with Half-Space, Convex, and Non-Convex Dominating Cones: Quasi-Concave and Quasi-Convex Multiple Attribute Utility Functions", *Comput. Opns. Res.*, **16**, 117–127.

[19] **Rietveld, P.** (1980). *Multiple Objective Decision Methods and Regional Planning.* North Holland, Amsterdam.

[20] **Ríos Insua, D.** (1990). *Sensitivity Analysis in Multiobjective Decision Making.* LNEMS 347, Springer, Berlin.

[21] **Ríos, S., M.J. Ríos Insua** and **S. Ríos-Insua** (1989). *Procesos de Decisión Multicriterio.* Eudema, Madrid.

[22] **Ríos-Insua, S.** (1980). "Decisiones Multicriterio con Ordenaciones Parciales", *Ph. D. Thesis,* Editorial de la Universidad Complutense, Madrid.

[23] **Roberts, F.S.** (1979). *Measurement Theory,* Addison Wesley, New York.

[24] **Sarin, R.K.** (1977). "Interactive Evaluation and Bound Procedure for Selecting Multi-Attributes Alternatives". In M. K. Starr, M. Zeleny (Eds.)., *Multiple Criteria Decision Making.* North Holland, New York.

[25] **Steuer, R.** (1986). *Multiple Criteria Optimization: Theory, Computation and Application.* Wiley, New York.

[26] **Tamura, K.** (1976). "A Method for Constructing the Polar Cone of a Polyhedral Cone, with Applications to Linear Multicriteria Decision Problems", *J. Optim. Theory Appl.*, **19**, 547–564.

[27] **Weber, M.** (1985). "A Method for Multi-Attribute Decision Making with Incomplete Information", *Mgmt. Sci.*, **31**, 1365–1371.

[28] **Weber, M.** (1987). "Decision Making with Incomplete Information", *Eur. J. Opl. Res.*, **28**, 44–57.

[29] **White III, C. C.**, **A.P. Sage** and **S. Dozono** (1984). "A model of Multiattribute Decision Making and Trade-off Weight Determination Under Uncertainty", *IEEE Trans. Syst. Man and Cybernet.* **SMC-14**, 223–229.

QÜESTIIÓ, Vol 18, 2 pp. 173-188, 1994

P–INSESGADEZ ASINTÓTICA Y ROBUSTEZ EN LA ESTIMACIÓN LINEAL CON MODELOS DE SUPERPOBLACIONES: UN CRITERIO DE SELECCIÓN

JOSÉ MIGUEL CASAS SÁNCHEZ* y MARTA GUIJARRO GARVI†

Se analiza la estimación lineal de la media poblacional desde el punto de vista de los modelos de superpoblaciones con parámetros desconocidos y correlación no nula. La posible existencia de errores de especificación en el modelo determina el estudio de propiedades tales como la insesgadez asintótica respecto al diseño de muestreo y la robustez débil, deseables para asegurar la robustez de los estimadores frente a este tipo de errores. Se establece, así, un criterio de selección entre estimadores lineales asintóticamente insesgados respecto al diseño y débilmente robustos.

Asymptotic design unbiasedness and robustness of linear estimation under superpopulation models: a selection procedure.

Key words: Modelo de superpoblación; Estimación lineal; P–insesgadez asintótica; Robustez débil.

Clasificación AMS: 62D05.

* José Miguel Casas Sánchez. Universidad de Alcalá de Henares.

† Marta Guijarro Garvi. Universidad de Cantabria.

–Article rebut el setembre de 1993.

–Acceptat el març de 1994.

1. INTRODUCCIÓN

En la estimación de la media poblacional bajo la suposición de un modelo de superpoblación que ligue la variable de interés a una o más variables auxiliares, resulta obligado considerar estimadores robustos frente a fallos en la formulación del modelo de trabajo. Si los parámetros del modelo son desconocidos (hecho, por otro lado, habitual), la estimación de dichos parámetros lleva consigo la dificultad para obtener estimadores de la media poblacional que sean insesgados según el diseño de muestreo. En este sentido, adoptaremos el punto de vista clásico, defendido por Hansen, Madow y Tepping (1983), exigiendo la insesgadez asintótica respecto del diseño de muestreo, o $p-$insesgadez asintótica, de los estimadores, como primera garantía de robustez; este hecho requiere la definición de un marco asintótico.

Sin embargo, aunque la $p-$insesgadez asintótica es necesaria, no es suficiente para que una estrategia de muestreo sea óptima en el sentido de minimizar el error cuadrático medio esperado, criterio que utilizaremos para medir la calidad de un estimador (Godambe, 1955; Särndal, 1980). Por ello, Tam (1988) considera otras propiedades de interés, tales como la insesgadez respecto al modelo y la robustez débil, o robustez frente a un error de especificación en la matriz de covarianzas del modelo.

En este trabajo, generalizamos la definición de estimador débilmente robusto debida a Tam (1988), con objeto de utilizarla en contextos con correlación no nula, considerando, únicamente, estimadores lineales.

El resultado que presentamos sugiere que, dados dos estimadores lineales asintóticamente $p-$insesgados y débilmente robustos, deberíamos elegir aquel con menor sesgo según el modelo porque, asintóticamente, es el de menor error cuadrático medio esperado. La extensión del resultado obtenido por Tam (1988) a un contexto con correlación conlleva, sin embargo, la necesidad de exigir una serie de hipótesis adicionales.

2. DESCRIPCIÓN DEL MODELO

Sea $\{i\}$ una sucesión de elementos, estando, cada uno de ellos, asociado a un número desconocido, y_i, y a un vector conocido, x_i, de dimensión $q \times 1$. Adoptando el marco de trabajo de Isaki y Fuller (1982), definimos una sucesión de poblaciones finitas, $\{U_t\}$, de tamaño N_t con $0 < N_1 < \ldots$, tal que U_1 está formada por las N_1 primeras unidades de $\{i\}$, $U_2 \supset U_1$ contiene los N_2 primeros elementos de $\{i\}$, etc. Así, U_t se expande haciendo tender N_t a infinito, cuando t tiende a infinito.

Para cada población, U_t, consideraremos que $y_t = (y_1, \ldots, y_{N_t})'$ es una realización del vector aleatorio $Y_t = (Y_1, \ldots, Y_{N_t})'$ relacionado con la matriz $X_t = (x_1, \ldots, x_{N_t})'$ a través del modelo de superpoblación ξ dado por

$$\begin{aligned} &E_\xi(Y_t) = X_t\beta \\ &E_\xi[(Y_t - X_t\beta)(Y_t - X_t\beta)'] = \sigma^2 V_t \end{aligned}$$

donde $E_\xi(\cdot)$ denota la esperanza respecto al modelo ξ, β es un vector de dimensión $q \times 1$ desconocido, σ^2 constante conocida y $V_t = (v_{ik})$, matriz simétrica y definida positiva con la siguiente estructura:

$$v_{ik} = \begin{cases} v_i & i = k \\ \rho(v_i v_k)^{1/2} & i \neq k \end{cases}$$

con $v_i > 0$ conocido ($i = 1, \ldots, N_t$) y ρ conocido verificando la condición $-(N_t - 1)^{-1} \leq \rho < 1$. Supondremos que X_t es de rango completo q.

Sea $\{s_t\}$ una sucesión de muestras obtenidas a partir de $\{U_t\}$ mediante una secuencia de diseños de tamaño efectivo fijo $\{n_t\}$, de modo que s_1 está formada por n_1 elementos distintos de U_1, s_2 está formada por n_2 elementos distintos de U_2, etc.; $n_1 < n_2 < \ldots$ y $n_t < N_t$ $\forall t$. El hecho de que también n_t tiende a infinito está implícito en la condición C.3 de la siguiente sección. Destaquemos la no exigencia de que n_t crezca con la misma rapidez que N_t.

Llamaremos π_{it} a la probabilidad de que la $i-$ésima unidad esté incluida en la $t-$ésima muestra, e I_{it} a la variable aleatoria que vale 1 si la unidad $i-$ésima está en la muestra $t-$ésima y 0 en caso contrario.

Sin pérdida de generalidad listaremos primero las unidades de la muestra, realizando, así, las siguientes particiones:

$$\begin{aligned} &Y_t = (Y'_{s_t}, Y'_{r_t})' \\ &X_t = (X'_{s_t}, X'_{r_t})' \\ &\Pi_t = \operatorname{diag}(\pi_{1t}, \ldots, \pi_{N_t t}) = \begin{pmatrix} \Pi_{s_t} & 0 \\ 0 & \Pi_{r_t} \end{pmatrix} \\ &1_t = (1'_{s_t}, 1'_{r_t})' \end{aligned}$$

donde 1_t es el vector de dimensión $N_t \times 1$ y s_t y r_t indican el conjunto de unidades de la población que están y no están en la muestra, respectivamente.

Denotaremos por $\{e_t\}$, una sucesión de estimadores lineales homogéneos, es decir,

$$e_t = L_t' Y_{s_t} = N_t^{-1}(l_{1s_t}, \ldots, l_{n_t s_t}) Y_{s_t}$$

Definición 1

Diremos que una sucesión de estimadores lineales, $\{e_t\}$, construida a partir de la sucesión de poblaciones, es asintóticamente insesgada según un diseño, p, o asintóticamente p−insesgada para $\bar{Y}_t$ si

$$\lim_{t\to\infty} [E_p(e_t) - \bar{Y}_t] = 0$$

donde $E_p(\cdot)$ es la esperanza respecto al diseño de muestreo p e $\bar{Y}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} Y_i$.

Por simplicidad, hablaremos de estimadores asintóticamente insesgados según el diseño o asintóticamente p−insesgados.

Definición 2

Un estimador, e_t, es insesgado respecto al modelo, o ξ−insesgado, si

$$E_\xi(e_t - \bar{Y}_t) = 0 \qquad \forall s.$$

Admitiremos que se cumplen las siguientes condiciones generales, habituales en la literatura del muestreo en superpoblaciones (Robinson y Särndal, 1983; Wright, 1983).

C.1 $\limsup_{t\to\infty} N_t^{-1}\sum_{i=1}^{N_t} x_{ij}^2 < \infty \qquad j = 1, \ldots, q$

C.2 $\limsup_{t\to\infty} E_p(\hat{\beta}_{jt})^2 \qquad j = 1, \ldots, q$

C.3 $\liminf_{t\to\infty} \frac{N_t}{n_t} \min_{1\le i\le N_t} \pi_{it} > 0$

C.4 $\limsup_{t\to\infty} \frac{N_t^2}{n_t} \max_{i\neq k} |\pi_{ikt} - \pi_{it}\pi_{kt}| < \infty$ donde $\pi_{ikt} = p(I_{it} = I_{kt} = 1)$ son las probilidades de inclusión de segundo orden

C.5 $\limsup_{t\to\infty} \frac{1}{N_t}\sum_{i=1}^{N_t} v_i < \infty$

C.6 Dado un diseño de muestreo p, existe una constante k tal que para un t suficientemente grande,

$$n_t \sum_{j=1}^{q} E_\xi(\beta_j - \hat{\beta}_{jt})^2 < k < \infty$$

C.7 $\limsup_{t\to\infty} N_t^{-1} \sum_{i=1}^{N_t} Y_i < \infty$

El estimador de regresión generalizado[1]:

$$e_{RG}(Q) = \frac{1}{N_t} 1'_{s_t} \Pi_{s_t}^{-1} Y_{s_t} + \frac{1}{N_t}(1'X_t - 1'_{s_t}\Pi_{s_t}^{-1} X_{s_t})\hat{\beta}(Q_{s_t})$$

con $\hat{\beta}(Q_{s_t}) = (X'_{s_t} Q_{s_t} X_{s_t})^{-1} X'_{s_t} Q_{s_t} Y_{s_t}$ y Q_{s_t}, matriz simétrica y definida positiva, es asintóticamente $p-$insesgado bajo las condiciones C.1-C.4 (Casas y Guijarro, 1993)[2].

Con objeto de simplificar notaciones, prescindiremos, en ocasiones, del subíndice t.

Definición 3

Un estimador lineal, e, de $\bar{Y}$ es robusto frente a un error en la especificación de la matriz de covarianzas, o débilmente robusto, si, para cada s con $p(s) > 0$, se cumple:

$$Q_s^{-1}\left(L - \frac{1}{N}\Pi_{0s}^{-1} 1_s\right) \in C(X_s)$$

para alguna matriz Q_s, simétrica y definida positiva. $C(X_s)$ denota el espacio generado por las columnas de X_s y Π_{0s}=diag $(\pi_{01}, \ldots, \pi_{0n})$= $n\left(\sum_{i=1}^N v_i^{1/2}\right)^{-1}$ diag$(v_1^{1/2}, \ldots, v_n^{1/2})$.

Tam (1988) demuestra que una estrategia de muestreo, es decir, un par (e, p_0) donde e es un estimador lineal de $\bar{Y}$ y p_0 un diseño de muestreo con probabilidades de inclusión $\pi_{0i} = n\left(\sum_{i=1}^N v^{1/2}\right)^{-1} v_i^{1/2}$ $(\forall i)$, es óptima, en el sentido de minimizar el error cuadrático medio esperado, si el estimador es $\xi-$insesgado y

[1] Elemento de la *clase de estimadores QR* (Wright, 1983):

$$e_{QR} = \frac{1}{N} 1'_s R_s Y_s + \frac{1}{N}(1'X - 1'_s R_s X_s)\hat{\beta}(Q_s)$$

para $R_s = \Pi_s^{-1}$.

[2] De hecho Casas y Guijarro prueban la $p-$insesgadez asintótica del estimador de regresión generalizado bajo las condiciones C.1, C.2 y

$$\liminf_{t\to\infty} \frac{N_t}{n_t} \min_{1\le i\le N_t} \pi_{it} > 0$$

$$\liminf_{t\to\infty} \left| \frac{\pi_{ikt}}{\pi_{it}\pi_{kt}} - 1 \right|$$

menos restrictivas que las condiciones C.3 y C.4.

débilmente robusto. Este hecho justifica, por sí solo, la introducción del concepto de robustez débil.

3. ESTIMADORES LINEALES p–INSESGADOS Y DÉBILMENTE ROBUSTOS

En el siguiente resultado probamos que, dados dos estimadores lineales asintóticamente p–insesgados y débilmente robustos, aquel cuyo ξ–sesgo sea menor tiene, también, menor error cuadrático medio esperado.

Teorema

Sean $L'_m Y_s$ $(m = 1, 2)$ dos estimadores lineales de $\bar{Y}$, asintóticamente p–insesgados y débilmente robustos, verificando las hipótesis siguientes:

H.1 $\limsup_{t\to\infty} \frac{n_t}{N_t} \max_{1\le i\le N_t} E_p(l^2_{m_{is}} I_{it}) < \infty$

H.2 $\limsup_{t\to\infty} \max_{1\le i\le N_t} |E_p(l_{m_{is}} I_{it}) - 1| < \infty$

H.3 $\limsup_{t\to\infty} n_t \max_{i\neq k} |E_p(l_{m_{is}} l_{m_{ks}} I_{it} I_{kt}) - 1| < \infty$

con $L'_m = N_t^{-1}(l_{m_{1s}}, \ldots, l_{m_{ns}})$.[3]

Si el ξ–sesgo de $L'_1 Y_s$ es mayor que el de $L'_2 Y_s$ (ambos en valor absoluto), entonces

$$\lim_{t\to\infty} n_t[E_p E_\xi(L'_1 Y_s - \bar{Y}_t)^2 - E_p E_\xi(L'_2 Y_s - \bar{Y}_t)^2] \ge 0$$

es decir, se verifica asintóticamente:

$$E_p E_\xi(L'_1 Y_s - \bar{Y}_t)^2 \ge E_p E_\xi(L'_2 Y_s - \bar{Y}_t)^2$$

para un diseño de muestreo dado p.

Demostración

El hecho de que los estimadores considerados sean débilmente robustos nos permite la siguientes descomposición:

[3] Deberíamos escribir $L'_{m_t} Y_{s_t} = N_t^{-1}(l_{m_{1s_t}}, \ldots, l_{m_{n_t s_t}}) Y_{s_t}$, estimador asintóticamente p–insesgado de $\bar{Y}_t$; mantendremos, sin embargo, la notación inicial a fin de simplificar las expresiones.

$$L'_m Y_s = e_{RG} + \left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) \hat{\beta}$$

para alguna matriz Q, simétrica y definida positiva. [4]

Así, podemos escribir

$$n_t E_p E_\xi (L'_m Y_s - \bar{Y}_t)^2 = n_t E_p E_\xi \left[e^*_{RG} + \left(L' X_s - \frac{1'_t X_t}{N_t} \right) \beta - \bar{Y}_t \right]^2 +$$
$$n_t E_p E_\xi \left[e_{RG} - e^*_{RG} - \left(L'_m X_s - \frac{1' X}{N_t} \right) (\beta - \hat{\beta}) \right]^2 +$$
$$+2 n_t E_p E_\xi \left[e^*_{RG} + \left(L'_m X_s - \frac{1' X}{N_t} \right) \beta - \bar{Y}_t \right] \cdot$$
$$\cdot \left[e_{RG} - e^*_{RG} - \left(L'_m X_s - \frac{1' X}{N_t} \right) (\beta - \hat{\beta}) \right]$$

donde e^*_{RG} es la variable aleatoria que resulta de sustituir $\hat{\beta}$ en e_{RG} por β, parámetro desconocido.

Acotemos cada uno de los sumandos:

$$e^*_{RG} + \left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) \beta - \bar{Y}_t = e^*_{RG} - \bar{Y}_t + \frac{1}{N_t} \sum_{j=1}^{q} \beta_j \sum_{i=1}^{N_t} (l_{m_{is}} I_{it} - 1) x_{ij}$$

con β_j componente j−ésima del vector β.

Elevando al cuadrado y tomando esperanzas tendremos que

$$n_t E_p E_\xi \left[e^*_{RG} + \left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) \beta - \bar{Y}_t \right]^2$$

puede expresarse como

$$n_t E_p E_\xi (e^*_{RG} - \bar{Y}_t)^2 + \frac{n_t}{N_t^2} E_p \left(\sum_{j=1}^{q} \beta_j c_{m_{jt}} \right)^2$$

con

$$c_{m_{jt}} = \sum_{i=1}^{N_t} (l_{is} I_{it} - 1) x_{ij}$$

[4]Para dar mayor fluidez a las notaciones escribiremos e_{RG} y $\hat{\beta}$ en vez de $e_{RG}(Q)$ y $\hat{\beta}(Q_s)$.

Bajo las condiciones C.3-C.5:

$$n_t E_p E_\xi (e^*_{RG} - \bar{Y}_t)^2 < \infty$$

cuando $t \to \infty$ (Casas y Guijarro, 1993).

Además,

$$\frac{n_t}{N_t^2} E_p \left(\sum_{j=1}^{q} \beta_j c_{m_{jt}} \right)^2 \leq \frac{n_t}{N_t^2} \sum_{j=1}^{q} \beta_j^2 \sum_{j=1}^{q} E_p(c^2_{m_{jt}})$$

Pero,

$$\frac{n_t}{N_t^2} E_p(c^2_{m_{jt}}) = \frac{n_t}{N_t^2} E_p \left[\sum_{i=1}^{N_t} (l_{m_{is}} I_{it} - 1) x_{ij} \right]^2 =$$

$$= \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} E_p(l_{m_{is}} I_{it} - 1)^2 x_{ij}^2 + \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} \sum_{i \neq k} E_p(l_{m_{is}} I_{it} - 1)(l_{m_{ks}} I_{kt} - 1) x_{ij} x_{kj}$$

Desarrollando cada sumando:

$$\frac{n_t}{N_t^2} \sum_{i=1}^{N_t} E_p(l_{m_{is}} I_{it} - 1)^2 x_{ij}^2 \leq \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} E_p(l^2_{m_{is}} I_{it}) x_{ij}^2 +$$

$$+2 \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} [1 - E_p(l_{m_{is}} I_{it})] x_{ij}^2 \leq$$

$$\leq \frac{n_t}{N_t} \max_{1 \leq i \leq N_t} E_p(l^2_{m_{is}} I_{it}) \left(\frac{1}{N_t} \sum_{i=1}^{N_t} x_{ij}^2 \right) +$$

$$+2 \frac{n_t}{N_t} \max_{1 \leq i \leq N_t} |E_p(l_{m_{is}} I_{it}) - 1| \left(\frac{1}{N_t} \sum_{i=1}^{N_t} x_{ij}^2 \right) < \infty$$

cuando $t \to \infty$ por las condiciones H.1, H.2 y C.1.

Con respecto al segundo sumando:

$$\frac{n_t}{N_t^2} \sum_{i=1}^{N_t} \sum_{i \neq k} E_p(l_{m_{is}} I_{it} - 1)(l_{m_{ks}} I_{kt} - 1) x_{ij} x_{kj} =$$

$$= \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} \sum_{i \neq k} [E_p(l_{m_{is}} l_{m_{ks}} I_{it} I_{kt}) - E_p(l_{m_{is}} I_{it}) - E_p(l_{m_{ks}} I_{kt}) + 1] x_{ij} x_{kj} =$$

$$= \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} \sum_{i \neq k} [E_p(l_{m_{is}} l_{m_{ks}} I_{it} I_{kt}) - 1] x_{ij} x_{kj} +$$

$$+ \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} \sum_{i \neq k} [1 - E_p(l_{m_{is}} I_{it})] x_{ij} x_{kj} +$$

$$+ \frac{n_t}{N_t^2} \sum_{i=1}^{N_t} \sum_{i \neq k} [1 - E_p(l_{m_{ks}} I_{kt})] x_{ij} x_{kj} \leq$$

$$\leq n_t \max_{i \neq k} |E_p(l_{m_{is}} l_{m_{ks}} I_{it} I_{kt}) - 1| \frac{1}{N_t^2} \left(\sum_{i=1}^{N_t} |x_{ij}| \right)^2 +$$

$$+ 2n_t \max_{1 \leq i \leq N_t} |1 - E_p(l_{m_{is}} I_{it})| \frac{1}{N_t^2} \left(\sum_{i=1}^{N_t} |x_{ij}| \right)^2 \leq$$

$$\leq n_t \max_{i \neq k} |E_p(l_{m_{is}} l_{m_{ks}} I_{it} I_{kt}) - 1| \frac{1}{N_t} \left(\sum_{i=1}^{N_t} x_{ij}^2 \right) +$$

$$+ 2n_t \max_{1 \leq i \leq N_t} |1 - E_p(l_{m_{is}} I_{it})| \frac{1}{N_t} \left(\sum_{i=1}^{N_t} x_{ij}^2 \right) < \infty$$

cuando t tiende a infinito, aplicando H.2 y H.3.

Acotemos ahora la expresión,

$$n_t E_p E_\xi \left[e_{RG} - e^*_{RG} - \left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) (\beta - \hat{\beta}) \right]^2 =$$
$$= n_t E_p E_\xi (e_{RG} - e^*_{RG})^2 +$$
$$n_t E_p E_\xi \left[\left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) (\beta - \hat{\beta}) \right]^2 +$$
$$+ 2n_t E_p E_\xi (e_{RG} - e^*_{RG}) \left[\left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) (\beta - \hat{\beta}) \right]$$

Teniendo en cuenta que

$$\left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) (\beta - \hat{\beta}) = \frac{1}{N_t} \sum_{j=1}^{q} (\beta_j - \hat{\beta}_j) \sum_{i=1}^{N_t} (l_{m_{is}} I_{it} - 1) x_{ij}$$

Elevando al cuadrado y tomando esperanzas respecto al modelo resultará:

$$\frac{n_t}{N_t^2}E_\xi\left[\sum_{j=1}^{q}(\beta_j-\hat{\beta}_j)c_{m_{jt}}\right]^2 \leq \frac{n_t}{N_t^2}\sum_{j=1}^{q}E_\xi(\beta_j-\hat{\beta}_j)^2\sum_{j=1}^{q}c_{m_{jt}}^2$$

Por la condición C.6, dado un diseño p, existe una constante K tal que, a partir de un t suficientemente grande:

$$\begin{aligned}
\frac{K}{N_t^2}\sum_{j=1}^{q}E_p(c_{m_{jt}}^2) &= \frac{K}{N_t^2}\sum_{j=1}^{q}E_p\left[\sum_{i=1}^{N_t}(l_{m_{is}}I_{it}-1)^2x_{ij}^2\right]+\\
&+ \frac{K}{N_t^2}\sum_{j=1}^{q}E_p\left[\sum_{i=1}^{N_t}\sum_{i\neq k}(l_{m_{is}}I_{it}-1)(l_{m_{ks}}I_{kt}-1)x_{ij}x_{kj}\right]\leq\\
&\leq \sum_{j=1}^{q}\left[\frac{K}{n_t}\frac{n_t}{N_t}\max_{1\leq i\leq N_t}E_p(l_{m_{is}}^2I_{it})\left(\frac{1}{N_t}\sum_{i=1}^{N_t}x_{ij}^2\right)\right]+\\
&+ \sum_{j=1}^{q}\left[2\frac{K}{N_t}\max_{1\leq i\leq N_t}|E_p(l_{m_{is}}I_{it})-1|\left(\frac{1}{N_t}\sum_{i=1}^{N_t}x_{ij}^2\right)\right]+\\
&+ k\sum_{j=1}^{q}\left[\max_{i\neq k}|E_p(l_{m_{is}}l_{m_{ks}}I_{it}I_{kt})-1|\left(\frac{1}{N_t}\sum_{i=1}^{N_t}x_{ij}^2\right)\right]+\\
&+ k2\frac{1}{N_t}\sum_{j=1}^{q}\left(\max_{1\leq i\leq N_t}|E_p(l_{m_{is}}I_{it})-1|\sum_{i=1}^{N_t}x_{ij}^2\right)
\end{aligned}$$

Expresión que tiende a 0 cuando $t\to\infty$ por las condiciones H.1 y H.3.

Además, las condiciones C.1, C.3, C.4 y C.6 permiten asegurar que

$$\lim_{t\to\infty}n_tE_pE_\xi(e_{RG}-e_{RG}^*)^2=0$$

Por último, por la desigualdad de Schwartz se demuestra que

$$2n_tE_pE_\xi(e_{RG}-e_{RG}^*)\left[\left(L_m'X_s-\frac{1_t'X_t}{N_t}\right)(\beta-\hat{\beta})\right]$$

y

$$\begin{aligned}
2n_tE_pE_\xi&\left[e_{RG}^*+\left(L_m'X_s-\frac{1_t'X_t}{N_t}\right)\beta-\bar{Y}_t\right]\left[e_{RG}-e_{RG}^*-\right.\\
-&\left.\left(L_m'X_s-\frac{1_t'X_t}{N_t}\right)(\beta-\hat{\beta})\right]
\end{aligned}$$

convergen a 0 cuando $t\to\infty$.

De todo lo visto se deduce que

$$n_t E_p E_\xi (L'_m Y_s - \bar{Y}_t)^2 = n_t E_p E_\xi (e^*_{RG} - \bar{Y}_t)^2 + n_t E_p \left[\left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) \beta \right]^2 + O_t$$

con $\lim_{t\to\infty} O_t = 0$, es decir, asintóticamente,

$$E_p E_\xi (L'_m Y_s - \bar{Y}_t)^2 = E_p E_\xi (e^*_{RG} - \bar{Y}_t)^2 + E_p \left[\left(L'_m X_s - \frac{1'_t X_t}{N_t} \right) \beta \right]^2$$

Por tanto, si

$$\left| \left(L'_1 X_s - \frac{1'_t X_t}{N_t} \right) \beta \right| \geq \left| \left(L'_2 X_s - \frac{1'_t X_t}{N_t} \right) \beta \right|$$

esto es, si el $\xi-$sesgo de $L'_1 Y_s$ es mayor que el de $L'_2 Y_s$ (ambos en valor absoluto), resulta

$$E_p \left[\left(L'_1 X_s - \frac{1'_t X_t}{N_t} \right) \beta \right]^2 \geq E_p \left[\left(L'_2 X_s - \frac{1'_t X_t}{N_t} \right) \beta \right]^2$$

ya que el valor medio esperado respecto al diseño del cuadrado del $\xi-$sesgo es una función creciente. Se concluye, así, la prueba del teorema.

4. APLICACIÓN A UN MODELO ESPECÍFICO

Sea el modelo

$$\begin{aligned} &E_\xi(Y_i) = \beta_0 + \beta_1 x_i \\ &E_\xi[(Y_i - \beta_0 - \beta_1 x_i)^2] = \sigma^2 x_i \\ &E_\xi[(Y_i - \beta_0 - \beta_1 x_i)(Y_k - \beta_0 - \beta_1 x_k)] = \sigma^2 \rho (x_i x_k)^{1/2} \end{aligned}$$

con $\beta_0 \neq 0$, $\beta_1 \neq 0$ y σ constantes desconocidas y $x_i > 0$ para $i = 1, \ldots N$.

Consideremos las estrategias de muestreo $(L'_1 Y_s, p_0)$ y $(L'_2 Y_s, p_0)$, donde

$$L'_1 Y_s = \bar{x} \frac{\sum_{i\in s} Y_i/\pi_i}{\sum_{i\in s} x_i/\pi_i}$$

es el estimador de razón generalizado (Brewer, 1963),

$$L'_2 Y_s = \frac{1}{N} \sum_{i\in s} \frac{Y_i}{\pi_i}$$

es el estimador de Horvitz-Thompson y, por último, p_0 es un diseño de muestreo con probabilidades de inclusión

$$\pi_{i0} = \frac{n x_i^{1/2}}{\sum_{i=1}^{N} x_i^{1/2}}$$

Se demuestra de modo sencillo que $L_1' Y_s$ es asintóticamente insesgado según el diseño de muestreo p_0 (Särndal, 1980). El estimador de Horvitz-Thompson, como ya es sabido, es insesgado para cualquier diseño de muestreo.

Las condiciones C.3 y C.4, junto con la aplicación del teorema de Slutsky (Särndal, 1980), mediante el cual sustituimos en el límite una función de las medias muestrales por la misma función de sus valores esperados, nos permiten demostrar que $L_1' Y_s$ y $L_2' Y_s$ cumplen las condiciones H.1-H.3.

El estimador $L_1' Y_s$ es débilmente robusto, sin más que considerar la matriz Q_s=diag $(x_1^{1.5}, \ldots, x_n^{1.5})$ (Tam, 1988). Además, como

$$L_2' Y_s = \frac{1}{N} 1_s' \Pi_s^{-1} Y_s$$

basta tomar el estimador de Horvitz-Thompson con el diseño de muestreo óptimo p_0 para que, trivialmente, se cumpla la definición de estimador débilmente robusto.

Estamos, pues, en condiciones de aplicar el resultado que nos sugiere la elección del estimador con el menor ξ−sesgo. Sencillas operaciones nos conducen a

$$E_\xi(L_1' Y_s - \bar{Y}) = \left[\bar{x} \left(\sum_s x_i^{-1/2} \right) \left(\sum_s x_i^{1/2} \right)^{-1} - 1 \right] \beta_0$$

y

$$\begin{aligned} E_\xi(L_2' Y_s - \bar{Y}) &= \left[\frac{\sum_{i=1}^{N} x_i^{1/2}}{Nn} \left(\sum_s x_i^{-1/2} \right) - 1 \right] \beta_0 + \\ &+ \left[\frac{\sum_{i=1}^{N} x_i^{1/2}}{Nn} \left(\sum_s x_i^{1/2} \right) - \bar{x} \right] \beta_1 \end{aligned}$$

con $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$.

El hecho de que el ξ−sesgo del estimador de Horvitz-Thompson, $L_2' Y_s$, dependa, no sólo de β_0, sino también de $\beta_1 \neq 0$, nos lleva a descartarlo, prefiriendo, en este caso, el estimador de razón generalizado, $L_1' Y_s$.

5. BIBLIOGRAFÍA

[1] **Azorín, F.** y **Sánchez-Crespo, J.L.** (1986). *Métodos y aplicaciones del muestreo.* Madrid: Alianza.

[2] **Brewer, K.R.W.** (1963). "Ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process". *Australian Journal of Statistics,* **5**, 93–105.

[3] **Brewer, K.R.W.** (1979). "A class of robust sampling designs for large-scale surveys". *Journal of American Statistical Association,* **74**, 911–915.

[4] **Cassel, C., Särndal, C.** y **Wretman, J.H.** (1977). *Foundations of Inference in Survey Sampling.* New York: John Wiley.

[5] **Casas, J.M.** y **Guijarro, M.** (1993). "El estimador de regresión generalizado en el modelo de superpoblación: $p-$insesgadez asintótica y robustez". *Estadística Española,* **35**, 425–437.

[6] **Fuller, W.A.** e **Isaki, C.T.** (1982). "Survey design under the regression superpopulation model". *Journal of American Statistical Association,* **77**, 89–96.

[7] **Godambe, V.P.** (1955). "A unified theory of sampling from finite populations". *Journal of the Royal Statistical Society,* Ser. B, **17**, 369–378.

[8] **Godambe, V.P.** (1982). "Estimation in survey sampling: robustness and optimality". *Journal of American Statistical Association,* **77**, 393–406.

[9] **Godambe, V.P.** y **Thompson, M. E.** (1977). "Robust near optimal estimation in survey practice". *Bulletin of the International Statistical Institute,* **47**, 129–146.

[10] **Hansen, M.H., Madow, W.G.** y **Tepping, B. J.** (1983). "An evaluation of model-dependent and probability-sampling inferences in sample surveys". *Journal of American Statistical Association,* **78**, 776–807.

[11] **Herson, J.** y **Royall, R.M.** (1973). "Robust estimation in finite populations". *Journal of American Statistical Association,* **68**, 880–893.

[12] **Robinson, P.M.** y **Särndal, C.E.** (1983). "Asymptotic properties of the generalized regression estimator in probability sampling". *Sankyā.* Ser. B, **45**, 240–248.

[13] **Särndal, C.E.** (1980a). "On $\pi-$inverse weighting versus best linear unbiased weighting in probability sampling". *Biometrika,* **67**, 639–650.

[14] **Tam, S.M.** (1988b). "Some results on robust estimation in finite population sampling". *Journal of American Statistical Association,* **83**, 242–248.

[15] **Wright, R.L.** (1983). "Finite population sampling with multivariate auxiliar information". *Journal of American Statistical Association,* **78**, 879–884.

ENGLISH SUMMARY:

ASYMPTOTIC DESIGN UNBIASEDNESS AND ROBUSTNESS OF LINEAR ESTIMATION UNDER SUPERPOPULATION MODELS: A SELECTION PROCEDURE

José Miguel Casas Sánchez and Marta Guijarro Garvi

1. INTRODUCTION

Estimation of means under superpopulation models leads to the necessity of looking for robust estimators when the model is misspecified.

This work assumes a superpopulation model with correlated residuals. In such context, a rule for choosing among weakly robust and asymptotically design unbiased linear estimators, is suggested.

2. MODEL DESCRIPTION

Following the asymptotic framework of Isaki and Fuller, a generalized definition of weakly robust estimator is presented.

This definition assumes $y_t = (y_1, \ldots, y_{N_t})'$ to be the realized outcome of a random vector $Y_t(Y_1, \ldots, Y_{N_t})'$ related to the matrix $X_t = (x_1, \ldots, x_{N_t})'$ through the superpopulation model ξ

$$E_\xi(Y_t) = X_t\beta$$
$$E_\xi[(Y_t - X_t\beta)(Y_t - X_t\beta)'] = \sigma^2 V_t$$

where $V_t = (v_{ik})$ is a positive definite matrix with non cero correlation coefficient.

Weakly robustness, asymptotic design unbiasedness, model unbiasedness and C.1-C.7 conditions, play a central role in this paper.

3. ASYMPTOTICALLY DESIGN UNBIASED AND WEAKLY ROBUST LINEAR ESTIMATORS

Under H.1-H.3 and given two veackly robust and asymptotically design unbiased linear estimators of $\bar{Y}, L'_m Y_s (m = 1, 2)$:

$$E_p E_\xi (L'_1 Y_s - \bar{Y}_t)^2 \geq E_p E_\xi (L'_2 Y_s - \bar{Y}_t)^2$$

asymptotically and for all sampling designs p, if the model bias of $L'_1 Y_s$ is bigger than that of $L'_2 Y_s$ (in absolute terms).

4. APPLICATION

Under the model

$$E_\xi(Y_i) = \beta_0 + \beta_1 x_i$$

$$E_\xi[(Y_i - \beta_0 - \beta_1 x_i)^2] = \sigma^2 x_i$$

$$E_\xi[(Y_i - \beta_0 - \beta_1 x_i)(Y_k - \beta_0 - \beta_1 x_k)] = \sigma^2 \rho (x_i x_k)^{1/2}$$

it can be easily proved that

$$E_\xi(L'_1 Y_s - \bar{Y}) = \left[\bar{x} \left(\sum_s x_i^{-1/2} \right) \left(\sum_s x_i^{1/2} \right)^{-1} - 1 \right] \beta_0$$

and

$$E_\xi(L_2'Y_s - \bar{Y}) = \left[\frac{\sum_{i=1}^{N} x_i^{1/2}}{Nn}\left(\sum_s x_i^{-1/2}\right) - 1\right]\beta_0 + \\ + \left[\frac{\sum_{i=1}^{N} x_i^{1/2}}{Nn}\left(\sum_s x_i^{1/2}\right) - \bar{x}\right]\beta_1$$

with $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and

$$\pi_{i0} = \frac{n x_i^{1/2}}{\sum_{i=1}^{N} x_i^{1/2}}$$

where

$$L_1'Y_s = \bar{x}\frac{\sum_{i\in s} Y_i/\pi_i}{\sum_{i\in s} x_i/\pi_i}$$

is the generalized ratio estimator and

$$L_2'Y_s = \frac{1}{N}\sum_{i\in s}\frac{Y_i}{\pi_i}$$

the Horvitz-Thompson estimator.

Since $E_\xi(L_2'Y_s - \bar{Y})$ depends not only on B_0, but in $B_1 \neq 0$, application of the theorem shows that $L_1'Y_s$ is better than $L_2'Y_s$.

QÜESTIIÓ, Vol 18, 2 pp. 189-208, 1994

CONDICIONES NECESARIAS DE OPTIMALIDAD EN PROGRAMACIÓN SEMI-INFINITA LINEAL: CUALIFICACIONES DE RESTRICCIONES Y PROPIEDADES DEL CONJUNTO POSIBLE

TERESA LEÓN y ENRIQUETA VERCHER

Universitat de València

En este trabajo se establece una caracterización de las soluciones óptimas para el problema continuo de Programación Semi-Infinita Lineal, donde el conjunto de índices es un compacto de $\mathbb{R}^p$. Para la demostración de la condición necesaria de optimalidad se ha utilizado una extensión de la cualificación de restricciones de Mangasarian-Fromovitz. Hemos probado que dicha cualificación es imprescindible para asegurar que no hay desigualdades inestables en el conjunto posible y para que existan puntos extremos no degenerados. Se estudian asimismo otras cualificaciones y su relación con aquélla. Incluimos numerosos ejemplos que clarifican esas relaciones.

Necessary optimality conditions in Semi-Infinite Linear Programming: constraint qualifications and properties of the feasible set.

Key words: Semi-Infinite Programming, Constraint qualifications, Structure of the feasible set, Optimality conditions.

Teresa León y Enriqueta Vercher. Departament d'Estadística i Investigació Operativa. Facultat de Matemàtiques. Universitat de València.

–Article rebut el juliol de 1992.

–Acceptat el desembre de 1993.

1. INTRODUCCIÓN Y FORMULACIÓN DEL PROBLEMA

Consideremos el problema continuo de Programación Semi-Infinita Lineal:

$$
\begin{array}{ll}
\text{(PSI1)} & \text{Min } c^T x \\
& \text{s.a. } a^T(s)x \geq b(s) \quad s \in \mathcal{S}
\end{array}
$$

donde el conjunto $\mathcal{S}$ es un compacto de $\mathbb{R}^p$ y las componentes $a_1, \ldots, a_n$ de la función a y la función b son continuas en $\mathcal{S}$. F denota el conjunto de soluciones posibles de (PSI1) y $v(P)$ su valor óptimo. Una de las aplicaciones clásicas del problema (PSI1) se da en la Teoría de la Aproximación, aunque muchos otros problemas pueden ser formulados también de ese modo.

En la Sección 3 hemos establecido una condición necesaria de optimalidad de tipo Kuhn-Tucker para el problema (PSI1), con una condición de regularidad que es una extensión de la cualificación de restricciones de Mangasarian-Fromovitz (CRMF). Dicha condición es también esencial en el estudio de algunas características del conjunto posible.

En la Sección 2 se ha comprobado que, al igual que ocurre en PL finita, los candidatos a óptimo son puntos posibles con alguna restricción activa. Además, para un conjunto posible no necesariamente compacto, hemos estudiado cómo afecta a su estabilidad el hecho de que el vector nulo sea un elemento del conjunto de los gradientes activos o de su envoltura convexa. La estabilidad de todas las desigualdades que definen el conjunto posible es también una propiedad de los problemas que satisfacen la cualificación de restricciones de Mangasarian-Fromovitz.

En la Sección 4 hemos comparado la CRMF con algunas de las cualificaciones que han sido establecidas en PSI lineal y convexa. Hemos demostrado que para una amplia clase de problemas de PSI lineal no es necesario verificar ninguna cualificación de restricciones —por ejemplo, aquellos que tengan al menos un punto extremo no degenerado— ya que satisfacen trivialmente la condición de Mangasarian-Fromovitz. Por otra parte, cuando en un problema no se satisface esa condición todos los puntos extremos son degenerados, lo que puede complicar considerablemente la resolución numérica del mismo. Hemos estudiado también otras cualificaciones más débiles que permiten caracterizar las soluciones óptimas aplicando las condiciones de Kuhn-Tucker para el problema de PSI lineal.

Podemos escribir el problema (PSI1) en forma estándar introduciendo la función de holgura $z(s)$:

$$
\begin{array}{llll}
\text{(PSI2)} & \text{Min } c^T x & & \\
& \text{s.a. } a^T(s)x \; - & z(s) = b(s) & s \in \mathcal{S} \\
& x \in \mathbb{R}^n & z(s) \geq 0 & s \in \mathcal{S}
\end{array}
$$

Nótese que z es una función continua sobre $\mathcal{S}$, i.e. $z \in \mathbb{C}[\mathcal{S}]$. Supondremos a lo largo de todo el trabajo que $F \neq \emptyset$. Dado un punto $x^* \in F$ y su función de holgura $z^*(s)$, el par $(x^*; z^*)$ es una solución posible del problema (PSI2). Denotaremos por $\text{constr}(x^*) = \{s \in \mathcal{S}: a^T(s)x^* = b(s)\} = \{s \in \mathcal{S}: z^*(s) = 0\}$ y por $A(x^*) = \{a(s): s \in \text{constr}(x^*)\}$ los conjuntos de restricciones y gradientes activos en x^*, respectivamente.

En Nash (85) se dan las siguientes definiciones para las soluciones posibles del problema (PSI2). Sea $(x; z)$ una *solución posible*, se dice que es *básica* si se satisface que $B((x; z)) \cap N(A) = \{\theta\}$, siendo $N(A) = \{(d; h) \in \mathbb{R}^n \times \mathbb{C}[\mathcal{S}]: h(s) = a^T(s)d$ para todo $s \in \mathcal{S}\}$, $B((x; z)) = \{(d; h) \in \mathbb{R}^n \times \mathbb{C}[\mathcal{S}]$: existe $\lambda > 0$ tal que $(x+\lambda d; z+\lambda h) \in X_+$ y $(x-\lambda d; z-\lambda h) \in X_+\}$, $X_+ = \{(y; f) \in \mathbb{R}^n \times \mathbb{C}[\mathcal{S}]: f(s) \geq 0$ para todo $s \in \mathcal{S}\}$ y donde θ es el elemento nulo de $\mathbb{R}^n \times \mathbb{C}[\mathcal{S}]$. Además, una *solución posible básica* $(x; z)$ es *no degenerada* si $B((x; z)) \oplus N(A) = \mathbb{R}^n \times \mathbb{C}[\mathcal{S}]$.

Un problema dual para el problema (PSI1) es (ver p.e. Charnes, Cooper y Kortanek (63)):

$$\text{(PSI1}^*\text{)} \quad \begin{array}{l} \text{Max } \sum_{s \in \mathcal{S}} \lambda(s) b(s) \\ \text{s.a. } \sum_{s \in \mathcal{S}} \lambda(s) a(s) = c \\ \lambda \in \mathbb{R}_+^{(S)} \end{array}$$

donde $\mathbb{R}_+^{(S)}$ es el cono convexo de las sucesiones finitas generalizadas no negativas: $\mathbb{R}_+^{(S)} = \{\alpha: \mathcal{S} \longrightarrow \mathbb{R}_+ / \alpha(s) = 0$ para todo s excepto para un número finito$\}$.

2. ESTABILIDAD EN EL CONJUNTO POSIBLE

Veremos que solo son candidatos a óptimo aquellos puntos posibles para (PSI2) que tienen alguna restricción activa y estudiaremos las características del conjunto posible cuando, para alguna solución $x \in F$, el vector nulo es un gradiente activo o un elemento de la envoltura convexa de $A(x)$, que denotaremos por $\text{co}(A(x))$.

Lema 2.1

Sea $(x; z)$ un punto posible para (PSI2). Si $\text{constr}(x) = \emptyset$, entonces $x \in \text{int}(F)$.

Demostración

Veamos que existe $\lambda > 0$ tal que la bola abierta $B(x, \lambda) \subseteq F$. Por hipótesis constr$(x) = \emptyset$, luego $z(s) > 0$ para todo $s \in \mathcal{S}$. Por compacidad de $\mathcal{S}$, sabemos que existe $\epsilon > 0$ tal que la función continua $z(s) > \epsilon$ en $\mathcal{S}$. Sea $d \in \mathbb{R}^n$, distinguimos dos situaciones:

(1) Si $a^T(s)d \geq 0 \; \forall s \in \mathcal{S}$, entonces $\overline{x} = x + \lambda d \in F$ para cualquier $\lambda > 0$.

(2) Si para d existe algún s tal que $a^T(s)d < 0$, definimos $\varphi(d) = \max\{-a^T(s)d : s \in \mathcal{S}\}$, $\varphi(d) > 0$. Sea $\varphi = \sup \varphi(d) > 0$. Sea $\gamma(d) = \inf\{-z(s)/a^T(s)d$: $s \in \mathcal{S}, a^T(s)d < 0\} \geq 0$, veamos que $\gamma(d) > 0$. Como para cualquier $s \in$ constr(x) tal que $a^T(s)d < 0$ se tiene que $-(z(s)/a^T(s)d) > -\epsilon/a^T(s)d \geq \epsilon/\varphi(d) \geq \epsilon/\varphi$, entonces $\gamma(d) \geq \epsilon/\varphi > 0$. Así, pues, definiendo $\lambda = \epsilon/2\varphi$, por construcción $x + \lambda d \in F$.

Entonces, $B(x, \lambda) \subseteq F$.

■

Nota. De hecho, sí existe una solución posible x de (PSI1) tal que constr$(x) = \emptyset$, en Goberna y López (88) se prueba que $\dim F = n$ y que int $F = \{x \in \mathbb{R}^n : a^T(s)x > b(s)$ para todo $s \in \mathcal{S}\}$. Además, el hecho de que int $F \neq \emptyset$ permite asegurar que el interior algebraico cor $F =$ int F (véase, p.e. Holmes (75)).

El recíproco del Lema 2.1 no es cierto, como se pone de manifiesto en el siguiente ejemplo:

$$\begin{array}{l} \text{Min } x_1 + x_2 \\ \text{s.a. } sx_1 + s^2x_2 \geq 0 \quad s \in [0, 1] \end{array}$$

Es fácil comprobar que: $F = \{x \in \mathbb{R}^2 : x_1 \geq 0,\ x_1 + x_2 \geq 0\}$. Sea $\overline{x} = (1,1)' \in$ int(F), tenemos que constr$(\overline{x}) = \{0\} \neq \emptyset$ siendo $a(0) = 0_2$.

Ahora bien, esta situación es "única", es decir, si $x \in$ int(F) y constr$(x) \neq \emptyset$, entonces necesariamente $A(x) = \{0_n\}$. En otro caso, tomando $d = a(s)$ para $s \in$ constr(x) y para cualquier valor positivo de λ se tendría que $x + \lambda d \notin F$, lo que contradice que $x \in$ int(F).

Corolario 2.1.1

Sea $x \in F$ tal que constr$(x) = \emptyset$, entonces x no es un mínimo.

Dado que los puntos que no tienen restricciones activas no pueden ser soluciones óptimas del (PSI1), en lo que sigue consideraremos soluciones posibles cuyo $\text{constr}(x) \neq \emptyset$.

Ya hemos visto que 0_n es el único gradiente activo posible para los puntos de $\text{int}(F)$. En el ejemplo anterior puede, además, comprobarse que $0_n \in A(x^*)$ para todas las soluciones óptimas x^*, que tenían otros gradientes activos. También podemos encontrar problemas cuyas soluciones óptimas tengan al vector nulo como único gradiente activo:

Ejemplo 2.1 (pp. 31 en GLASHOFF y GUSTAFSON (83))

$$(P_1) \quad \begin{array}{ll} \text{Min } x_1 & \\ \text{s.a. } sx_1 + s^2 x_2 \geq s^2 & s \in [0,1] \end{array}$$

Sea $\overline{x} = (0,2)' \in F$, $z(s) = s^2$ $s \in [0,1]$, $\text{constr}(\overline{x}) = \{0\}$ y se tiene que $A(\overline{x}) = \{0_2\}$. Es fácil ver que $v(P_1) = 0$, luego $\overline{x}$ es óptimo para (P_1).

De hecho, si un punto posible tiene al vector nulo como uno de sus gradientes activos, entonces todas las demás soluciones lo tendrán como gradiente activo.

Lema 2.2

Si existe un $\overline{s} \in \mathcal{S}$ tal que $a(\overline{s}) = 0_n$, entonces o $\overline{s} \in \text{constr}(x)$, para todo $x \in F$, o $\overline{s} \notin \text{constr}(x)$, para ningún $x \in F$.

Demostración

Sea x una solución posible cualquiera, su función de holgura será $z(s) = a^T(s)x - b(s) \geq 0$, para todo $s \in \mathcal{S}$. En particular se tiene que $z(\overline{s}) = -b(\overline{s}) \geq 0$. Así que:

(i) $b(\overline{s}) = 0 = z(\overline{s})$, y $\overline{s} \in \text{constr}(x)$, o

(ii) $b(\overline{s}) < 0$ y $z(\overline{s}) > 0$, con lo que $\overline{s} \notin \text{constr}(x)$.

■

Nota. Según Eckhardt (75) la *desigualdad $a^T(s^*)x \geq b(s^*)$ es inestable en F* si para todo $x \in F$ se tiene que $a^T(s^*)x = b(s^*)$. Es decir, si $s^* \in \text{constr}(x)$, para todo $x \in F$.

Cuando para un cierto $x_0 \in F$ se satisface la condición $0_n \in A(x_0)$, tenemos que existe $s_0 \in \text{constr}(x_0)$ tal que simultáneamente $a(s_0) = 0_n$ y $z(s_0) =$

$a^T(s_o)x_o - b(s_o) = 0$, luego también $b(s_o) = 0$. Del lema anterior se sigue que la desigualdad $a^T(s_o)x \geq b(s_o)$ es inestable en F. El recíproco no es cierto (véase Ejemplo 2.3).

En el Ejemplo 2.1 la desigualdad inestable está asociada a $s_o = 0$, y es una desigualdad trivial, pero esa restricción no puede suprimirse ya que perderíamos la compacidad de $\mathcal{S}$. Podríamos plantearnos redefinir el conjunto posible, pero en general eso no es sencillo ni aconsejable.

Ejemplo 2.2

(P_2) Min $x_1 + (1/2)x_2 + (1/3)x_3$

s.a. $(s_2^2+s_1)x_1+(s_1s_2-s_2^2)x_2+(s_1s_2+s_2^2+s_2)x_3 \geq 2s_1s_2-s_1^2 \quad s \in [0,1]\times[0,1]$

Sea $x = (1,0,0)'$ una solución posible cuyo $z(s_1,s_2) = s_1+(s_2-s_1)^2$, constr$(x) = \{0_2\}$ y $A(x) = \{0_3\}$. De nuevo tenemos una desigualdad que corresponde a $s_o = 0_2$, que no podemos eliminar. En este caso, la discusión sobre la optimalidad de las soluciones puede ser bastante más difícil, como se verá en las siguientes secciones.

Otra situación interesante es la de que $0_n \in$ co$(\mathcal{A}(x))$ —veremos más adelante que la negación de esta condición es una cualificación de restricciones para el (PSI1)—. De nuevo, si para una solución posible x se tiene que $0_n \in$ co$(A(x))$, entonces todos los puntos de F verifican esta condición, que es equivalente a la existencia de desigualdades inestables en el conjunto posible. Para demostrar esos resultados haremos uso del lema siguiente.

Lema 2.3

Sea $(x;z)$ un punto posible para el (PSI2) cuyo constr$(x) \neq \emptyset$, entonces co$(A(x))$ es un conjunto compacto.

Demostración

El conjunto de índices activos constr(x) es cerrado, pues es la antiimagen mediante una función continua de un cerrado: constr$(x) = \{s \in \mathcal{S}: z(s) = 0\} = z^{-1}(\{0\})$. Además, constr$(x)$ es un subconjunto de $\mathcal{S}$, compacto, luego constr(x) es compacto.

Dado que la función a es continua, el conjunto de los gradientes activos $A(x) = a(\{\text{constr}(x)\})$ será un compacto y, en consecuencia, $\text{co}\{A(x)\}$ es un compacto.

■

Teorema 2.4

Si existe una solución posible x^* de (PSI1) tal que $0_n \in \text{co}(A(x^*))$, entonces cualquier otra solución posible satisface que $0_n \in \text{co}(A(x))$.

Demostración

Por reducción al absurdo. Supongamos que existe un $x \in F$ tal que $0_n \notin \text{co}(A(x))$, dicha condición es equivalente a que exista $d \in \mathbb{R}^n$ tal que $a(s)^T d > 0$ para todo $s \in \text{constr}(x)$, pues $\text{co}(A(x))$ es cerrado (Lema 3.7 en López y Vercher (83)).

Definimos $x^1 = x + \mu d$, su función de holgura es $z^1(s) = z(s) + \mu a^T(s)d$. Para $s \in \text{constr}(x)$ tenemos que $z^1(s) > 0$ para todo $\mu > 0$. Por continuidad, $a(\cdot)^T d$ está acotado en $\mathcal{S}$, y como $z(s) > 0$ para $s \notin \text{constr}(x)$, tomando $\mu > 0$ suficientemente pequeño, llegamos a que $z^1(s) > 0$.

Sea $d^1 = x^1 - x^*$, entonces $a^T(s)d^1 = a^T(s)x^1 - b(s) = z^1(s) > 0$ en $\mathcal{S}$. En particular, hemos encontrado una dirección d^1 en la que $a^T(s)d^1 > 0$, para $s \in \text{constr}(x^*)$. Por tanto, $0_n \notin \text{co}(A(x^*))$, lo que contradice la hipótesis.

■

Teorema 2.5

Si existe una solución posible x^* de (PSI1) tal que $0_n \in \text{co}(A(x^*))$, entonces existe una desigualdad inestable en el sistema de restricciones. El recíproco también es cierto.

Demostración

Si $0_n \in \text{co}(A(x^*))$ existirán $\lambda_i > 0$ con $\sum_{i=1}^m \lambda_i = 1$, tales que $0_n = \sum_{i=1}^m \lambda_i a(s_i)$, para $s_i \in \text{constr}(x^*)$, $i = 1, \ldots, m$. Tomemos $a(s_1) = -(1/\lambda_1) \sum_{i=2}^m \lambda_i a(s_i)$ y veamos que $a^T(s_1)x = b(s_1)$ para todo $x \in F$, i.e. que s_1 es inestable. Por reducción al absurdo, si existe una solución posible x tal que $a^T(s_1)x > b(s_1)$, entonces

$$b(s_1) < a^T(s_1)x = -(1/\lambda_1)\sum_{i=2}^{m}\lambda_i a^T(s_i)x \leq -(1/\lambda_1)\sum_{i=2}^{m}\lambda_i b(s_i) = b(s_1),$$

pues $\sum_{i=1}^{m} \lambda_i b(s_i) = \sum_{i=1}^{m} \lambda_i a^T(s_i)x = 0$.

Recíprocamente, sea s^* tal que $a^T(s^*)x = b(s^*)$ para todo $x \in F$. Si existe cualquier solución posible x tal que $0_n \notin \text{co}(A(x))$, siguiendo el razonamiento del Teorema 2.4, podemos construir un punto x^1 cuya función de holgura sea $z^1(s) > 0$ en $\mathcal{S}$, luego $a^T(s^*)x^1 > b(s^*)$. Por lo tanto, $0_n \in \text{co}(A(x))$ para cualquier solución posible x.

■

Veamos algunos ejemplos en los que quedan reflejados los resultados anteriores:

Ejemplo 2.3

Consideremos el problema

$$(P_3) \quad \begin{array}{l} \text{Min } x_1 \\ \text{s.a. } (s-1)x_1 + s(s-2)x_2 \geq 2s(2-s)^2 \quad s \in [0.2] \end{array}$$

Es fácil ver que $F = \{(0,\delta)' : \delta \leq -4\}$ y que F coincide con el conjunto de soluciones óptimas, aunque solo la solución $\tilde{x} = (0,-4)'$ es un punto extremo. Consideremos $x(\delta) = (0,\delta)'$ con $\delta \leq -4$, entonces $z(\delta) = (\delta + 4 - 2s)s(s-2) \geq 0$, $\text{constr}(x(\delta)) = \{0,2\}$ y $\text{co}(A(x(\delta))) = \{x : -1 \leq x_1 \leq 1,\ x_2 = 0\}$.

Tenemos, pues, dos desigualdades inestables correspondientes a los índices $\{0,2\}$. Además, $\text{co}(A(x))$ contiene el vector nulo para cada solución óptima, aunque $\{0_2\}$ no es un gradiente activo.

El dual asociado (PSI1*) tiene como soluciones posibles (y óptimas) las siguientes: $\lambda(2) = \alpha$, $\lambda(0) = \alpha - 1$, $\alpha \geq 1$, $\lambda(s) = 0$ para $s \in (0,2)$. Nótese que los índices correspondientes a las desigualdades inestables son los que caracterizan las soluciones del problema dual. Así, pues, la eliminación de las restricciones asociadas con $s = 0$ y $s = 2$ mantiene inalterado el conjunto posible F, pero convierte en inconsistente el problema dual.

Ejemplo 2.4

Consideremos el problema

(P_4) Min $x_1 + (1/2)x_2 + (1/3)x_3$

s.a. $(s_1 + s_2^2 + 1)x_1 + (s_1 s_2 - s_2^2)x_2 + (s_1 s_2 + s_2^2 + s_2)x_3 \geq 1 \quad s \in [0,1] \times [0,1]$

Sea $x_o = (1, -1, 1)'$, cuya función de holgura es $z_o(s) = s_1 + 3s_2^2 + s_2$, $\text{constr}(x_o) = \{0_2\}$, siendo $a(0_2) = (1, 0, 0)'$. Como consecuencia del Teorema 2.4, podemos afirmar que $0_3 \notin \text{co}(A(x))$ para cualquier solución posible de (P_4).

3. UN TEOREMA DE CARACTERIZACIÓN DE ÓPTIMO

La teoría de Kuhn y Tucker ha dado muy buenos resultados en el estudio de la optimalidad en PSI convexa diferenciables (véase p.e. Goberna *et al.* (81), Hettich y Zencke (82)). Es bien sabido que las condiciones necesarias exigen introducir una condición de regularidad sobre las restricciones, salvo en el caso finito lineal. Presentamos en esta sección una condición necesaria de optimalidad de tipo Kuhn-Tucker para el problema (PSI2) basada en una extensión de la cualificación de restricciones de Mangasarian-Fromovitz.

Definición 3.1

Sea $(x; z)$ un punto posible para el (PSI2) tal que $\text{constr}(x) \neq \emptyset$. Diremos que x satisface la *cualificación de restricciones de Mangarasian-Fromovitz (CRMF)* si existe $d \in \mathbb{R}^n$ tal que $a(s)^T d > 0$ para todo $s \in \text{constr}(x)$.

La condición anterior es una cualificación de restricciones "local" para los problemas de programación semi-infinita lineal y permite asegurar que podemos movernos partiendo de una solución x a lo largo de una dirección d sin perder la factibilidad. Además, ya hemos visto en la demostración del Teorema 2.4 que la cualificación de Mangasarian-Fromovitz es equivalente a la condición $0_n \notin \text{co}(A(x))$. Si $\text{card}(\text{constr}(x))$ es finito la equivalencia se sigue del Teorema de alternativa de Gordan y, en ese caso, dicha condición puede ser fácilmente comprobada resolviendo un PL.

Teorema 3.1

Sea $(x; z)$ un punto posible para el (PSI2) cuyo $\text{constr}(x) \neq \emptyset$, y supongamos que verifica la cualificación CRMF. Entonces $(x; z)$ es óptimo para el (PSI2) si y sólo si existen un subconjunto $\{s_1, s_2, \ldots, s_m\}$ de constr(x) y unos escalares $\lambda_i \geq 0$, $i = 1, \ldots, m$ tales que $c = \sum_{i=1}^{m} \lambda_i a(s_i)$.

Demostración

Sea x una solución posible para (PSI1) cuya función de holgura es $z(s) = a^T(s)x - b(s)$. Si suponemos que $c = \sum_{i=1}^{m} \lambda_i a(s_i)$, para $\lambda_i \geq 0$ y $s_i \in$ constr(x) $i = 1, \ldots, m$, entonces $\{s_1, s_2, \ldots, s_m, \lambda_1, \lambda_2, \ldots, \lambda_m\}$ es una solución posible para el (PSI1*) y $\lambda_i(a^T(s_i)x - b(s_i)) = \lambda_i z(s_i) = 0 \quad i = 1, \ldots, m$. Del Teorema de Holgura Complementaria (Glashoff-Gustafson (83)) se sigue que x es solución óptima para (PSI1) y $\{s_1, s_2, \ldots, s_m, \lambda_1, \lambda_2, \ldots, \lambda_m\}$ solución óptima para (PSI1*).

Recíprocamente, co$(A(x))$ es un compacto que no contiene a 0_n, por hipótesis, luego el cono convexo generado por el conjunto $A(x)$, $K\{A(x)\}$, es un cerrado (Corolario 9.6.1 en Rockafellar (70)). Además, por la CRMF existe $d \in \mathbb{R}^n$ tal que $a(s)^T d > 0$ para todo $s \in$ constr(x).

Supongamos que $c \notin K\{A(x)\} =$ cl$K\{A(x)\}$, y veamos que eso contradice que x sea solución óptima de (PSI1). Por el Teorema de Farkas generalizado (Goberna *et al.* (84)) existe $u \in \mathbb{R}^n$ tal que $a(s)^T u \geq 0$ para todo $s \in$ constr(x) y $c^T u < 0$. Definimos $y = u + \gamma d$. Para $\gamma > 0$ suficientemente pequeño se tiene que $c^T y < 0$ y que $a(s)^T y > 0$, para todo $s \in$ constr(x).

Sea $z(s)$ la función de holgura de x, definimos $\mu =$ inf$\{-z(s)/a^T(s)y$, para $s \in \mathcal{S}$ tales que $a^T(s)y < 0\}$, $\mu \geq 0$. Veamos que $\mu > 0$ por reducción al absurdo. Supongamos que $\mu = 0$, entonces, para $K \in \mathbb{N}$ suficientemente grande y $j \geq K$, existirá $t_j \in \mathcal{S}$, tal que $-(z(s_j)/a^T(s_j)y) < 1/j$, siendo $a^T(s_j)y < 0$. Realmente, como $a^T(\cdot)y$ es una función continua y acotada en el compacto $\mathcal{S}$, podemos considerar sin pérdida de generalidad que $z(t_j) < 1/j$. Entonces existirá una subsucesión $\{t_{jk}\}_{k \geq 1}$ en $\mathcal{S}$ convergente a $t^o \in \mathcal{S}$. Por continuidad de $z(s)$ y $a(s)$:

(i) $z(t^o) = \lim z(t_{jk}) = 0$, por tanto $t^o \in$ constr(x),

(ii) $a^T(t^o)y = \lim a^T(t_{jk})y \leq 0$.

Llegamos a una contradicción, puesto que $a(s)^T y > 0$, para todo $s \in$ constr(x). Luego, necesariamente $\mu > 0$.

Sea $x_1 = x + \mu y$, por construcción $x_1 \in F$ y $c^T x_1 < c^T x$. Por tanto, x no es una solución óptima.

■

Nota. Tal como acabamos de ver, con la condición de Mangasarian-Fromovitz se asegura que $K\{A(x)\}$ es cerrado, aun con infinitas restricciones activas, lo que cuando $v(P)$ es finito permite afirmar que el dual tiene solución, no habiendo fallo de dualidad (Teorema 10.19 en Glashoff y Gustafson (83)). Estos autores utilizan como condición de regularidad que el cono de los momentos $K\{[a^T(s), b(s)]' : s \in \mathcal{S}\}$ sea cerrado, lo que para nuestro problema es una implicación de la CRMF.

Cuando un punto posible satisface la CRMF y tiene un número finito de restricciones activas la aplicación del anterior teorema permite contrastar la optimalidad de una solución posible mediante la resolución de un problema lineal finito (León y Vercher (92)).

En general no es sencillo comprobar si 0_n pertenece o no a la envoltura convexa de un conjunto infinito de vectores, aunque verificar la CRMF en esos casos puede ser innecesario si se ha comprobado la condición $0_n \notin \operatorname{co}(A(x))$ en otro punto posible. En el ejemplo siguiente comprobamos que se verifica la cualificación de restricciones sobre una solución posible no óptima.

Ejemplo 3.1

Sea

$$\begin{array}{ll} (P_5) & \text{Min } x_1 + (1/2)x_2 + (1/3)x_3 \\ & \text{s.a. } s_1x_1 + s_2x_2 + x_3 \geq s_1^2 + s_2 \quad s \in [0,1] \times [0,1] \end{array}$$

Consideremos la solución posible $x = (1,1,0)'$ cuyos $z(s_1, s_2) = s_1 - s_1^2$ y $\operatorname{constr}(x) = \left\{ \begin{pmatrix} 0 \\ s_2 \end{pmatrix}, \begin{pmatrix} 1 \\ s_2 \end{pmatrix} : s_2 \in [0,1] \right\}$. Dado que la última componente de cualquier vector $a(s)$ es igual a 1, puede comprobarse fácilmente que $0_3 \notin \operatorname{co}\{(A(x)\}$, donde $A(x) = \left\{ \begin{pmatrix} 0 \\ s_2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ s_2 \\ 1 \end{pmatrix} : s_2 \in [0,1] \right\}$. Además, se tiene que $c = (1, 1/2, 1/3)' \notin K\{A(x)\}$ y, por tanto, x no es solución óptima.

Nota. Cuando la función $b(s) \geq 0$, para todo $s \in \mathcal{S}$, la no acotación del problema semi-infinito lineal puede ser fácilmente detectada.

Lema 3.2

Supongamos un problema (PSI1) tal que $b(s) \geq 0$ para todo $s \in \mathcal{S}$. Sea x una solución posible, si $c^T x < 0$, entonces el problema es no acotado.

Demostración

Dado que x es una solución posible $a(s)^T x \geq b(s) \geq 0$, para todo $s \in \mathcal{S}$, luego x es una dirección de recesión de F y la función objetivo disminuye estrictamente a lo largo de esa dirección. Por consiguiente, el problema no está acotado.

■

Volviendo al Ejemplo 3.1, consideremos ahora la solución posible $x = (-3/2, 1, 5/2)'$ con función de holgura $z(s_1, s_2) = -(3/2)s_1 + 5/2 - s_1^2$ y $\mathrm{constr}(x) = \left\{ \begin{pmatrix} 1 \\ s_2 \end{pmatrix} : s_2 \in [0,1] \right\}$, siendo $c^T x = -1/6 < 0$. Del lema anterior se sigue que el problema primal (P_5) es no acotado. Puede comprobarse fácilmente que su dual es inconsistente.

Para el Ejemplo 2.4, donde $b(s) = 1$, se comprueba la no acotación de (P_4) tomando $x = (1.1, -2, -1/3)'$. En ese punto $z(s_1, s_2) = 1/10 + 83/30 s_2^2 - 1/3 s_2 - 7/3 s_1 s_2 + 11/10 s_1$ y se puede comprobar que $z(s_1, s_2) > 0$ en $[0,1] \times [0,1]$. Además $c^T x = -1/90 < 0$.

4. OTRAS CONDICIONES DE REGULARIDAD

Para caracterizar las soluciones óptimas de los problemas que no satisfacen la condición de Mangasarian-Fromovitz necesitamos utilizar alguna cualificación más débil. Hemos llevado a cabo un estudio comparativo de la CRMF y algunas de las condiciones que han sido utilizadas en PSI convexa.

Cuando $\mathcal{S}$ es un intervalo de $\mathbb{R}$ y las funciones a y b son analíticas, Anderson y Lewis (89) establecen una caracterización de optimalidad para puntos extremos no degenerados del problema (PSI2). Veamos que la condición de no degeneración implica la CRMF, también para cualquier subconjunto $\mathcal{S}$ compacto de $\mathbb{R}^p$, con $p > 1$.

Teorema 4.1

Sea $(x; z)$ una solución posible para el (PSI2), tal que $\mathrm{constr}(x) \neq \emptyset$. Si $(x; z)$ es un punto extremo no degenerado entonces se satisface la CRMF.

Demostración

Sea $(x;z)$ un punto extremo no degenerado de (PSI2), entonces se cumplirá que: $X = \mathbb{R}^n \times \mathbb{C}[\mathcal{S}] = B((x;z)) \oplus N(A)$.

Sea $(d;h) \in B((x;z))$, entonces $(z+\lambda h)(s) \geq 0$ y $(z-\lambda h)(s) \geq 0$ para todo $s \in \mathcal{S}$. Para $s \in$ constr(x) tenemos que $z(s)=0$, entonces $\lambda h(s) \geq 0$ y $-\lambda h(s) \geq 0$, para todo $\lambda > 0$. En consecuencia, $h(s)=0$ para $s \in$ constr(x).

Sean $p \in \mathbb{R}^n$ y M constante positiva tales que $(p;M) \in X$. Entonces existirá $(d;h) \in B((x;z))$ tal que $(p-d;M-h) \in N(A)$, con lo que $a^T(s)(p-d) = M - h(s)$, para todo $s \in \mathcal{S}$. En particular, para $s \in$ constr(x): $a^T(s)(p-d) = M > 0$. Luego existe una dirección $u \in \mathbb{R}^n$ tal que $a^T(s)u > 0$, para $s \in$ constr(x).

■

Nota. En consecuencia, aquellos problemas que no satisfacen la CRMF solo tienen puntos extremos degenerados, lo que dificultará su resolución mediante cualquier método primal tipo-simplex. El recíproco no es cierto (véase Ejemplo 4.1).

Cuando tenemos una solución posible con un número finito de restricciones activas podemos utilizar la caracterización que han establecido Anderson y Lewis (89) de punto extremo no degenerado. Dicha caracterización implica que los gradientes activos son linealmente independientes y, en ese caso, la demostración del anterior resultado es obvia. Pero, para aquellos puntos $x \in F$ tales que $A(x) \neq \emptyset$ y card(constr$(x) \leq n$, la CRMF no implica que los vectores de $A(x)$ sean linealmente independientes. Veamoslo:

Ejemplo 4.1

(P_6) Min $x_1 + (1/2)x_2$
s.a. $x_1 + s^2 x_2 \geq 0 \quad s \in [-1,1]$

Sea $x = (2,-2)'$ cuya función de holgura es $z(s) = 2(1-s^2)$ y constr$(x) = \{-1,1\}$. Entonces, $A(x) = \{a(1), a(-1)\}$, co$(A(x)) = \{(1,1)'\}$ y los gradientes activos son linealmente dependientes.

Puesto que $F = \{x \in \mathbb{R}^2 : x_1 \geq 0$ y $x_1 + x_2 \geq 0\}$, el único punto extremo es 0_2, cuyo constr$(0_2) = [-1,1]$. Esta es la solución óptima del problema y un punto extremo degenerado. Nótese que $z(s) = 0$, para $s \in [-1,1]$, entonces $B(x;z) = \{(d;h) \in \mathbb{R}^n \times \mathbb{C}[\mathcal{S}] : h(s) = 0$ para todo $s \in [-1,1]\}$. Además,

$N(A) = \{(y; g) \in \mathbb{R}^n \times \mathbb{C}[\mathcal{S}]: g(s) = y_1 + s^2 y_2$ para todo $s \in [-1, 1]\}$, así pues, es fácil ver que $\mathbb{R}^n \times \mathbb{C}[\mathcal{S}] \neq B(x; z) \oplus N(A)$.

Hemos probado también que la CRMF es equivalente a la generalización al caso semi-infinito lineal de la cualificación de Cottle. La condición $G_\circ \neq \emptyset$, para aquellos $x \in F$ cuyo $\text{constr}(x) \neq \emptyset$, ha sido también utilizada como condición de regularidad en Krabs (79) para problemas semi-infinitos no lineales.

Definición 4.1

Sea x un punto posible cuyo $\text{constr}(x) \neq \emptyset$. Diremos que satisface la *cualificación de restricciones de Cottle* si y sólo si $\mathbb{C} \subseteq \text{ cl } G_\circ$, donde $\mathbb{C} = \{d: a^T(s)d \geq 0 \ s \in \text{ constr}(x)\}$ y $G_\circ = \{d: a^T(s)d > 0 \ s \in \text{ constr}(x)\}$.

Teorema 4.2

Para el problema (PSI2), la cualificación de restricciones de Mangasarian-Fromovitz es equivalente a la cualificación de Cottle.

Demostración

Sea $x \in F$ tal que x satisface la cualificación de Cottle. Si x no cumpliese la CRMF el sistema $\{a^T(s)d > 0, \ s \in \text{ constr}(x)\}$ sería inconsistente y, por tanto, $G_\circ = \emptyset$. En ese caso cl $G_\circ = \emptyset$ y por hipótesis $\mathbb{C} = \emptyset$. Pero $0_n \in \mathbb{C}$, con lo que llegaríamos a una contradicción. Así pues, x satisface la CRMF.

Recíprocamente, si x satisface CRMF tenemos que $G_\circ \neq \emptyset$. Ahora bien, $G_\circ = \cap\{C_s: s \in \text{ constr}(x)\} \neq \emptyset$, siendo $C_s = \{d: a^T(s)d > 0\}$, con int $C_s = C_s$. Por Teorema 6.5 en Rockafellar (70) cl $G_\circ = $ cl $[\cap\{C_s: s \in \text{ constr}(x)\}] = \cap\{\text{cl } C_s: s \in \text{ constr}(x)\} = \{d: a^T(s)d \geq 0: s \in \text{ constr}(x)\} = \mathbb{C}$ y, por tanto, en x se satisface la cualificación de Cottle.

■

En un reciente trabajo León y Vercher (92) han demostrado, de forma constructiva, la equivalencia entre la condición de Slater y la CRMF para el problema (PSI1). La cualificación de Slater es la condición de regularidad más utilizada en PSI convexa. Así, pues, hemos establecido condiciones locales, fácilmente verificables, que tienen la misma potencia que la cualificación de Slater en Programación Semi-Infinita Lineal. Si esta última no se cumple hay que recurrir a condiciones más débiles: cualificación de Farkas-Minkowski y regularidad Lagrangiana.

Para PSI convexa las relaciones entre la condición de Slater y las anteriores cualificaciones han sido estudiadas en López y Vercher (83), donde se demuestra:

(i) que la condición de Slater implica que el conjunto posible satisface la propiedad de Farkas-Minkowski y que todos los puntos posibles son regulares Lagrangianos, y

(ii) que ambas cualificaciones permiten también asociar a cada solución óptima un punto de Kuhn-Tucker.

La condición de punto regular Lagrangiano es una cualificación local para el problema de PSI convexa. Recordemos la definición de *cono de las tangentes a* F *en* $\overline{x}$, para $\overline{x} \in \mathrm{cl}F$:

$$T(F, \overline{x}) = \left\{ z : z = \lim_{k \to \infty} \lambda_k (x_k - \overline{x}), \quad \text{donde } \lambda_k > 0,\ x_k \in F \text{ y } \lim_{k \to \infty} x_k = \overline{x} \right\}$$

Sean $B(\overline{x}) = -A(\overline{x}) = \{-a(s): s \in \mathrm{constr}(\overline{x})\}$, $B(\overline{x})^* = \{d \in \mathbb{R}^n: a(s)^T d \geq 0,$ $s \in \mathrm{constr}(\overline{x})\}$ es el cono polar de $B(\overline{x})$, y $K\{B(\overline{x})\}$ es el cono convexo generado por $B(\overline{x})$.

Definición 4.2

Para el problema (PSI1) decimos que $\overline{x} \in F$ es un *punto regular Lagrangiano* si:

(i) $T(F, \overline{x}) = \mathbb{R}^n$, cuando $\mathrm{constr}(\overline{x}) = \emptyset$

(ii) (a) $T(F, \overline{x}) \supseteq B(\overline{x})^*$ y
(b) $K\{B(\overline{x})\}$ es cerrado, cuando $\mathrm{constr}(\overline{x}) \neq \emptyset$.

Definición 4.3

Se dice que el sistema consistente $\{a^T(s)x \geq b(s): s \in S\}$ es un *sistema de Farkas-Minkowski* si toda relación consecuente del sistema lo es de un subsistema finito.

Definición 4.4

Se dice que un problema (PSI1) satisface la *cualificación de restricciones de Farkas-Minkowski* si el sistema $\{a^T(s) \geq b(s): s \in \mathcal{S}\}$ es un sistema de Farkas-Minkowski.

Una caracterización y una excelente descripción de las propiedades de los sistemas de Farkas-Minkowski en PSI puede encontrarse en Goberna *et al.* (81).

En cuanto a la relación que existe entre las cualificaciones de regularidad Lagrangiana y de Farkas-Minkowski en PSIL podemos encontrar diferentes situaciones.

Ejemplo 4.2

$$\begin{array}{ll}(P_7) & \text{Min } x_1 \\ & \text{s.a. } (s-1)x_1 + s(s-2)x_2 \geq 0 \quad s \in [0,2]\end{array}$$

Es fácil ver que $F = \{(0,\delta)' : \delta \leq 0\}$ y que todas las soluciones son óptimas, aunque solo $\tilde{x} = 0_2$ es un punto extremo. Para $\tilde{x}$ tenemos que $\text{constr}(\tilde{x}) = [0,2]$, con lo que $\text{co}(A(\tilde{x})) = \{x : -1 \leq x_1 \leq 1,\ x_1^2 - 1 \leq x_2 \leq 0\}$. Consideremos $x(\delta) = (0,\delta)'$ con $\delta < 0$, entonces $z(\delta) = \delta s(s-2) \geq 0$, $\text{constr}(x(\delta)) = \{0,2\}$ y $\text{co}(A(\delta))) = \{x : -1 \leq x_1 \leq 1,\ x_2 = 0\}$. Así pues, (P_7) no satisface la CRMF, por lo que el punto extremo $\tilde{x}$ es necesariamente degenerado (Teorema 4.1).

El sistema $\{(s-1)x_1 + s(s-2)x_2 \geq 0 \quad s \in [0,2]\}$ es de Farkas-Minkowski, pues el cono $K = K\left\{\begin{pmatrix} a(s) \\ 0 \end{pmatrix} : s \in [0,2]\right\} = \left\{\begin{pmatrix} \alpha \\ \beta \\ 0 \end{pmatrix} : \beta \leq 0\right\}$ es cerrado (Corol. 3.1.1 en Goberna *et al.* (81)).

Todas las soluciones óptimas son puntos regulares Lagrangianos, pues para $\tilde{x}$ tenemos que $B(\tilde{x})^* = \{d : d_1 = 0,\ d_2 \leq 0\} = T(F,\tilde{x})$ y $K\{B(\tilde{x})\} = \{d : d_2 \geq 0\}$ es cerrado y para $x(\delta)$, con $\delta < 0$, $B(x(\delta))^* = \{d : d_1 = 0\} = T(F, x(\delta))$.

Puede verse fácilmente que en F tenemos dos desigualdades inestables, las correspondientes a los índices $\{0,2\}$. Así, las anteriores cualificaciones permiten la existencia de desigualdades inestables en el conjunto posible.

Como ya se había dicho antes, a cada solución óptima podemos asociarle un punto de Kuhn-Tucker. Dado que $c = a(2)$, la solución del problema dual es $\lambda(2) = 1$ y $\lambda(s) = 0$ para $s \neq 2$.

En el Ejemplo 4.2 se satisfacen simultáneamente las cualificaciones de Farkas-Minkowski y la regularidad Lagrangiana de todas las soluciones posibles. Sin embargo, en el siguiente no se verifica la cualificación de Farkas-Minkowski, aunque algunas de las soluciones óptimas son puntos regulares Lagrangianos. La cualificación local de regularidad Lagrangiana es la más débil de todas las que hemos estudiado para el problema (PSI1), la única desventaja de trabajar con ella es que debe comprobarse sobre cada candidato a óptimo.

Ejemplo 4.3

$$\begin{array}{ll}(P_8) & \text{Min } x_1 + x_2 \\ & \text{s.a. } sx_1 + s^2x_2 \geq s^3 \quad s \in [0,1]\end{array}$$

Es fácil ver que la desigualdad $x_1 \geq 0$ es una relación consecuente del sistema $\{sx_1 + s^2x_2 \geq s^3 : s \in [0,1]\}$. Sin embargo, no lo es para ningún subsistema finito: $\{s_ix_1 + s_i^2x_2 \geq s_i^3, \quad i = 1, 2, \ldots, k\}$. Podemos suponer que $s_i \neq 0$ y este subsistema finito es equivalente a: $\{x_1 \geq s_i^2 - s_1x_2, \quad i = 1, 2, \ldots, k\}$. En particular, tomando $x_2 = 2$ y $x_1 = \max\{s_j(s_j - 2)\}$ tenemos una solución del subsistema tal que $x_1 < 0$. Por tanto (P_8) no satisface la cualificación de restricciones de Farkas-Minkowski.

Se puede comprobar que $F = \{x \in \mathbb{R}^2 : x_1 \geq 0, \ x_1 + x_2 \geq 1\}$. Las soluciones posibles de la forma $x(\alpha) = \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$ con $\alpha > 1$, $\mathrm{constr}(x(\alpha)) = \{0\}$, no son óptimas y tampoco son puntos regulares Lagrangianos, ya que $B(x(\alpha))^* = \mathbb{R}^2$ pero $T(F, x(\alpha)) = \{y \in \mathbb{R}^2 : y_1 \geq 0\}$. Los puntos $x(\beta) = \begin{pmatrix} \beta \\ 1-\beta \end{pmatrix}$ con $\beta \geq 0$, son todos óptimos, y puesto que $z(\beta) = s(1-s)(s+\beta), z(\beta) = 0$ si y sólo si $s = 0$ ó $s = 1$. Luego $A(x(\beta)) = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ y $B(x(\beta))^* = \{y \in \mathbb{R}^2 : y_1 + y_2 \geq 0\} \forall \beta \geq 0$. Para $\beta > 0, x(\beta)$ es regular Lagrangiano, pues $T(F, x(\beta)) = \{y \in \mathbb{R}^2 : y_1 + y_2 \geq 0\}$. Ahora bien, el único punto extremo $\tilde{x} = (0,1)'$ —que también es solución óptima, pues $c = 1.a(1)$— no es regular Lagrangiano, ya que $T(F, x(0)) = \{y \in \mathbb{R}^2 : y_1 + y_2 \geq 0, \ y_1 \geq 0\}$.

5. REFERENCIAS

[1] **Anderson, E.J.** and **A.S. Lewis** (1989). "An extension of the simplex algorithm for semi-infinite linear programming". *Mathematical Programming*, **44**, 247–269.

[2] **Charnes, A., Cooper, W.W.** and **K.O. Kortanek** (1963). "Duality in semi-infinite programs and some works of Haar and Caratheodory". *Management Science*, **9**, 209–228.

[3] **Eckhardt U.** (1975). "Theorems on the dimension of convex sets". *Linear Algebra and its Applications*, **12**, 63–76.

[4] **Glashoff, K.** and **S.A. Gustafson** (1983). *Linear Optimization and Approximation.* Springer-Verlag. New York.

[5] **Goberna, M.A.** and **M.A. López** (1988). "A theory of linear inequality systems". *Linear Algebra and its Applications*, **106**, 77–115.

[6] **Goberna, M.A., López, M.A.** and **J. Pastor** (1981). "Farkas-Minkowski systems in semi-infinite programming". *Applied Mathematics and Optimization*, **7**, 295–308.

[7] **Goberna, M.A., López, M.A., Pastor, J.** and **Vercher, E.** (1984). "Alternative Theorems for Infinite Systems with Applications to Semi-Infinite Games". *Nieuw Archief voor Wiskunde*, **4**, 218–234.

[8] **Hettich, R.** and **P. Zenke** (1982). *Numerische Methoden der Approximation und Semi-infiniten Optimierung.* B.G. Teubner, Stuttgart.

[9] **Holmes, R.** (1975). *Geometric Functional Analysis and its Applications.* Springer-Verlag. New York.

[10] **Krabs, W.** (1979). *Optimization and Approximation.* Wiley, Chichester.

[11] **León, T.** and **E. Vercher** (1992). "An Optimality Test for Semi-Infinite Linear Programming". *Optimization*, **26**, 51–60.

[12] **López, M.A.** and **E. Vercher** (1983). "Optimality conditions for non-differentiable convex semi-infinite programming". *Mathematical Programming*, **27**, 307–319.

[13] **Nash, P.** (1985). "Algebraic fundamentals of linear programming". In: *Infinite Programming, Proceedings.* Springer-Verlag. Berlin, 37–52.

[14] **Rockafellar, R.T.** (1970). *Convex Analysis.* Princeton University Press. Princeton, New Jersey.

ENGLISH SUMMARY:

NECESSARY OPTIMALITY CONDITIONS IN SEMI-INFINITE LINEAR PROGRAMMING: CONSTRAINT QUALIFICATIONS AND PROPERTIES OF THE FEASIBLE SET

Teresa León and Enriqueta Vercher

Consider the problem of minimizing a linear function $c^T x$ subject to an arbitrary family of linear constraints $a^T(s) \geq b(s), s \in \mathcal{S}$. When the index set $\mathcal{S}$ is a compact set in $\mathbb{R}^p$ and the functions $a_1, \ldots, a_n$ and b are continuous on $\mathcal{S}$ we have a continuous problem of Semi-Infinite Linear Programming (SILP).

We can state the above problem in the standard LP form by using the slack function $z(s)$:

$$\begin{array}{lllll} \text{(PSI2)} & \text{Min } c^T x & & & \\ & \text{s.t. } a^T(s)x & - & z(s) = b(s) & s \in \mathcal{S} \\ & x \in \mathbb{R}^n & & z(s) \geq 0 & s \in \mathcal{S} \end{array}$$

Let us assume that the problem is consistent. Note that the set of active constraints at a feasible point coincides with the set of the zeros of its associated slack function.

First of all we prove that the candidate points for optimality are those feasible points with some active constraints. Even some inner points may have the null vector as its unique active gradient.

In SILP the null gradient, active or not, plays a crucial role in the study of some properties of the feasible set and in the statement of the necessary optimality conditions.

We have found that if a feasible point has the null vector as an active gradient then all the remaining feasible points also have it. And the same situation occurs when the null vector belongs to the convex hull of the active gradients set for some feasible point. Moreover, in that case we have shown that there exists an unstable inequality of the feasible set (the converse also holds).

We establish a necessary Kuhn-Tucker type optimality condition for SILP based on an extension of the Mangasarian-Fromovitz constraint qualification (MFCQ). Although the statement of this constraint qualification seems local, in fact it is a global condition (the proof of the last result makes use of Gordan's Alternative Theorem). So, we have to check this CQ only once.

Working with the MFCQ has a major advantage: if you have a feasible point with a finite number of active constraints, you simply have to solve an LP in order to test it.

There are other properties of the problem which guarantee that MFCQ holds, for instance, the existence of a nondegenerate extreme point. Therefore, those problems for which MFCQ does not hold only have degenerate extreme points, and this would be noted for designing appropiate methods to deal with them.

It is well known that Kuhn-Tucker's criterion fails to characterize the optimality unless a CQ holds. Then we have studied other conditions weaker than MFCQ: Lagrangian regularity and the Farkas-Minkowski constraint qualification. We show the relationships among them using suitable examples.

QÜESTIIÓ, Vol 18, 2 pp. 209-228, 1994

SENSITIVITY EXAMINATION OF THE SIMULATION RESULT OF DISCRETE EVENT DYNAMIC SYSTEMS WITH PERTURBATION ANALYSIS*

TAMAS KOLTAI†, JUAN LARRAÑETA, LUIS ONIEVA and SEBASTIÁN LOZANO‡

Universitat de Sevilla

Simulation completed with perturbation analysis provides a new approach for the optimal control of queuing network type systems. The objective of this paper is to calculate the sensitivity range of finite zero-order perturbation, that is, to determine the maximum and minimum size of perturbation within which zero-order propagation rules can be applied. By the introduction of the concept of virtual queue and first and second level no-input and full-output matrices, an algorithm is provided which can solve this task efficiently in transfer lines and in relatively small general networks when short simulation run is required and the sensitivity of the individual sample path is in question. The implementation of the algorithm with the help of conventional simulation languages is also discussed and presented in an example.

Key words: Simulation, perturbation analysis, queuing networks.

---

*This research is part of the research project 'Expert Systems for the Scheduling and Dynamic Production Control in a Steel making Factory BOF/Secondary Metallurgy/Continuous Casting', sponsored by the European Coal and Steel Community, and carried out at the University of Seville, Department of Industrial Organization.

† Tamas Koltai. Associate Professor at The Technical University of Budapest, Visiting Professor at The University of Seville.

‡ Juan Larrañeta. Professor.
Luis Onieva. Associate Professor.
Sebastián Lozano. Associate Professor.
} University of Seville, School of Engineering, Department of Industrial Organization.

–Article rebut el desembre de 1992.
–Acceptat el desembre de 1993.

1. INTRODUCTION

The complex nature of design, planning and control of modern, highly automated production systems raises various challenging problems in the field of operations management. The difficulty of their examination comes from the fact that these are generally discrete event dynamic systems (DEDS) in the majority of the cases with a stochastic nature. Queuing network representation seems to provide appropriate modelling framework but their exact analytical examination can be performed just in very limited cases. Usually the combined application of simulation and exact optimization algorithm is required.

One of the most challenging solution of this problem seems to be the recently developed perturbation analysis (PA) which can provide gradient information from a single simulation experiment [4]. The idea is to perform a simulation experiment, and via an algorithm an estimate can be derived about the derivative of a performance measure of the system with respect to one of its parameters [6]. This gradient information can be used for iterative improvement of system performance [8],[11].

Various intriguing problems have been solved since the first presentation of the method. Propagation rules for infinitesimal and finite perturbations [5], examination of multi-class networks [1], various suggestions for avoiding or at least smoothing the effect of discontinuities are extending the applicability of the method [7]. Researchers of this field, however, have mostly concentrated on generating and/or propagating perturbations, but have avoided the examination of validity range within which the gradient information is correct. The infinitesimal approach deals with this problem by simply saying that the size of the perturbation is small enough not to hurt the deterministic similarity. The finite approach calculates accurately the effect on the performance measures or on other activities with higher order propagation rules, but it also fails to provide information about the validity [5]. The effect of a specific perturbation is calculated correctly but if the perturbation changes the calculation has to be performed again.

The objective of this paper is to calculate the sensitivity range of zero-order perturbation, that is, to determine the maximum and minimum size of perturbation within which zero-order propagation rules can be applied. The practical significance of this problem is, that within the determined sensitivity limits a linear extrapolation of the change of the examined performance measure based on the gradient information is accurate [3].

An intuitive illustration of this problem can be given by linear programming (LP). The basic idea of the optimization process in LP is that an initial basis solution is generated and an iterative improvement is performed with the help of the shadow prices. The shadow price is actually the gradient of the objective function with respect to the right hand side parameters. The validity range of the shadow price imposes limits on the improvement and also provides important information about the limiting constraints [2]. With PA we may try to construct a similar process. The simulation results can be considered as initial solutions. The gradient information can be gained by the propagation rules. The objective of this paper is to derive the validity range of the gradient.

2. PROBLEM DEFINITION

Suppose we have a transfer line type queuing network consisting of M resources and the same number of queues, illustrated in Fig. 1. Every resource (R_j) is preceded by a queue (Q_j) with $c(j)$ capacity. ($c(j)$ includes the place in R_j as well.) At the queues the first-in-first-out (FIFO) rule is applied.

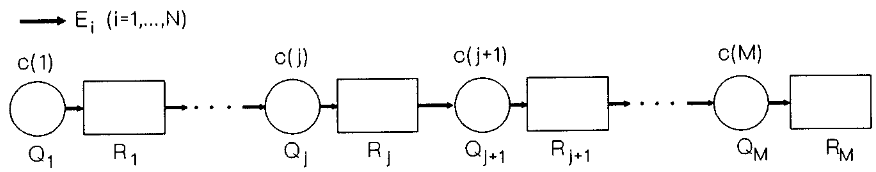

Figure 1.
Transfer line type queuing network.

N entities (E_i) enter the network and follow a flow-shop type process. The order of entities at every resource is the same. The operation time of E_i at R_j in a specific sample path is $t^{\omega}_{i,j}$ (shortly $t_{i,j}$)[1].

A simulation experiment is performed. As a result a time schedule of the activities is gained and summarized in a matrix B where $b_{i,2j-1}$ denotes the beginning time and $b_{i,2j}$ the ending time of the operation of E_i at R_j.

A single finite perturbation is introduced at E_x on $R_y(\delta_{x,y})$ and its effect is examined on the throughput time (T), that is, on the total operation time

[1] Since all the calculations introduced here, refer to a specific sample path we abandon ω when it is not misleading.

necessary to finish the manufacturing of N entities. We would like to determine the gradient of the throughput time with respect to the change of the operation time $t_{x,y}$ and the validity range of this gradient.

Our analysis is based on the **event sequence table** introduced by Ho and Cassandras [6]. The principle of this table is that every resource participate in one of three mutually exclusive events, that is,

- performing an operation on an entity (OP),
- being idle and waiting for the arrival of an entity (NI),
- being blocked by the proceeding part of the process (FO).

The event sequence table contains the order of these events at the resources. The event sequence table changes if any of its event disappears or new one appears as a consequence of any change of the system control variables. **Deterministic similarity** means that the event sequence table of the original and perturbed sample path are equal.

Based on these definitions our problem can be expressed on the following way:

We suppose that $\delta_{x,y}$ is not infinitesimal, but small enough not to hurt the deterministic similarity. We want to determine the set of $\delta_{x,y}$ which satisfies this requirement.

3. EVENT SEQUENCE AND VIRTUAL QUEUE

We will assign numeric values to the activities of the event sequence table and group NI (no-input) and FO (full-output) activities in different matrices. First, however, we have to introduce the concept of the 'virtual queue'.

Entities may wait for a resource either in a queue or, in case of blocking, in an other resource. Let's suppose that E_i is blocked in R_j (Fig. 2.). This practically means that E_i joins to the $c(v)$ entities waiting for the release of R_v, by this way transforming R_j into a part of the queue of R_v. If in the meantime entities are entering in Q_j then they will join a line which now consists of the entities waiting in Q_v, R_j and Q_j for the release of R_v. If Q_j is full and blocks R_p then the entities in R_p and in Q_p join the same line, transforming R_p and Q_p as well into part of the virtual queue of R_v.

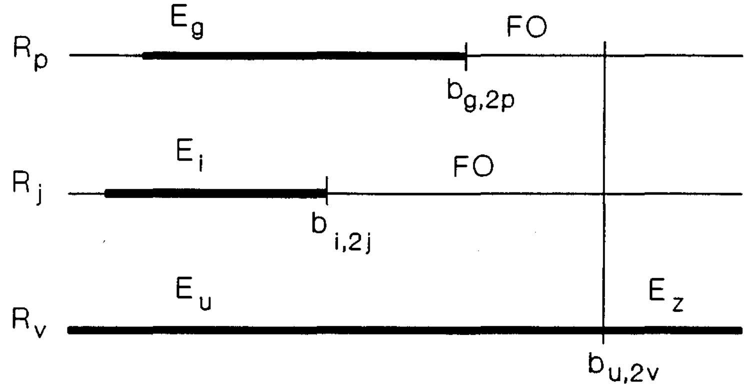

Figure 2.
Illustration of the virtual queue.

Definition 1:

The virtual queue consists of all those resources and queues in which the entities are waiting at the same time for the release of the same resource.

In further discussions, let $\boldsymbol{V(t)} \subset \{\boldsymbol{1,\ldots,M}\}$ denote the index set of the resources belonging to a virtual queue of $\boldsymbol{R_v}$ $\boldsymbol{v} \in \boldsymbol{V(t)}$, as $\boldsymbol{E_u}$ causes blocking in it when $b_{u,2v-1} \leq t \leq b_{u,2v}$. If we move upstream in this virtual queue, then the utmost resource which is blocked (by $\boldsymbol{E_g}$) will be $\boldsymbol{R_p}$ $\boldsymbol{p} \in \boldsymbol{V(t)}$.

A state of a queue changes if the number of entities in the queue changes [6]. This will be true for the virtual queue as well. The only difference is that even if the capacity of all the resources is one, the state of the virtual queue may change by more than one. It may occur when a queue joins the virtual queue and contains more than one entity. For example in Fig. 2. all the entities in Q_p join the virtual queue of R_v at $b_{g,2p}$. It may also occur that more than one resource and their queues join or leave the virtual queue at the same time.

The capacity of the virtual queue is the sum of the capacity of all the queues participating in it (In Fig. 2. $c'(v) = c(v) + c(j) + c(p)$). An extreme case of the virtual queue is a transfer line where the last workstation is the bottleneck of the system and blocks all the other machines.

The pseudo code of the algorithm to find the resource and entity (R_v, E_u) which are responsible for the appearance of the virtual queue containing R_j, E_i is represented in Fig. 3.

```
fin:= 0
WHILE j + 1 ≤ M AND i − c(j + 1) > 0 AND fin:= 0 DO
          BEGIN
          IF b_{i−c(j+1),2(j+1)} > b_{i,2j} THEN
             BEGIN
             j := j + 1
             i := i − c(j)
             END
          ELSE
             BEGIN
             u := i + c(j)
             v := j − 1
             fin:= 1
             END
          END
store (v, u)
```

Figure 3. Pseudo code of the algorithm for searching E_u, R_v.

The last resource element of the virtual queue and the entity residing in it (R_p, E_g) can be found by the algorithm given in Fig. 4. It is assumed that R_v and E_u is already known and the search is started at E_i in R_j $j \in V(t)$.

```
fin:= 0
WHILE j − 1 > 0 AND i + c(j) ≤ N AND fin:= 0 DO
          BEGIN
          IF b_{i,2j} < b_{u,2v} THEN
             BEGIN
             i := i + c(j)
             j := j − 1
             END
          ELSE
             BEGIN
             g := i − c(j + 1)
             p := j + 1
             fin:= 1
             END
          END
store (p, g)
```

Figure 4. Pseudo code of the algorithm for searching E_g, R_p.

4. QUANTITATIVE ANALYSIS OF THE FO AND NI EVENTS

We will define two FO and two NI matrices based on whether the event (FO or NI) in question is **experienced at** or **caused by** E_i in R_j.

4.1. Full-output matrices

Definición 2

The first level full-output matrix (FO(1)) expresses the duration of the FO R_j has to **endure**, while it contains E_i.

A FO event exists when an entity can not leave the resource because the queue (either real or virtual) it has to enter is full. On Fig. 5(a). and 5(b). it can be seen that E_i wants to leave R_j and go to R_{j+1} but Q_{j+1} is full. Since the capacity of Q_{j+1} is $c(j+1)$, R_j will be blocked until $E_{i-c(j+1)}$ can not leave R_{j+1} and free a place in Q_{j+1}. This event may occur at $b_{i-c(j+1),2(j+1)}$ if $j+1 \notin V(t)$ or at $b_{u,2v}$ if $j+1 \in V(t)$. The first level full-output matrix ($fo_{i,j,1}$) denotes the length of the FO event and is calculated as follows,

$$fo_{i,j,1} = \begin{cases} b_{i-c(j+1),2(j+1)} - b_{i,2j} & \text{if } j+1 \not\subset V(t) \\ b_{u,2v} - b_{i,2j} & \text{otherwise} \end{cases} \tag{1}$$

If $fo_{i,j,1} > 0$ then FO really exists when E_i wants to leave R_j and the blocking of R_j lasts $fo_{i,j,1}$. If $fo_{i,j,1} \leq 0$ then there is no FO and $fo_{i,j,1}$ expresses the available time to avoid it.

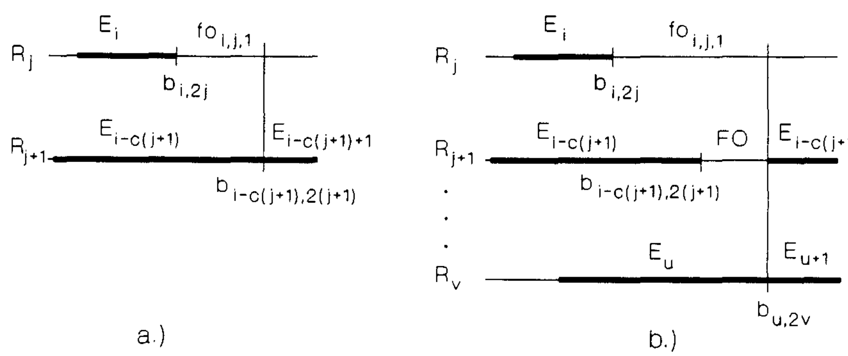

Figure 5.
Calculation of the FO(1) matrix.

Definition 3

The second level full-output matrix (FO(2)) expresses the duration of the FO R_j **causes**, while it contains E_i. The calculation of $fo_{i,j,2}$ is as follows (fig 6(a), 6(b)):

If, $(j-1) \notin V(t)$ then,

$$fo_{i,j,2} = b_{i+c(j),2(j-1)} - b_{1,2j} \tag{2}$$

If, $(j-1) \in V(t)$ then,

$$fo_{i,j,2} = \begin{cases} b_{g,2p} - b_{i,2j} & \text{if, } b_{g+1,2(p-1)} \leq b_{u,2v} \\ b_{g+1,2(p-1)} - b_{i,2j} & \text{otherwise} \end{cases} \tag{3}$$

where (R_p, E_g) is the last in this virtual queue.

If $fo_{i,j,2} < 0$, then FO really exists when E_g wants to leave R_p. If $fo_{i,j,2} \geq 0$, then there is no FO and $fo_{i,j,2}$ expresses the available time to avoid it.

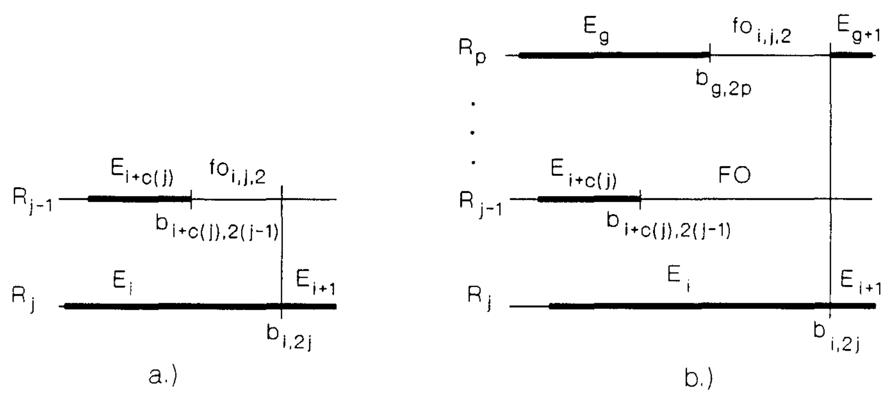

Figure 6. Calculation of the FO(2) matrix.

4.2. No-input matrices

Definition 4

The first level no-input matrix (NO(1)) expresses the duration of the NI R_j has to **endure** after completing the operation on E_i.

A no-input event exists when a resource is empty and waits for an incoming entity. It may occur just after the termination of an operation (Fig. 7(a).) or after the release of the resource from blocking (Fig. 7(b).).

If, $(j-1) \notin V(t)$ and $j \notin V(t)$ then,

$$ni_{i,j,1} = b_{i+1,2(j-1)} - b_{i,2j} \tag{4}$$

If, $(j-1) \notin V(t)$ and $j \in V(t)$ then,

$$ni_{i,j,1} = b_{i+1,2(j-1)} - b_{i,2j} - fo_{i,j,1} \tag{5}$$

If $ni_{i,j,1} > 0$, then NI really exists when E_i leaves R_j and $ni_{i,j,1}$ expresses the time R_j has to wait for E_{i+1}. If $ni_{i,j,1} \leq 0$, then there is no NI and $ni_{i,j,1}$ expresses the available time to avoid it. When both R_j and R_{j+1} belong to a virtual queue, then $ni_{i,j,1}$ has no meaning because E_{i+1} can never cause NI at R_j.

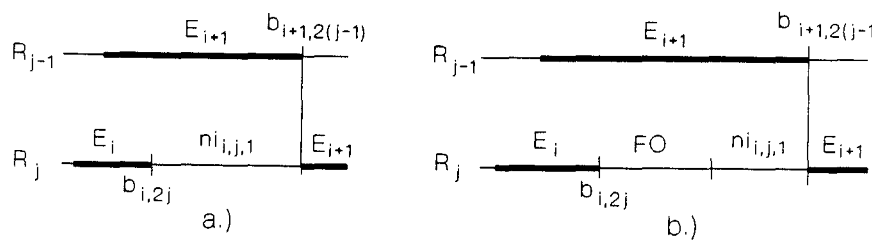

Figure 7.
Calculation of the NI(1) matrix.

Definition 5

The second level no-input matrix (NO(2)) expresses the duration of the NI **caused** by R_j while performing operation on E_i.

The calculation of $ni_{i,j,2}$ is as follows (fig 8(a), 8(b)):

$$ni_{i,j,2} = \begin{cases} b_{i-1,2(j+1)} - b_{i,2j} & \text{if } j+1 \notin V(t) \\ b_{u,2v} - b_{i,2j} & \text{otherwise} \end{cases} \tag{6}$$

If $ni_{i,j,2} < 0$, then NI really exists when E_{i-1} (E_u) leaves R_{j+1} (R_v) and $ni_{i,j,2}$ expresses the time R_{j+1} (R_v) has to wait for E_i (E_{u+1}). If $ni_{i,j,2} \geq 0$, then there is no NI and $ni_{i,j,2}$ expresses the time we have to avoid it.

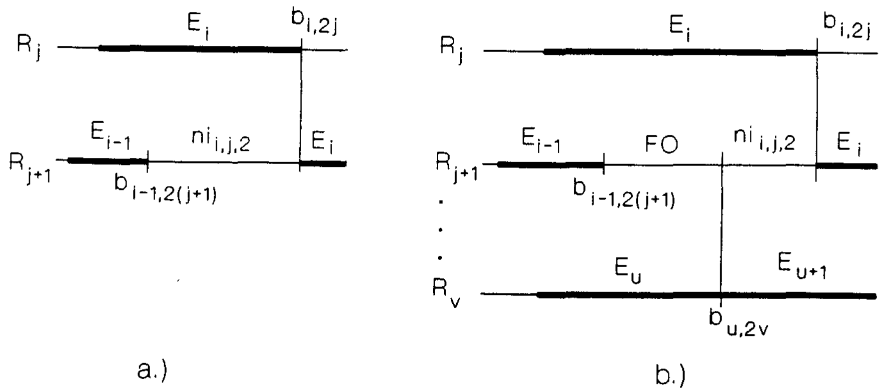

Figure 8.
Calculation of the NI(2) matrix.

5. CALCULATION OF THE VALIDITY RANGE

In case of zero-order perturbation the size of $fo_{i,j,k}$ and $ni_{i,j,k}$ may change if the perturbation appears only at one of the two entities which participate in the calculation of $fo_{i,j,k}$ and $ni_{i,j,k}$. If it appears at both entities then the size of $fo_{i,j,k}$ and $ni_{i,j,k}$ will not change.

Let $\mathbf{Z}$ denote a set of the (i,j) index-pairs belonging to those E_i at R_j which $fo_{i,j,k}$ and $ni_{i,j,k}$ $(k = 1,2)$ change due to the effect of the introduced perturbation.

Theorem 1

Q_j, R_j $j = 1,\ldots,M$ and E_i $i = 1,\ldots,N$ determine a transfer line type queuing network with $c(j)$ queue capacities and FIFO queuing disciplines. A perturbation $(\delta_{x,y})$ is introduced on E_x at R_y, and has an effect of $\delta_{i,j}$ on E_i at R_j. Deterministic similarity holds as long as

$$
\begin{array}{ccc}
\text{MAX}(ni_{i,j,k}, fo_{i,j,k}) \leq & \delta_{x,y} & \leq \text{MIN}(ni_{i,j,k}, fo_{i,j,k}) \\
ni_{i,j,k} \leq 0 & & ni_{i,j,k} > 0 \\
fo_{i,j,k} \leq 0 & & fo_{i,j,k} > 0
\end{array}
\tag{7}
$$

$$
\begin{gathered}
i, x \in \{1,\ldots,N\}; \quad j, y \in \{1,\ldots,M\}; \\
i, j \in \mathbf{Z}; \quad k = 1,2;
\end{gathered}
$$

Proof

Due to the propagation of $\delta_{x,y}$ the resulting $\delta_{i,j}$ may have the following effects on the FO and NI events,

If $\delta_{i,j} > 0$, then while R_j is working on E_i

- FO and NI may disappear after R_j,
- FO may appear on the preceding resources,
- NI may appear on the subsequent resources.

If $\delta_{i,j} < 0$, then while R_j is working on E_i

- FO and NI may appear after R_j,
- FO may disappear on the preceding resources,
- NI may disappear on the subsequent resources.

Identifying these cases in the FO(1), FO(2), NI(1) and NI(2) matrices directly comes (7).

■

6. EXTENSION OF THE PROBLEM

The calculation introduced in the previous sections is valid only in transfer line type networks. It is easy, however, to generalize the method to be valid for general networks (GN). Fig. 9. shows the type of network we want to deal with. We kept the notations of Fig. 1., but completed it with the routing parameter ($r_{i,j}$). There are two basic problem we have to cope with,

- entities may follow different routes in the network,
- entities may overtake other entities.

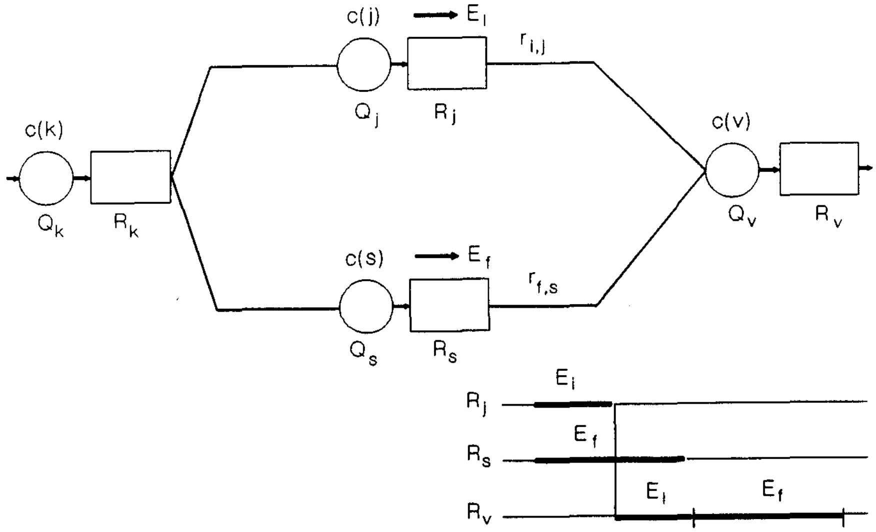

Figure 9.
Illustration of a general queuing network.

6.1. The effect of different routes

The possibility of the different routes makes necessary to revise that simple index identification of entities and resources we introduced at the transfer line type queuing networks.

Since in GN entities may follow different routes, the order of entities at the various resources may be distinct. Let us suppose that E_i is the $i-$th entity on R_j and leaving R_j it goes to R_v. Since R_v can be visited by entities from other resources as well we have no guarantee that E_i will also be the $i-$th at R_v. This phenomena causes the difficulty. In transfer lines we can identify entities by their place in the order at the various resources. At the calculation of the validity range we implicitly used this fact. To cope with GN we have to introduce a general entity identification.

Let $n_{i,j}$ denote the number in the order of E_i at R_j $i = 1, \ldots, N$, $j = 1, \ldots, M$ and $n^*_{k,j}$ $k = 1, \ldots, N$ the identification of the $k-$th entity at R_j.

The $n_{i,j}$ and $n^*_{i,j}$ matrices can be obtained from the simulation output. From the B matrix we can easily derive $n_{i,j}$ if we put in increasing order $b_{i,2j-1}$ or $b_{i,2j}$ $i = 1, \ldots, N$. In a transfer line case $n_{i,j} = i$ and $n^*_{(i,j)}$ identifies E_i.

A similar approach should be applied for the identification of resources. At the transfer line E_i always moves from R_j to R_{j+1}, in GN, however, it may move to any R_s $s = 1, \ldots, M$. The information of the routing in a sample path will be expressed in the routing matrix.

Let $r_{i,j}$ denote the index of the resource E_i enters after completing the service at R_j and $r^*_{i,j}$ the index of the resource E_i left before arriving to R_j.

The $r_{i,j}$ and $r^*_{i,j}$ matrices can be obtained from the simulation output. We can derive $r_{i,j}$ from the B matrix if we put in increasing order $b_{i,2j-1}$ or $b_{i,2j}$ $j = 1, \ldots, M$. The serial of j determines the route of E_i. In the transfer line case $r_{i,j} = j + 1$ and $r^*_{i,j} = j - 1$.

Applying these notations (1), (2), (3), (4), (5) and (6) can be transformed to be valid for general queuing networks. The transformation can be seen in the appendix.

6.2. The effect of overtake

Overtake of entities may occur as an effect of perturbation at assembly type queues. Let's suppose that both E_i from R_j and E_f from R_s goes to R_v and $b_{i,2j} < b_{f,2s}$ (Fig. 9.). As a consequence of the FIFO rule $n_{i,v} < n_{f,v}$, that is, E_i will get to R_v before E_f. If the perturbation changes the relationship of $b_{i,2j}$ and $b_{f,2s}$ the relationship of $n_{i,v}$ and $n_{f,v}$ will also change, violating this way the deterministic similarity. To avoid this, we have to ensure that the perturbation be small enough not to cause overtake of entities.

Let $\mathbf{A}$ denote the index set of the assembly type queues and $v \in \mathbf{A}$. If Q_v visited by E_i from R_j and E_f from R_s, then $ot_{i,f,v}$ will denote the time of E_f for not to be passed by E_i on R_v as a consequence of perturbation.

If $ot_{i,f,v} < 0$, then $n_{f,v} < n_{i,v}$, that is, E_f can overtake E_i. If $ot_{i,f,v} \geq 0$, then $n_{f,v} > n_{i,v}$, that is, E_i can overtake E_f. The overtake matrix is calculated as follows,

(8) $$ot_{i,f,v} = b_{f,2s} - b_{i,2j}$$

Let $\mathbf{W}$ denote a set of the (i, f) index-pairs belonging to those E_i and E_f which $ot_{i,f,v}$ $v \in \mathbf{A}$ changes due to the effect of the introduced perturbation.

From the point of view of the change of order of entities at the resources, deterministic similarity holds as far as the perturbation is small enough not to cause overtake, that is, if

$$
\begin{array}{ccc}
\text{(9)} & \operatorname{MAX}(ot_{i,f,v}) \le \delta_{x,y} \le \operatorname{MIN}(ot_{i,f,v}) \\
& ot_{i,f,v} \le 0 \qquad\qquad\qquad ot_{i,f,v} > 0
\end{array}
$$

6.3. Validity range for general queuing networks

The extension of theorem 1 for general queuing networks is as follows,

Theorem 2

Q_j, R_j $j = 1, \dots, M$ and E_i $i = 1, \dots, N$ determines a general queuing network with $c(j)$ queue capacities and FIFO queuing disciplines. A perturbation $(\delta_{x,y})$ is introduced on E_x at R_y, and has an effect of $\delta_{i,j}$ on E_i at R_j. Deterministic similarity holds as long as

$$
\begin{array}{ll}
& \operatorname{MAX}(ni_{i,j,k}, fo_{i,j,k}, ot_{i,f,v}) \le \delta_{x,y} \le \operatorname{MIN}(ni_{i,j,k}, fo_{i,j,k}, ot_{i,f,v}) \\
& \quad ni_{i,j,k} \le 0 \qquad\qquad\qquad\qquad\qquad ni_{i,j,k} > 0 \\
\text{(10)} & \quad fo_{i,j,k} \le 0 \qquad\qquad\qquad\qquad\qquad fo_{i,j,k} > 0 \\
& \quad ot_{i,f,v} \le 0 \qquad\qquad\qquad\qquad\qquad ot_{i,f,v} > 0 \\
& \qquad i, f, x \in \{1, \dots, N\}; \quad j, v, y \in \{1, \dots, M\}; \\
& \qquad\quad i, j \in \mathbf{Z}; \quad i, f \in \mathbf{W}; \quad k = 1, 2
\end{array}
$$

Proof

Completing the proof of theorem 1 with the limits imposed by the overtake restrictions, we can get (10).

■

7. IMPLEMENTATION OF THE ALGORITHM

Based on theorem 1 the validity range can be calculated by the following algorithm:

1. Various preliminary calculations are performed to obtainthe B^*, R, R^*, N, N^* furthermore the NI, FO, OT matrices and the $\mathbf{A}$ set.
2. The perturbation is propagated and the $\mathbf{Z}$ and $\mathbf{W}$ set is determined.

3. With theorem 1 or 2 the validity range is calculated.

The algorithm is polynomial type but, due to the matrices necessary to the calculation, a careful data organization is required. If N is the number of entities, M is the number of resources and F is the number of assembly type nodes then, the required number of input data is approximately $N(8M + FN) + F$. This formula shows that we have a quadratically increasing data requirement. In case of transfer line the quadratic tag is canceled out ($F = 0$).

As a consequence, in systems where N is high (i.g. in communication systems where the number of arriving entities may easily increase 10^4) data management may slow down the calculation. In manufacturing systems, especially when technology intensive parts are produced in small lots (e.g. FMS), N may not exceed even 10^2 and the algorithm can be organized easily. The computational time is insignificantly small.

The results provided by the algorithm are valid only for a specific sample path. Considering the stochastic similarity, the convergence properties of a single perturbation should be regarded here as well [3].

The algorithm was tested on various examples. We provide the results of a small system similar to the one in Fig. 9. The queue capacities used in the calculation are the following, $c(k) = 1$, $c(j) = 1$, $c(s) = 2$, $c(v) = 1$. 10 entity with identifications from 1 to 10 enter the system at Q_k and may randomly choose between R_j and R_s. They leave the system at R_v. The simulation was performed with the SIMAN simulation language. The identification of entities and the resulting B matrix of a sample path can be seen in Table 1. Note that the change of E_8 and E_7 is not a coincidence. It reflects that the order of entities at the entrance of the system are not the same as at the exit.

Table 1.

The simulation output (B matrix)

ID	BEG_k	END_k	BEG_j	END_j	BEG_s	END_s	BEG_v	END_v
E_1	0.0	3.0	3.0	8.0	___	___	8.0	20.0
E_2	3.0	6.0	___	___	6.0	12.0	20.0	24.0
E_3	6.0	10.0	10.0	17.0	___	___	24.0	30.0
E_4	10.0	14.0	___	___	20.0	25.0	30.0	33.0
E_5	14.0	21.0	24.0	27.0	___	___	33.0	35.0
E_6	24.0	29.0	___	___	30.0	38.0	38.0	43.0
E_8	34.0	39.0	39.0	44.0	___	___	44.0	57.0
E_7	29.0	34.0	___	___	38.0	46.0	57.0	61.0
E_9	39.0	42.0	44.0	47.0	___	___	61.0	66.0
E_{10}	44.0	52.0	___	___	57.0	64.0	66.0	71.0

Table 2.

Results of the calculations of perturbation analysis

Introduction		SLOPE	Validity range results				
RES.	IDEN.		LIMIT	IDEN.	ID/R	TYPE	
R_k	E_1	1.0	$-\infty$	—	—	—	LL
			∞	—	—	—	UL
R_k	E_2	0.0	-2.0	E_3	R_k	FO(1)	LL
			3.0	E_5	R_k	FO(1)	UL
R_k	E_3	0.0	-2.0	E_3	R_k	FO(1)	LL
			3.0	E_5	R_k	FO(1)	UL
R_k	E_4	0.0	$-\infty$	—	—	—	LL
			3.0	E_5	R_k	FO(1)	UL
R_k	E_5	0.0	$-\infty$	—	—	—	LL
			3.0	E_5	R_k	FO(1)	UL
R_k	E_6	1.0	-1.0	E_9	E_7	OT(v)	LL
			1.0	E_6	R_k	NI(2)	UL
R_k	E_7	1.0	-1.0	E_9	E_7	OT(v)	LL
			2.0	E_8	E_7	OT(v)	UL
R_k	E_8	1.0	-1.0	E_9	E_7	OT(v)	LL
			2.0	E_8	E_7	OT(v)	UL
R_k	E_9	0.0	$-\infty$	—	—	—	LL
			2.0	E_9	R_k	FO(1)	UL
R_k	E_{10}	0.0	-14.0	E_{10}	R_k	FO(1)	LL
			5.0	E_{10}	R_k	NI(2)	UL

Since SIMAN can directly communicate with FORTRAN [10] the programs for calculating the validity range were written in this programming language. The results can be seen in Table 2. The effect of 10 perturbation was studied. The perturbations were introduced in R_k at every entity independently. The SLOPE column shows the gradient of the through put time with respect to the operation time of the respective entities at R_k. These slopes are certainly valid only within the limits provided by the LIMIT columns. If the perturbation exceeds these limits the deterministic similarity is hurt, that is, higher order propagation rule should be applied. The IDEN. and ID/R columns of the result section identify the place, while the TYPE column identifies the reason of the limits. For example if we introduce a perturbation at E_6 in R_k then the lower limit will be caused by an overtake between E_9 and E_7 while the upper limit will be the result of a NI caused by E_6 when it is in R_k.

The algorithm was implemented in a practical case as well. The gradient of the throughput time and of the waiting time in queues were calculated for production control purposes in a continuous steel manufacturing process. The approximately 40 entities (steel slabs) and four resources (machines) did not cause any memory management problem and the insignificant computation time allowed on line gradient calculation. For further details about this project see Koltai *et al.* [9].

8. CONCLUSIONS

An algorithm was provided to calculate the validity range of deterministic similarity at the simulation of general queuing networks. The calculation provides information about the validity of the zero order gradient of the throughput time with respect to the operation time of entities at resources. The algorithm requires only the time schedule of the resources (B matrix) which can be provided easily by any simulation language in use today. The information gained by the calculation can be attached to the simulation output, providing useful information for parametric analysis of the throughput time.

The results can also provide basis for the calculation of the validity range of higher order perturbations, for the computation of multi variable gradient information, and for the examination of the gradient information concerning other parameters and performance measures, which are all subject of our further research. We expect that any result in these areas might support the improvement of the planning and dynamic control of discrete event dynamic systems.

9. REFERENCES

[1] **Cao, X.R.** (1988)."Realization Probability in Multi-class Closed Queuing Networks". *European Journal of Operations Research*, **36**, 393–401.

[2] **Gál, T.** (1979). *Postoptimal Analyses, Parametric Programming and Related Topics.* MacGraw-Hill Inc.

[3] **Heidelberg, P., Cao, X.R., Zazanis, M.A.** & **Suri, R.** (1988). "Convergence Properties of Infinitesimal Perturbation Analysis Estimates". *Management Science*, **34**, 1281–1302.

[4] **Ho, Y.C.** (1987). "Performance Evaluation and Perturbation Analysis of Discrete Event Dynamic Systems". *IEEE Transactions on Automatic Control*, **AC-32**, 563–572.

[5] **Ho, Y.C., Cao, X.R.** & **Cassandras, C.** (1983). "Infinitesimal and Finite Perturbation Analysis for Queuing Networks". *Automatica*, **19**, 439–445.

[6] **Ho, Y.C.** & **Cassandras, C.** (1983). "A New Approach to the Analysis of Discrete Event Dynamic Systems". *Automatic*, **19**, 149–167.

[7] **Ho, Y.C.** & **Li, S.** (1988). "Extension of Infinitesimal Perturbation Analysis". *IEEE Transaction on Automatic Control*, **AC-33**, 427–438.

[8] **Ho, Y.C., Suri, R., Cao, X.R., Diehl, G.W., Dille, L.W.** & **Zazanis, M.** (1984). "Optimization of Large Multi-class (non-product-form) Queuing Networks Using Perturbation Analysis". *Large Scale Systems*, **7**, 165–180.

[9] **Koltai, T., Onieva, L.** & **Larrañeta, A.** (1993). "Examination of the Sensitivity of an Operation Schedule with Perturbation Analysis". *International Journal of Production Research*, **31**, 2777–2787.

[10] **Pedgen, D.C., Shannon, R.E.** & **Sadowski, R.P.** (1990). *Introduction to Simulation Using SIMAN*. McGraw-Hill. New York.

[11] **Rubistein, R.Y.** (1986). *Monte Carlo Optimization, Simulation and Sensitivity of Queuing Networks*. John Wiley & Sons. New York.

10. APPENDIX

a) The examination of virtual queue

The transformed version of the algorithm to look for R_v, E_u is presented in Fig. 10, and the algorithm for finding R_p, E_g belonging to R_j, E_i when R_v, E_u is already known, is given in figure 11.

To facilitate further discussion let $b^*_{i,2j-1}$ denote the beginning and $b^*_{i,2j}$ the ending time of the operation of the $i-$th entity that is, entity $n_{i,j}$ on R_j.

```
nr := r_{i,j}                  (next resource)
ec := n*_{n_{i,nr}-c(nr),nr}   (entity for comparison)
fin:= 0
WHILE nr ∈ {1,...,M} AND ec ∈ {1,...,N} AND fin:= 0 DO
          BEGIN
          IF b_{ec,2nr} < b_{i,2j} THEN
             BEGIN
             j := nr
             i := ec
             nr := r_{i,j}
             ec := n*_{n_{i,nr}-c(nr),nr}
             END
          ELSE
             BEGIN
             v := j
             u := i
             fin:= 1
             END
          END
store (v, u)
```

Figure 10.
Pseudo code of the algorithm for searching E_u, R_v in general networks.

```
ec := n*_{n_{i,j}+c(j),j}      (entity for comparison)
pr := r*_{ec,j}                (previous resore)
fin:= 0
WHILE pr ∈ {1,...,M} AND ec ∈ {1,...,N} AND fin:= 0 DO
          BEGIN
          IF b_{ec,2pr} < b_{u,2v} THEN
             BEGIN
             i := ec
             j := pr
             ec := n*_{n_{i,j}+c(j),j}
             pr := r*_{ec,j}
             END
          ELSE
             BEGIN
             g := i
             p := j
             fin:= 1
             END
          END
store (p, g)
```

Figure 11.
Pseudo code of the algorithm for searching E_g, R_p in general networks.

b) Calculation of the FO(1) matrix

Recall the elements in figure 5a, 5b. In addition, let's suppose that $r_{i,j} = s$, that is E_i goes to R_s after completing the operation at R_j.

(1) $$fo_{i,j,1} = \begin{cases} b^*_{n_{i,s}-c(s),2s} - b^*_{n_{i,j},2j} & s \notin V(t) \\ b^*_{n_{u,v},2v} - b^*_{n_{i,j},2j} & \text{otherwise} \end{cases}$$

c) Calculation of the FO(2) matrix

Recall the elements in figure 6a, 6b. In addition, let's suppose that $r^*_{i,j} = s$, that is E_i visited R_s before arriving to R_j.

If, $s \notin V(t)$ then,

(2) $$fo_{i,j,2} = b^*_{n_{i,s}+c(j),2s} - b^*_{n_{i,j},2j}$$

If, $s \in V(t)$ then

(3) $$fo_{i,j,2} = \begin{cases} b^*_{n_{g,p},2p} - b^*_{n_{i,j},2j} & \text{if } b^*_{n_{g,p}+1,2(p-1)} \geq b^*_{u,2v} \\ b^*_{n_{g,p}+1,2p} - b^*_{n_{i,j},2j} & \text{otherwise} \end{cases}$$

d) Calculation of the NI(1) matrix

Recall the elements in figure 7a, 7b. In addition, let's suppose that $r^*_{i,j} = s$, that is E_i visited R_s before arriving to R_j.

If, $s \notin V(t)$ and $j \notin V(t)$ then,

(4) $$ni_{i,j,1} = b^*_{n_{i,s}+1,2s} - b^*_{n_{i,j},2j}$$

If, $s \notin V(t)$ and $j \in V(t)$ then,

(5) $$ni_{i,j,1} = b^*_{n_{i,s}+1,2s} - b^*_{n_{i,j},2j} - fo_{i,j,1}$$

If, $s \in V(t)$ then $ni_{i,j,1}$ has no meaning.

e) Calculation of the NI(2) matrix

Recall the elements in figure 8a, 8b. In addition, let's suppose that $r_{i,j} = s$, that is E_i goes to R_s after completing the operation at R_j.

(6) $$ni_{i,j,2} = \begin{cases} b^*_{n_{i,s}-1,2s} - b^*_{n_{i,j},2j} & s \notin V(t) \\ b^*_{n_{u,v},2v} - b^*_{n_{i,j},2j} & \text{otherwise} \end{cases}$$

Estadística Oficial

EUROSTAT, LA OFICINA ESTADÍSTICA DE LA COMISIÓN EUROPEA

FERNANDO DE ESTEBAN ALONSO*

La creación y el desarrollo de Eurostat están íntimamente ligados a la historia y a los avances de la integración europea. En un principio, con el fin de asistir a la gestión supranacional del carbón y el acero dentro del marco de la Comunidad Europea del Carbón y del Acero (CECA), la Alta Autoridad creó la "Oficina de Estadística". Esta obtenía, directamente de los obligados a suministrar información estadística de los seis Estados miembros, los datos necesarios para la actuación de la CECA.
Desde entonces, los métodos de trabajo han cambiado radicalmente. En aquella época, por ejemplo, se trataba de llegar a "un acuerdo entre caballeros" de las partes en el Tratado; hoy día, debido a la complejidad creciente de la economía, se trata sistemáticamente de establecer un fundamento jurídico legal para institucionalizar los acuerdos y compromisos. En este sentido, Eurostat y sus interlocutores de los institutos nacionales de estadística forman el sistema estadístico comunitario, organizado con arreglo al principio de subsidiariedad.

1. LAS MISIONES DE EUROSTAT

1.1. Estadísticas y democracia

El objetivo primario de la estadística es servir al proceso de toma de decisiones, integrarse en él a fin de alcanzar un óptimo de eficacia en las decisiones

* Fernando de Esteban Alonso. Director de Difusión y Relaciones Públicas, Informática Estadística y Relaciones con los paises de África, Caribe y Pacífico y paises en vía de desarrollo.
–Article rebut el gener de 1994.
–Acceptat l'abril de 1994.

(reduciendo la incertidumbre sobre sus consecuencias), y permitir una evaluación de los resultados. La estadística, instrumento de evaluación y de transparencia política, permitirá a los ciudadanos y a todos los agentes de la vida política, económica y social, juzgar las actuaciones realizadas.

Así, el trabajo estadístico tiene por objetivo responder a una política determinada, desde su concepción inicial a su valoración final.

Para ello, la estadística ha de ser científica, independiente y de calidad.

El carácter científico de los trabajos queda garantizado mediante una colaboración intensa y permanente entre Eurostat, las organizaciones internacionales y los institutos nacionales de estadística. Por otro lado, Eurostat recurre a expertos independientes (universitarios, investigadores, consultores, etc.).

La independencia de Eurostat queda garantizada por su estatuto de dirección general de la Comisión, garante del interés comunitario.

La calidad del trabajo de Eurostat queda garantizada ante todo por la cientificidad y la independencia. Por otro lado, los obligados a suministrar información estadística (empresas, etc.) tienen la obligación de responder (a cambio se les garantiza una estricta confidencialidad de los datos individuales).

1.2. Las cuatro misiones de Eurostat

Para alcanzar este objetivo general, Eurostat organiza sus tareas en torno a cuatro misiones:

- *Proporcionar a la Comisión información estadística*: el campo estadístico que cubre Eurostat es el de las líneas políticas comunes, como consecuencia de determinadas opciones políticas. La concepción, elaboración y gestión de las líneas políticas comunes se basan en datos objetivos y de calidad. En este sentido, Eurostat es un servicio de información para la Comisión Europea, pero también para el conjunto de las instituciones europeas (Parlamento Europeo, etc.).
- *Construir el sistema estadístico comunitario*: las líneas políticas comunes deben basarse en una información estadística armonizada entre los doce Estados miembros de la Unión Europea. Con ese fin, Eurostat, en estrecha colaboración con el conjunto de los institutos nacionales de estadística, va creando un lenguaje estadístico común que comprende normas y definiciones armonizadas, clasificaciones comunes (NACE: Nomenclatura de Actividades Económicas de la Comunidad Europea, NUTS: Nomenclatura

de Unidades Territoriales Estadísticas, etc.) y métodos armonizados para la elaboración de encuestas (Encuesta sobre las fuerzas de trabajo, etc.).

- *Organizar el acceso a la información estadística*: para que dicha información desempeñe el papel que le corresponde en un sistema democrático, debe llegar a los usuarios (económicos, políticos, sindicales, etc.). Eurostat utiliza cuatro soportes para la difusión:
  - Las publicaciones sobre papel (más de cien títulos al año).
  - Las bases de datos (comercio exterior, regiones, sociales, etc.).
  - Los soportes electrónicos (CD-ROM, disquetes informáticos, etc.).
  - Los servicios de información de la Comisión (redes de información tales como los Euro Info Centros, los CDE-DEP o centros de documentación europea, etc.).
- *Cooperar técnicamente*:
  - Con los interlocutores mundiales que son las organizaciones internacionales (AELC, Naciones Unidas, FMI, etc.) y los países con un sistema estadístico avanzado (Estados Unidos, Canadá, etc.). El nuevo sistema de cuentas nacionales constituye el ejemplo más reciente de dicha cooperación.
  - Con los países que desean desarrollar un sistema estadístico que responda a las mismas exigencias del sistema comunitario. Se trata sobre todo de los países de Europa Central y Oriental (en el marco del programa PHARE) y de los países signatarios del Convenio de Lomé.

2. LA ORGANIZACIÓN DEL SISTEMA ESTADÍSTICO COMUNITARIO

La organización del sistema estadístico comunitario se basa en el principio de subsidiariedad.

2.1. El principio de subsidiariedad

El Tratado de Maastricht define este principio en su artículo 3 B:

> "La Comunidad actuará dentro de los límites de las competencias que le atribuye el presente Tratado y de los objetivos que éste le

asigna. En los ámbitos que no sean de su competencia exclusiva, la Comunidad intervendrá, conforme al principio de subsidiariedad, sólo en la medida en que los objetivos de la acción pretendida no puedan ser alcanzados de manera suficiente por los Estados miembros, y, por consiguiente, puedan lograrse mejor, debido a la dimensión o a los efectos de la acción contemplada, a nivel comunitario."

Los ámbitos de competencia de Eurostat quedan acotados con arreglo a este criterio: son, en definitiva, las políticas o actuaciones comunitarias.

El sistema estadístico comunitario comprende doce institutos nacionales de estadística más Eurostat. Cada instituto nacional de estadística tiene su propia cultura (estructura centralizada o federal, etc.). Es necesario, por ello, alcanzar compromisos entre todas las partes y organizar las diferentes fases del trabajo estadístico sobre principios admitidos por todos.

2.2. La división del trabajo estadístico dentro del sistema estadístico comunitario

Sentado ese principio, los Estados miembros y Eurostat se reparten las tareas del modo siguiente:

Los estados miembros:

- Participan en la definición de proyectos.
- Recogen, verifican y hacen un tratamiento previo de la información nacional.
- Transmiten dicha información a Eurostat.

Eurostat por su parte:

- Coordina los proyectos comunitarios.
- Verifica la información recibida.
- Procesa y da forma a la información.
- Divulga la información estadística.

2.3. Los instrumentos

El Programa estadístico (que cubre actualmente el período 1993-1997) es la traducción de las necesidades en proyectos. Dicho programa se elabora en estrecha colaboración con el conjunto de los intelocutores (para una presentación completa véase Sigma 1/1993).

Eurostat cuenta en su trabajo con la asistencia de una decena de comités (Comité del PNB, Comité del Secreto Estadístico, Comité de Comercio Exterior, etc).

A un nivel inferior se encuentran los grupos de trabajo (cerca de 80), que sirven de enlace permanente entre los Estados miembros y Eurostat.

2.4. El proceso de toma de decisiones

El proceso de toma de decisiones pasa por varias etapas.

A petición de la Comisión, Eurostat realiza un análisis de viabilidad de la demanda estadística. Dicho análisis se lleva a cabo en colaboración con los servicios estadísticos de diversas organizaciones internacionales y con expertos.

A continuación, las propuestas de Eurostat son estudiadas por el Comité del Programa Estadístico (CPE), integrado por los directores generales de los institutos nacionales de estadística (DGINE), que componen con Eurostat el sistema estadístico comunitario.

Tras recibir la conformidad del CPE, las propuestas son presentadas por la Comisión al Parlamento Europeo.

Paralelamente, esas mismas propuestas se someten al Consejo para que éste decida al respecto. Las propuestas se convierten en actos jurídicos (directivas o reglamentos).

Finalmente dichos actos jurídicos se trasponen a las legislaciones nacionales.

A partir de aquí, los institutos nacionales de estadística pueden proceder a la recogida de datos para transmitirlos seguidamente a Eurostat.

3. LAS REALIZACIONES DE EUROSTAT

El problema de toda organización es saber qué ha de producir, cómo y para quién. Eurostat trata de responder a la demanda creciente de información diversificando su producción.

3.1. Informar al ciudadano

Eurostat está en primer lugar al servicio de los órganos de la Comisión Europea encargados de las políticas comunes. Para hacer frente a esa demanda se producen con regularidad datos detallados sobre precios, producción agrícola, transportes, cuestiones sociales, siderurgia, etc.

Esta información es excesivamente compleja como para hacerla llegar directamente a los ciudadanos europeos. Con el fin de obviar este obstáculo, Eurostat empezó hace unos años a elaborar una serie de productos destinados al "gran público", concretamente *Europa en cifras, Retrato social o Retrato de las regiones.*

Por otro lado, Eurostat publica también regularmente un catálogo del conjunto de sus productos.

3.2. Las nuevas tecnologías: CD-ROM, redes transeuropeas

La estadística no escapa al progreso tecnológico de la información. Esos avances permitirán a Eurostat diversificar y mejorar sus productos.

Así, Eurostat produce actualmente dos CD-ROM. El primero lleva por título COMEXT y presenta, con periodicidad mensual, los intercambios comerciales de los Estados miembros de la Unión entre ellos y con el resto del mundo. La nomenclatura propuesta consta de ocho dígitos. El segundo CD-ROM (bienal) puede considerarse como un anuario electrónico, y presenta el conjunto de los ámbitos estadísticos de Eurostat (cuentas nacionales, economía regional, demografía, medio ambiente, etc.). Estos dos CD-ROM se han desarrollado buscando ante todo la facilidad de uso para quien no posea formación informática.

La importancia de las redes transeuropeas, especialmente las redes telemáticas para la transmisión de información, acaba de ser reafirmada en el *Libro blanco sobre el crecimiento, la competitividad y el empleo*, presentado por la Comisión en diciembre de 1993. Dichas redes ponen las condiciones para la creación de un "espacio informático común", de gran importancia para la realización del mercado único y la competitividad de las empresas europeas. El objetivo perseguido en el campo de las aplicaciones estadísticas es lograr, mediante el desarrollo de normas informáticas comunes, la conexión entre todos los ordenadores europeos para facilitar la circulación de la información.

Dicha circulación responde a una triple exigencia:

- Recoger más datos, dado que asistimos a un fuerte incremento de la demanda de información estadística.

- Reducir la carga sobre los responsables de facilitar la información, a saber, las empresas para las que la obligación de responder representa un coste en recursos personales y en tiempo.
- Ofrecer unos rendimientos informativos más rápidos y específicos.

Eurostat participa muy activamente en la creación de las mencionadas redes, especialmente en los proyectos siguientes:

- EDICOM (Electronic Data Interchange in Commerce), que permitirá una recogida informática de los datos sobre intercambios comerciales intracomunitarios.
- DSIS (Distributed Statistical Information Services), que versa, por ejemplo, sobre la normalización de los registros, el secreto estadístico y las normas relativas a la circulación de la informática de la información.
- SERT (Statistiques d'Entreprises et Réseaux Télématiques), cuyo objetivo es aumentar la competitividad de las empresas europeas reduciendo su carga administrativa como responsable del suministro de la información estadística, ofreciéndoles a la vez una información estadística más centrada en sus necesidades.

Estos proyectos son el resultado de la cooperación entre organizaciones internacionales, pero también con países no pertenecientes a la Unión Europea.

3.3. El nuevo sistema de cuentas nacionales (SCN)

El marco contable utilizado para los análisis económicos y estadísticos no se había modificado desde 1968. El nuevo SCN tiene en cuenta los cambios económicos, sociales y políticos ocurridos desde entonces. Permite integrar, por ejemplo, las cuentas de patrimonio, los servicios prestados por las administraciones públicas (v.g.: enseñanza, sanidad), las nuevas actividades económicas (arrendamiento con opción de compra, nuevos instrumentos financieros) y los costes ambientales.

El nuevo SCN puede calificarse de sistema mundial, pues su flexibilidad permite aplicarlo a los países en vías de desarrollo y a los países en transición hacia una economía de mercado (cuyas cuentas nacionales se basaban en la noción de "producto material").

El nuevo SCN facilitará la rápida obtención, la calidad y la comparabilidad de los datos. Los primeros datos (relativos a 1995) se prevén para 1997.

Dentro de la Unión Europea, el Tratado de Maastrich sobre la Unión Económica y Monetaria hace aún más necesario disponer de estadísticas con mayor grado comparabilidad.

4. LOS NUEVOS RETOS

4.1. Maastrich y sus exigencias en materia estadística

La "ardiente obligación" que constituye la aplicación del nuevo SCN es una de las condiciones de éxito de la UE. Las cuentas nacionales, a través de los criterios de convergencia, servirán para medir las condiciones de acceso a la Unión Económica y Monetaria. Así, las estadísticas constituirán en lo sucesivo una base para la toma de decisiones. Anteriormente, las cuentas nacionales se utilizaban ya para determinar la contribución de los Estados miembros al presupuesto comunitario (mediante el PIB, que sirve de base para calcular las contribuciones de los Estados miembros) y su derecho a la percepción de partidas de los fondos estructurales (FEDER, FES), creados para garantizar la cohesión económica y social de la Comunidad.

Para responder a las necesidades específicas de la UE, Eurostat y los Estados miembros "Sistema europeo revisado de cuentas económicas" (SEC), compatible con el SCN, pero más detallado y preciso, a fin de asegurar la armonización de las características peculiares de la UE.

Por otro lado, el Tratado de la Unión ha conferido nuevas responsabilidades a la Comisión Europea, especialmente en materia de medio ambiente (ya contempladas en el Acta Única, pero reforzadas en el Tratado), formación, educación, sanidad pública, etc.

Con la introducción de ciertos ámbitos de competencia y el refuerzo de otros, Eurostat deberá desarrollar un marco estadístico para conocer la situación existente en cada Estado miembro, para preparar a continuación una serie de actuaciones comunes.

4.2. La ampliación

La UE ha llevado, paralelamente a los debates sobre su profundización, negociaciones sobre la ampliación con cuatro países candidatos (Austria, Finlandia, Noruega y Suecia).

Dichos países habrán de adaptar sus sistemas estadísticos al sistema estadístico comunitario. Este movimiento de aproximación se ha iniciado ya en el marco del Acuerdo sobre el Espacio Económico Europeo (EEE) (organización que agrupa a la CE y a los países de la AELC, excepto Suiza). Los países del EEE participan en las labores de los grupos de trabajo y de los comités.

En caso de adhesión, los citados países deberán acelerar la armonización de los métodos y conceptos estadísticos y la adaptación a las nomenclaturas de Eurostat. Los institutos nacionales de estadística de esos países están representados entre el personal de Eurostat a fin de ir familiarizándose con los métodos de trabajo de Eurostat.

EL PROGRAMA ANUAL D'ACTUACIÓ ESTADÍSTICA

MONTSERRAT BIOSCA

Institut d'Estadística de Catalunya

En el marc de la legislació estadística de Catalunya, la programació de l'activitat estadística oficial es regula per unes normes generals i unes altres d'específiques que es concreten per a cada activitat, i es descriu de forma homogènia en les fitxes incloses en el Programa anual. A més d'aquests aspectes, l'article tracta dels trets més significatius que es deriven de l'anàlisi global del Programa.

Annual Programme of Statistical Actuation.

Key words: Òrgan gestor del sistema estadístic, planificació estadística, programació estadística, descripció normalitzada activitats estadístiques, normes generals, normes específiques.

A. L'INSTITUT D'ESTADÍSTICA DE CATALUNYA, GESTOR DEL SISTEMA ESTADÍSTIC

El sistema estadístic de Catalunya es pot definir com un conjunt ordenat i harmònic de mètodes, procediments i resultats estadístics dels diferents agents institucionals que participen en l'elaboració de l'estadística oficial. Aquest sistema és resultant d'un procés on s'ha anat precisant de forma progressiva el fonament i el contingut de l'activitat estadística, l'àmbit d'aplicació, les competències, requisits i procediments, la vertebració del sistema, la forma d'execució, l'aprovació de resultats, les interrelacions institucionals i altres aspectes bàsics, com queda reflectit en la legislació vigent en matèria estadística.

–Article rebut el febrer de 1994.

–Acceptat el maig de 1994.

Entre aquests últims aspectes, un de particularment substancial és el que estableix la titularitat de l'Institut d'Estadística de Catalunya com a organisme responsable del sistema estadístic i d'altra banda regula la forma d'exercir aquesta funció.

La Llei 14/1987, de 9 de juliol, d'estadística, en l'article 44, autoritza el Govern de la Generalitat a crear l'Institut d'Estadística de Catalunya. El Decret 341/1989, d'11 de desembre, estableix la creació de l'Institut d'Estadística de Catalunya com a òrgan estadístic de la Generalitat de Catalunya i, al mateix temps, el Consell Català d'Estadística com a òrgan col·legiat de consulta i d'assessorament. Així, l'Institut d'Estadística de Catalunya coordina el sistema estadístic perquè és l'òrgan de la Generalitat sobre el qual la Llei estableix l'obligatorietat d'aconseguir les estadístiques que faciliten el coneixement de la realitat catalana per a la presa de decisions.

Un aspecte substantiu de la Llei d'estadística, és el que determina, en l'article 6, que l'activitat estadística haurà de ser planificada i aprovada per Llei, obrint camí a l'elaboració de la proposta de Llei del pla estadístic. Complementàriament, l'article 10 obliga el Govern de la Generalitat a aprovar, a proposta del conseller d'Economia i Finances, un programa d'actuació estadística acomodat al Pla estadístic de Catalunya i del qual s'ha de donar compte al Parlament.

La Llei del pla estadístic que fou aprovada el 13 de desembre de 1991 per al període 1992-1995, estableix explícitament que l'Institut d'Estadística de Catalunya és l'organisme responsable de dur a terme aquest Pla i els programes anuals d'actuació estadística, directament o en col·laboració amb altres entitats públiques o privades, i per últim, en l'article 29 disposa que l'aprovació del programa anual d'actuació estadística per a l'any següent es farà per Decret del Govern de la Generalitat i n'estableix el procediment d'elaboració. En aplicació d'aquesta normativa, l'Institut d'Estadística de Catalunya elaborà l'esborrany del primer Pla estadístic referit al període 1992-1995 i porta fetes les propostes dels programes anuals d'actuació estadística per als anys 1992, 1993 i 1994.

El procediment per planificar en primer lloc i després programar l'activitat estadística oficial de Catalunya es fonamenta en un procés sistemàtic de consulta a les diferents institucions que integren el sistema estadístic de Catalunya. Les consultes fetes per l'Institut d'Estadística de Catalunya als departaments de la Generalitat i altres institucions van donar com a resultat la concreció i la definició de les activitats estadístiques existents i alhora el coneixement de les necessitats en aquesta matèria. Això va permetre configurar els objectius a assolir pel Pla estadístic i la seva concreció en les activitats estadístiques especificades en els programes anuals d'actuació estadística.

Un aspecte rellevant a destacar fou la participació dels interlocutors tècnics de cada departament, que eren tècnics nomenats pel seu secretari general, als quals s'encomanava la important tasca de fer arribar a l'Institut la informació sobre les activitats en curs i les necessitats estadístiques, per a l'elaboració en primer lloc de la Llei del pla estadístic i posteriorment els programes. Per altra banda i d'una manera periòdica els interlocutors tècnics notifiquen la situació dels treballs, per tal d'informar a la Junta de Govern i iniciar així el procediment d'aprovació dels resultats. Aquest procediment consisteix en l'aprovació, per part de la Junta de Govern de l'Institut, d'un informe on es descriu l'estat de situació de les activitats pel que fa al nivell d'acompliment, la publicació o la disponibilitat de les dades i la seva referència temporal. El procés legal d'aprovació de resultats s'inicia amb les resolucions de cadascun dels consellers dels departaments responsables de l'execució de les activitats estadístiques programades. Pel que fa al Programa de 1992, primer del Pla vigent, aquestes resolucions han estat publicades al DOGC i se n'ha fet el comunicat preceptiu al Parlament de Catalunya.

Per últim i des de la perspectiva de l'Institut com a òrgan gestor del sistema estadístic, cal fer esment també de les següents funcions. En primer lloc la d'elaborar i promoure l'aprovació dels projectes de normatives tècniques i prestar els serveis d'assistència tècnica als organismes inclosos en el sistema estadístic. En segon lloc verificar l'acompliment de la legalitat estadística i, especialment, vetllar per l'acompliment de les obligacions en matèria de secret estadístic.

B. DESCRIPCIÓ NORMALITZADA DE LES ACTIVITATS ESTADÍSTIQUES PROGRAMADES

La Llei del pla defineix els objectius a assolir en matèria estadística. En primer lloc estableix l'objectiu central, eix del sistema estadístic. Per al desplegament d'aquest objectiu, ressenyat en l'article 8 de la Llei, s'estableixen els objectius operatius, que són la definició més precisa per identificar de forma normalitzada les activitats estadístiques a assumir.

En el Programa anual d'actuació estadística es fa constar de forma precisa la descripció de les activitats estadístiques que, en compliment dels preceptes de la Llei del pla estadístic de Catalunya 1992-1995, han de fer-se durant l'any.

Les activitats estadístiques programades es regulen per normes generals, les quals regeixen totes les activitats estadístiques incloses en el Programa, i per normes específiques que s'apliquen a una activitat concreta.

Les normes generals establertes per a totes les activitats són:

- El compliment d'algun objectiu operatiu del Pla estadístic de Catalunya.
- L'elaboració d'un projecte tècnic o esquema metodològic que garanteixi la correcció tècnica i la legalitat de l'activitat estadística.
- La garantia de comparabilitat dels resultats numèrics i la compatibilitat de mètodes i normes tècniques amb altres de similars de Catalunya i de l'entorn estatal i europeu.
- L'obligació de mantenir el secret estadístic en els termes establerts per la Llei d'estadística en qualsevol activitat estadística que requereixi la utilització de dades individualitzades.
- El seguiment del procediment legalment establert per a l'aprovació o homologació dels resultats obtinguts per l'activitat.
- La publicitat dels resultats i productes de les activitats estadístiques i la garantia de la seva disponibilitat i periodicitat.
- La presentació, per part del personal que intervé en les operacions de camp, de la corresponent acreditació oficial, encara que aquestes tinguin caràcter experimental.
- La comunicació als subministradors d'informació de les finalitats i les característiques de la investigació estadística, dels seus drets i obligacions i de l'existència, si s'escau, d'infraccions sancionables.

Pel que fa a les normes específiques, per a cada activitat estadística, es tenen en compte els aspectes següents:

- L'explicitació de l'objectiu operatiu del Pla estadístic a què fa referència l'activitat.
- La designació de l'organisme responsable d'executar l'activitat i dels altres ens, la participació dels quals pugui ser necessària en algunes fases de l'execució.
- La formulació de les característiques tècniques de l'activitat a dur a terme.
- L'establiment de les persones físiques o jurídiques i els ens de dret públic obligats a subministrar la informació requerida.
- La determinació de la forma i els terminis en què ha de ser subministrada la informació.

- La determinació de la forma i els terminis en què els resultats estaran disponibles per a la seva difusió.
- La designació de l'organisme obligat a difondre els resultats de l'activitat estadística.

La descripció normalitzada de cada activitat estadística està contemplada en la fitxa d'activitat estadística i consisteix en la designació de les seves característiques tècniques i de les normes reguladores específiques.

La descripció normalitzada de la fitxa d'activitat estadística té els apartats següents:

1. Identificació de l'activitat.
2. Descripció de l'activitat.
3. Difusió de resultats estadístics.
4. Cost directe estimat.
5. Base legal.

1. Pel que fa a la identificació de l'activitat, es fa constar el seu nom, l'àrea temàtica, l'objectiu de la Llei al qual serveix, l'organisme responsable de la seva execució i els organismes públics o privats que hi col·laboren. La descripció d'aquestes característiques és:
   - Nom de l'activitat: denominació més usual en l'àmbit de l'estadística oficial.
   - Àrea temàtica: referència sintètica de la matèria tractada per l'activitat estadística.
   - Objectiu: transcripció de l'objectiu operatiu de la Llei 30/91, de 13 de desembre, del pla estadístic de Catalunya 1992-1995, en acompliment del qual es fa l'activitat estadística.
   - Organisme responsable: ens públic que té l'obligació d'executar l'activitat estadística en el nivell especificat.
   - Organismes col·laboradors: organismes públics o privats que participen en algunes fases de l'activitat o que tenen l'obligació per conveni o per disposició legal de subministrar la informació necessària.

2. Quant a la descripció de les activitats estadístiques es consideren els següents camps:

   - Ressenya: procés tècnic que s'ha de seguir per tal d'obtenir els resultats que es pretenen.
   - Font d'informació: procedència directa de les dades que han de ser tractades.
   - Unitat investigada: referent principal de les dades de l'activitat estadística.
   - Variables: característiques quantitatives i qualitatives, els valors de les quals s'han de determinar.
   - Informant inicial: persona o entitat que subministra les dades primàries.
   - Tècnica de recollida de les dades primàries: procediment que s'ha utilitzat, o s'utilitzarà, per tal d'obtenir les dades primàries.
   - Termini per informar: termini legal de què disposa l'informant inicial per contestar quan l'activitat requereix operacions de camp.
   - Nivell d'execució el 199..: compromís de l'organisme responsable respecte del grau d'acompliment de l'activitat.
   - Termini de disponibilitat: estimació del temps que ha de passar entre l'inici efectiu de l'activitat i la disponibilitat dels primers resultats. S'utilitza la classificació següent:
     - ▶ Disponibilitat immediata: quan l'activitat està produint resultats de manera continuada.
     - ▶ Disponibilitat a curt termini: si s'estima que el període de temps és inferior a un any.
     - ▶ Disponibilitat a mitjà termini: si s'estima que el període de temps és superior a un any i inferior a dos.
     - ▶ Disponibilitat a llarg termini: si s'estima que el període de temps és superior a dos anys.

3. En referència a la difusió dels resultats estadístics es tenen en compte els aspectes següents:

   - Organisme difusor: ens públic al qual s'assigna l'obligació de difondre els resultats.
   - Mitjà de difusió: principal suport a través del qual s'han de fer públics o accessibles els resultats obtinguts.
   - Periodicitat: freqüència amb la qual es farà la difusió.

- Nivell de desagregació territorial: unitats de màxima desagregació territorial dels resultats, referides al mitjà principal de difusió.

4. El cost directe estimat es defineix com el conjunt de recursos imputats directament a l'execució de l'activitat estadística i a la difusió de resultats. S'estableixen els següents intervals de cost:

   - Cost molt baix: si s'estima que el cost anual d'execució de l'activitat és fins a 10 milions de pessetes corrents.
   - Cost baix: si s'estima que el cost anual d'execució de l'activitat és de més de 10 milions i fins a 20 milions de pessetes corrents.
   - Cost moderat: si s'estima que el cost anual d'execució de l'activitat és de més de 20 milions i fins a 40 milions de pessetes corrents.
   - Cost alt: si s'estima que el cost anual d'execució de l'activitat és de més de 40 milions i fins a 70 milions de pessetes corrents.
   - Cost molt alt: si s'estima que el cost anual d'execució de l'activitat és de més de 70 milions i fins a 100 milions de pessetes corrents.
   - Cost extraordinari: si s'estima que el cost anual d'execució de l'activitat és de més de 100 milions de pessetes corrents.

5. El fonament legal de l'activitat estadística fa referència, com a mínim, a alguns dels objectius operatius que estableix la Llei del pla estadístic.

Com s'ha dit, el programa s'aprova per Decret i, naturalment, es publica al DOGC, el que fa que anualment s'editi un voluminós DOGC on es descriuen detalladament el que la Generalitat i les institucions que col·laboren en el camp estadístic es proposen dur a terme. Fins ara s'han publicat tres DOGC de l'11.6.92, 14.1.93 i 8.2.94 (núms. 1.605, 1 .694 i 1.857 respectivament).

C. CONTINGUT I ANÀLISI DEL PROGRAMA ANUAL D'ACTUACIÓ ESTADÍSTICA

L'estudi dels programes anuals ens porta a assenyalar alguns aspectes particularment significatius. Bàsicament ens centrarem en el Programa anual vigent, la qual cosa no implica que en casos determinats s'hagin d'examinar els programes anteriors.

1. Descentralització

L'Institut d'Estadística de Catalunya té com a funció primordial, com s'ha vist, coordinar i gestionar la globalitat del sistema estadístic, a més de tenir responsabilitat directa en l'elaboració de les activitats estadístiques pròpies. En aquesta elaboració l'Institut d'Estadística de Catalunya assumeix directament les activitats instrumentals, és a dir, aquelles que no són de producció però que són necessàries per garantir la fiabilitat, la comparabilitat, la preservació del secret estadístic, etc., de les dades produïdes. Així mateix, actua de forma subsidiària en les activitats de producció estadística en les quals tenen primacia les competències sectorials dels departaments. Es pot dir que l'Institut d'Estadística de Catalunya assumeix directament activitats de caràcter polivalent, les que són metodològicament complexes a nivell estadístic, les de síntesi i les generades per conveni amb altres institucions estadístiques com l'Institut Nacional d'Estadística (INE) per raons de secret estadístic. En aquestes situacions es troben les estadístiques demogràfiques, les resultants del cens agrari i els comptes econòmics de Catalunya, entre altres.

L'anàlisi del Programa anual d'actuació estadística de 1994 dóna la següent distribució de responsabilitats dels diferents organismes en l'execució de les estadístiques:

Organismes responsables	Activitats estadístiques	%
Departaments sense IEC	126	67.4
IEC	48	25.6
Total Generalitat	174	93.0
Entitats territorials	13	7.0
TOTAL	187	100

Una quarta part de les actuacions estadístiques previstes per a 1994 seran dutes a terme directament per l'Institut. Si a més deduïm les activitats instrumentals podem indicar que el nivell de participació de l'Institut en la producció estadística és d'un 21% sobre la totalitat, la qual cosa no fa més que confirmar el caràcter de model descentralitzat del sistema estadístic català. En l'annex 1 s'especifica per a cada organisme la relació nominal de les activitats estadístiques a dur a terme.

2. Abast temàtic

Les activitats estadístiques s'han classificat, en un primer nivell, en set àrees temàtiques. L'aplicació d'aquesta classificació a les activitats estadístiques del Programa anual d'actuació estadística per a 1994 dóna la distribució següent:

Activitats	Nombre	%
Econòmiques	53	28.3
Demogràfiques	11	5.9
Socials	81	43.3
Activitats administratives estrictes	16	8.5
Multidisciplinàries	14	7.5
Instrumentals legals	9	4.8
Instrumentals per a l'optimització del sistema estadístic	3	1.6
TOTAL	187	100

En un segon nivell d'ordenació s'han diferenciat les estadístiques econòmiques o demogràfiques de síntesi, és a dir, les que tenen un caràcter polivalent de les de caràcter específic. En l'annex 2 es presenta el quadre amb aquesta desagregació més àmplia.

3. Consolidació

Una de les característiques considerades en la descripció de l'activitat estadística fa referència al nivell d'execució previst per l'organisme responsable en el desenvolupament de l'activitat programada. En aquest sentit es diferencien les que estan en curs i les noves. Per a aquestes últimes es preveuen diverses alternatives: el projecte, el projecte i la primera execució o bé directament la primera execució perquè ja existeix el projecte.

L'evolució d'aquesta característica ens permet ressaltar la consolidació del sistema estadístic al llarg d'aquests tres anys. Com es pot veure en el quadre adjunt hi ha un augment de les activitats ja en curs i un desplaçament de les activitats previstes en el projecte cap a una primera execució, situació indicadora de la consolidació de les activitats a elaborar.

Nivell d'execució de les activitats estadístiques en els programes anuals d'actuació estadística de 1992, 1993 i 1994

	PROJECTE	PROJECTE I PRIMERA EXECUCIÓ	PRIMERA EXECUCIÓ	EN CURS	TOTAL
TOTAL PROGRAMA 1992	35	11	11	112	169
TOTAL PROGRAMA 1993	8	12	20	142	182
TOTAL PROGRAMA 1994	2	2	21	162	187

La consolidació del sistema es posa també de manifest en la implantació progressiva de nous objectius en el Pla estadístic. L'any 1992 es van abordar 143 objectius, els anys 1992 i 1993 ho van ser 153 i per als tres anys programats es preveu abordar 160 objectius. S'ha de matisar que abordar un objectiu no implica elaborar una sola activitat estadística. Així diverses activitats estadístiques poden preveure aspectes complementaris d'un mateix objectiu.

Cal esmentar també altres aspectes que contribueixen a la consolidació del sistema. En particular, l'aprovació de normes tècniques reguladores del sistema estadístic, i també, molt especialment, la incorporació de les entitats catalanes de caràcter territorial al sistema estadístic. Així, progressivament, s'ha anat incorporant al Program a, d'acord amb el que preveuen els articles 24 i 25 de la Llei del pla estadístic, les entitats territorials de Catalunya que estableixen convenis de col·laboració amb la Generalitat per tal que les seves activitats estadístiques esdevinguin oficials. En el Programa de 1994 s'han signat convenis amb tretze ajuntaments, deu consells comarcals i amb les quatre diputacions catalanes.

4. Fonts d'informació alternatives

Per últim, farem esment d'un aspecte al·ludit en l'objectiu central del Pla estadístic sobre la forma de treballar en la realització de les activitats estadístiques. Així, l'objectiu central del Pla estadístic de Catalunya 1992-1995 és aconseguir un conjunt coherent, fiable i actualitzat de dades estadístiques al mínim cost possible, aprofitant al màxim les fonts existents, que permeti el coneixement de la realitat econòmica, demogràfica i social de Catalunya i que sigui útil a la presa de decisions de les institucions públiques i dels agents socials, minimitzant les molèsties als ciutadans i garantint el secret estadístic.

L'aprofitament de les fonts existents, amb la utilització d'informació d'origen administratiu o bé estadístic, és una alternativa a operacions de camp complexes i costoses que permet, a la vegada, minimitzar les molèsties als ciutadans. L'acompliment d'aquest criteri es posa de manifest en els continguts dels programes. Pel que fa al programa de 1994, s'han classificat les activitats estadístiques d'acord amb la procedència directa de les dades que han de ser tractades.

A l'objecte de conèixer l'aprofitament de les fonts d'informació s'ha fet una classificació bastant agregada. Cal dir que en el cas de coexistir en una activitat estadística una operació de camp amb altres fonts existents, s'ha classificat a la primera categoria. Només en 25 de les 36 activitats, les dades tractades procedeixen exclusivament d'una operació de camp.

Fonts d'informació	Activitats	%
Tractament de dades d'una operació de camp	36	19.3
Tractament de dades primàries d'origen administratiu, estadístic o mixt	120	64.2
Recopilació de dades d'origen administratiu, estadístic o mixt i publicacions estadístiques	18	9.6
Instrumentals	12	6.4
A concretar (projecte)	1	0.5
Total activitats programades	187	100

La consideració més precisa sobre la procedència de les dades en la producció estadística ens porta a afirmar que la font d'informació més utilitzada, de forma exclusiva o bé conjuntament amb altres fonts és la d'origen administratiu. Certament, la major part de les activitats desenvolupades pels departaments de la Generalitat són resultats de la pròpia gestió administrativa. Hi ha 98 activitats de producció estadística que han utilitzat dades d'aquesta procedència. També pel que fa a les entitats territorials 10 de les 13 activitats estadístiques són d'origen administratiu.

Des del vessant d'aprofitament de fonts d'informació existents s'han d'assenyalar també els diferents nivells de col·laboració de l'Institut d'Estadística de Catalunya amb les operacions estadístiques fetes per l'INE, la qual cosa ha permès obtenir, a més, informació amb una major desagregació territorial que la prevista per a tot l'Estat, acompanyada, en ocasions, d'una major exhaustivitat i, fins i tot, de l'obtenció de variables específiques per a Catalunya complementàriament a les dades comunes a tot l'Estat. Els diferents nivells de participació en les diferents fases d'una operació estadística varien d'una operació a una altra i queden reflectits en els continguts dels convenis de col·laboració entre ambdues institucions.

Finalment, indiquem també l'existència de convenis signats amb les universitats catalanes pel que fa a diverses activitats instrumentals.

Acabem aquesta exposició assenyalant l'existència de dos articles que tenen en compte diferents aspectes del sistema estadístic. En l'article de José A. Casco, "Estadística oficial i coordinació del sistema estadístic", publicat en la revista *Qüestiió*, vol. 18 núm. 1, s'analitza de forma precisa el contingut del sistema estadístic de Catalunya. En el primer apartat es fa referència a conceptes bàsics de terminologia estadística. La interpretació dels termes *estadística, activitat estadística oficial i pública, resultats estadístics oficials* i altres conceptes són precisats en aquest estudi. Pel que fa al sistema estadístic català, es presenten en un primer lloc les característiques del sistema i els trets diferencials respecte a altres models existents, i després l'estructura organitzativa amb el detall de les institucions que la integren i les principals funcions assignades a cadascuna d'elles. Per últim, en un tercer apartat es presenten els instruments legals establerts per portar a terme l'actuació estadística. L'establiment del Pla estadístic i dels programes, la normativa tècnica pel que fa a la normalització de les activitats complementàries de caràcter instrumental per a la producció estadística i la formalització de la relació de cooperacions amb altres institucions alienes al sistema estadístic català, són els instruments necessaris per a la coordinació de l'activitat estadística.

Quant al segon article, publicat en aquest mateix número "Els convenis de col·laboració en el sistema estadístic de Catalunya" elaborat per un equip de tècnics de l'Institut d'Estadística de Catalunya, s'hi analitzen amb profunditat els diferents tipus de convenis que l'Institut d'Estadística de Catalunya ha signat amb institucions de l'Administració de l'Estat, de l'Administració local i les universitats catalanes.

ANNEX 1

Relació nominal de les activitats estadístiques del Programa anual d'actuació estadística per al 1992, per organismes responsables

Departament de la Presidència
Directori de clubs i associacions esportius
Estadística de radiodifusió sonora i televisió local
Estadística de l'ocupació de l'espectre radioelèctric
Estadística de l'atenció al ciutadà en matèria de recepció de senyals
Cens de clubs i associacions esportius

Comissionat per a Universitats i Recerca
Estadística de l'ensenyament universitari
Estadística de recerca i desenvolupament

Departament de Governació
Estadística d'incendis i salvaments
Estadística d'actuacions en matèria de protecció civil
Estadística de les actuacions del cos de la policia autonòmica
Estadística dels accidents de trànsit amb víctimes a Catalunya
Directori de cafeteries i restaurants amb llicències de joc
Estadística de pressupostos de les administracions locals
Estadística de resultats electorals
Estadística de les actuacions del Pla únic d'obres i serveis de Catalunya
Estadística del joc
Estadística de personal de l'Administració

Departament d'Economia i Finances
Estimació de les ràtios econòmiques i financeres de les empreses
Avanç de la taxa de variació del PIB
Conjuntura econòmica
Comptes econòmics de les administracions públiques

Institut d'Estadística de Catalunya
Estadística i comptes econòmics del sector comerç
Estadística de composició i característiques de les famílies
Estadística estructural de migracions
Estadística de la localització de l'ocupació
Estadística de característiques de la fecunditat

Estadística estructural del parc d'habitatges
Estadística de fluxos intramunicipals per treball i estudi
Estadística de naixements i matrimonis
Estadística de moviments migratoris
Base de dades estadístiques comarcals i municipals
Anuari estadístic de Catalunya
Estadística de síntesi
Sistemes interactius d'informació estadística per a la distribució en suports electrònics
Base de dades ESPAN
Biblioteca de l'Institut d'Estadística de Catalunya
Accés a bases de dades d'informació estadística i econòmica internacional
Distribució de bases de dades per teleprocés
Comptes econòmics de Catalunya
Taules input-output
Economies comarcals
Base de dades de sèries i taules estadístiques
Índex de consum privat
Estadística de l'estructura de les explotacions agràries
Comptes econòmics del sector industrial
Estadística industrial d'empreses
Estadística de la inversió estrangera autoritzada
Índex d'activitat industrial
Estadística i comptes econòmics del sector del transport de mercaderies per carretera
Economia de les famílies
Estadística del mercat de treball
Estadística del comerç exterior
Projeccions de població
Estadística de despeses i ingressos de l'ensenyament privat no universitari
Estadística de la despesa social pública
Estadística de biblioteques
Estadística comarcal i municipal
Codis oficials d'entitats territorials i administratives
Normalització de nomenclatures i classificacions estadístiques
Procediments per a la preservació del secret estadístic
Disseny de mostres estadístiques
Suport tècnic en l'execució del Programa anual d'actuació estadística
Verificació de l'acompliment del Pla estadístic i del Programa anual d'actuació estadística
Verificació de l'acompliment de la legalitat estadística
Elaboració de la proposta del Programa anual d'actuació estadística

Formació en procediments estadístics
Sistema de codificació automàtica
Modelització economètrica regional
Difusió i promoció de la recerca estadística a Catalunya

Departament d'Ensenyament
Directori de centres d'ensenyament
Estadística de l'ensenyament infantil i primari
Estadística de l'ensenyament secundari
Estadística de l'ensenyament musical, artístic i altres ensenyaments especialitzats
Estadística de l'educació especial
Estadística de la despesa pública en ensenyament

Departament de Cultura
Directori d'empreses editorials, periodístiques, cinematogràfiques i discogràfiques
Directori de cinemes i teatres
Estadística del cinema
Directori de museus
Estadística de teatre
Estadística de música

Departament de Sanitat i Seguretat Social
Directori d'establiments sanitaris
Estadística d'establiments hospitalaris
Estadística de causes de mort
Estadística de morbiditat hospitalària
Estadística de malalties de declaració obligatòria
Estadística del control sanitari d'aigües
Estadística dels programes de vacunació
Directori de les entitats d'assegurança lliure d'assistència mèdico-farmacèutica
Estadística dels centres integrats a la Xarxa d'Atenció a les Drogodependències
Estadística de la despesa sanitària pública
Estadística de salut de la població
Estadística d'interrupcions voluntàries d'embaràs

Departament de Política Territorial i Obres Públiques
Anuari estadístic de política territorial i obres públiques
Estadística de la construcció d'habitatges de protecció oficial
Estadística de la construcció residencial
Estadística dels préstecs protegits per a la compra i rehabilitació d'habitatges
Estadística de llicències d'edificació
Estadística de demanda d'habitatges

Estadística de l'oferta d'habitatges
Directori de les empreses de transport
Estadística del transport ferroviari de la Generalitat de Catalunya
Estadística de trànsit marítim
Estadística d'actuacions de la Generalitat en obres públiques
Estadística de vehicles matriculats
Estadística hidrogràfica
Estadística de finançament al sector de la construcció
Estadística del transport regular de viatgers per carretera
Estadística del transport de mercaderies per carretera
Estadística dels transports ferroviari, aeri i marítim
Directori d'empreses del sector de la construcció
Estadística de la xarxa viària de Catalunya
Estadística de la xarxa de transport

Departament d'Agricultura, Ramaderia i Pesca
Estadística de les superfícies i produccions agrícoles
Estadística de la ramaderia
Estadística de les superfícies i produccions forestals
Estadística de les captures pesqueres
Estadística dels preus agraris cotitzats en llotges i mercats
Estadística dels preus rebuts i pagats pels pagesos
Comptes econòmics del sector agrari
Xarxa comptable agrària
Enquesta i comptes de la indústria agroalimentària
Directori de les indústries agroalimentàries

Departament de Treball
Anuari estadístic de treball
Estadística de vagues i tancaments patronals
Estadística d'expedients de regulació d'ocupació
Estadística de conciliacions laborals individuals
Estadística de sancions en matèria laboral
Estadística de conciliacions laborals col·lectives
Estadística d'obertures de centres de treball
Estadística de sinistralitat laboral
Directori de mutualitats de previsió social voluntària
Directori de les societats cooperatives
Directori de les societats anònimes laborals
Directori d'associacions professionals
Directori d'empreses i de centres de treball
Estadística de l'ocupació assalariada

Estadística de convenis col·lectius
Estadística de la formació professional ocupacional

Departament de Justícia
Estadística penitenciària
Estadística de reinserció juvenil
Directori de fundacions i d'associacions
Directori d'edificis i locals judicials

Departament d'Indústria i Energia
Estadística de conjuntura elèctrica
Estadística municipal i comarcal del consum de gasos canalitzats i d'energia elèctrica
Estadística del gas natural
Estadística de l'energia elèctrica
Estadística del carbó i del coc de petroli
Estadística del petroli
Estadística del consum energètic del sector industrial
Balanç energètic
Estadística d'inversions en noves indústries i ampliacions
Estadística de les variacions del Registre Industrial
Estadística d'accidents industrials
Estadística de l'estat mecànic del parc de vehicles
Expectatives empresarials
Estadística d'indicadors de la seguretat de les instal·lacions industrials
Base de dades del parc de vehicles

Departament de Comerç, Consum i Turisme
Base de dades d'establiments turístics
Directori d'allotjaments hotelers, càmpings, agències de viatges i altres establiments turístics
Estadística del turisme entrat per la Jonquera
Estadística de la despesa turística

Departament de Benestar Social
Directori d'entitats, serveis i establiments socials
Directori de centres de formació d'adults
Estadística de l'alumnat de formació d'adults
Estadística del professorat de formació d'adults
Estadística sobre disminucions registrades a Catalunya

Departament de Medi Ambient
Estadística de residus sòlids urbans
Estadística de residus industrials
Indicadors de contaminació atmosfèrica
Indicadors de medi ambient
Indicadors de qualitat de l'aigua litoral
Estadística de la despesa pública en medi ambient
Estadística sobre entitats subministradores d'aigua

Activitats estadístiques de les entitats públiques catalanes de caràcter territorial
Desagregació conceptual i territorial dels censos d'edificis i locals de 1990 i de població i habitatge de 1991
Avanç d'estadístiques de resultats electorals del municipi de Barcelona
Estadística del moviment natural de la població del municipi de Barcelona
Estadística del moviment migratori del municipi de Barcelona
Anuari estadístic de la ciutat de Barcelona
Estadística de mortalitat de la població del municipi de Barcelona
Estadística de morbiditat de les malalties de declaració obligatòria de la ciutat de Barcelona
Cens de vehicles del municipi de Barcelona
Estadística del mercat de treball del municipi de Barcelona
Estadística del cadastre del municipi de Barcelona
Estadística de les activitats econòmiques del municipi de Barcelona
Índex de capacitat econòmica familiar a la ciutat de Barcelona
Anuari estadístic de la ciutat de l'Hospitalet de Llobregat

Distribució de les activitats estadístiques del Programa anual d'actuació estadística de 1994 per nivells d'execució i per àrees temàtiques

ÀREES TEMÀTIQUES	TOTAL
ECONÒMIQUES	**53**
Síntesi econòmica	10
Agricultura, ramaderia i pesca	9
Indústria i energia	14
Construcció	2
Serveis	6
Hisenda pública	1
Administració pública	1
Locals i establiments de caràcter econòmic	3
Economia familiar	2
Treball i salaris	3
Sector exterior	2
DEMOGRÀFIQUES	**10**
Síntesi demogràfica	1
Estructura i fluxos demogràfics	9
SOCIALS	**81**
Síntesi social	—
Situació social de la població	—
Edificis i habitatges	7
Ensenyament i recerca	10
Cultura	7
Esports	2
Salut	14
Condicions laborals	12
Benestar social	6
Infraestructura	4

ÀREES TEMÀTIQUES	TOTAL
Transports	8
Comunicacions	3
Medi ambient	8
ACTIVITATS ADMINISTRATIVES ESTRICTES	**17**
Justícia	4
Seguretat	8
Control administratiu del joc	2
Eleccions	2
Personal de l'Administració pública	1
MULTIDISCIPLINÀRIES	**14**
Base de dades	3
Anuari	7
Biblioteca	1
Altres	3
INSTRUMENTALS LEGALS	**9**
Homogeneïtzació estadística	2
Correcció tècnica	6
Secret estadístic	1
Assistència tècnica estadística	4
Fiabilitat estadística	1
Formació en procediments estadístics	1
INSTRUMENTALS PER A L'OPTIMITZACIÓ DEL SISTEMA ESTADÍSTIC	**3**
Desenvolupament tècnic	1
Perfeccionament metodològic	2
Formació en procediments estadístics	—
TOTAL ACTIVITATS 1994	**187**

ENGLISH SUMMARY:

ANNUAL PROGRAMME OF STATISTICAL ACTUATION

Montserrat Biosca

Within Catalonia's legislative framework, the programming of official statistical activity is regulated by certain general norms and by other specific ones. These are established for each activity and described uniformly on index cards included in the annual Programme. Aside from these aspects, the article deals with the most significative traits deriving from the Programme's overall analysis.

The purpose of the article is to deal with the programming of statistical action carried out by the Generalitat de Catalunya (Autonomous Government of Catalonia). There is first a brief reference to the legal foundation on which the operation of the Institut d'Estadística de Catalunya is established as the managing organization, the manner of regulating statistical action through the creation of the legal Bill of Statistical Plans, the successive annual programmes of statistical actions, the procedure of consulting institutions when carrying them out, and, finally, the responsibility assigned to the Institute in this process.

The programming of official statistical activity in Catalonia is based on a legal decree detailing descriptions of activities to be dealt with during the year. These are regulated by certain general norms which govern all activities included in the programme. Further regulations include specific norms applied to each concrete activity. These norms are specified in the article.

Each statistical activity is described on an index card with a uniform format. The technical characteristics and specific regulating norms are recorded on the cards. The descriptive card for each statistical activity deals with: identification of the activity, description, dissemination of statistical results, estimated direct cost, and legal base. Each of these categories is delimited in a set of concrete features. This facilitates reading of the technical data. The description of each of the concepts employed is also dealt with in the article, where possible alternative strategies are specified in some cases.

Finally, there is an analysis of the content of the annual programme of statistical action. The Institut d'Estadística de Catalunya serves a double purpose. On the one hand it is the managing organization of the statistical system, on the other hand it is one more agent in the system itself. The way in which the system's decentralized model operates is clearly seen when we specify the responsiblities of different organizations in carrying out statistical analyses. The Institute directly assumes instrumental activities. This includes in a subsidiary way activities of statistical production which, because of their own areas of responsibility, are carried out by other departments of the Generalitat. Determination of the significative aspects is merely to corroborate features of the statistical system described in other articles published in the same journal.

One section allowing analysis of the fulfilment of annual programmes is the one dealing with the level of execution. Monitoring this shows how the statistical system is consolidated. Activities in progress, i.e., the periodically updated data, are constantly growing as execution of new activities, or projects, is beginning. Also, the statistical system's regulating technical norms merely contribute to system consolidation, as does the incorporation of territorial type Catalan organizations.

A further significant feature is the priority Catalonia's statistical system awards the exploitation of existing sources. This allows data from the administration or statistical information to be used as an alternative to expensive and complex field operations, thus minimizing problems for the public. It also provides a diversity of subject matter, as has been seen in programmed statistical activities, fulfilling the primary aim of the Catalan Statistics Law 1992-1995 of achieving a coherent, reliable, up to date set of statistical data at the minimum possible cost. This optimally exploits existing sources, with a knowledge of Catalonia's economic, demographic and social reality. It is useful in decisions taken by public institutions and social agents, minimizing any problems encountered by the public and guaranteeing statistical confidentiality.

ELS CONVENIS DE COL·LABORACIÓ AMB L'INSTITUT D'ESTADÍSTICA DE CATALUNYA

JORDI BACARIA, JOAQUIM CAPELLADES,
ÀLEX COSTA, MANUEL FALGUERA

Institut d'Estadística de Catalunya

L'Institut d'Estadística de Catalunya en la línia pròpia de planificació de l'activitat estadística de Catalunya ha establert un seguit de convenis de col·laboració en matèria estadística amb diferents administracions per tal d'evitar duplicitats d'actuacions i molèsties innecessàries als subministrants d'informació, minimitzar els costos de l'activitat estadística amb la finalitat de contribuir a cohesionar i optimitzar el sistema estadístic de Catalunya. En l'article s'analitzen algunes de les modalitats de cooperació interinstitucional tant amb l'Administració central com amb altres administracions.

Agreements of collaboration with the Catalan Statistical Institute.

Key words: Col·laboració, conveni, estadística, administració.

1. INTRODUCCIÓ

En l'article del senyor Josep Casco "Estadística oficial i coordinació del sistema estadístic català", publicat en el número anterior d'aquesta mateixa revista, s'entenia per sistema estadístic el conjunt de normes jurídiques, els processos a complir pels agents que hi intervenen i la seva interrelació, els plans, programes i activitats ordenats en un conjunt harmònic que té com a finalitat l'assoliment

–Article rebut el febrer de 1994.

–Acceptat el maig de 1994.

d'uns objectius concrets. En el cas de Catalunya aquestes relacions estan definides per la Llei 14/1987 d'estadística de Catalunya i, molt especialment, per la Llei 30/1991 del pla estadístic de Catalunya 1992-1995.

El Pla estadístic de Catalunya és concebut com l'instrument d'ordenació i planificació de l'estadística de la Generalitat i dels seus organismes autònoms, i també com l'instrument marc de col·laboració institucional entre la Generalitat i les entitats públiques catalanes de caràcter territorial per a la progressiva constitució del sistema estadístic integral de Catalunya, així com amb les entitats públiques estatals o internacionals. L'esmentat Pla explicita els objectius operatius que seran objecte de les activitats estadístiques, les quals es concretaran anualment en els programes d'actuació estadística, i estableix un objectiu central que serà l'eix vertebrador de tota l'activitat: *aconseguir un conjunt coherent, fiable i actualitzat de dades estadístiques al mínim cost possible, aprofitant al màxim les fonts existents, que permeti el coneixement de la realitat econòmica, demogràfica i social de Catalunya i que sigui útil per a la presa de decisions de les institucions públiques i als agents socials, minimitzant les molèsties als ciutadans i garantint el secret estadístic.*

En l'estructura organitzativa de l'estadística oficial catalana, l'Institut d'Estadística de Catalunya té un paper central ja que és l'organ especialitzat del Govern de Catalunya en matèria estadística, responsable de la seva planificació i de la gestió del sistema estadístic en el seu conjunt. Dins d'aquesta activitat planificadora que li és pròpia, l'Institut d'Estadística de Catalunya ha establert un seguit de convenis de cooperació en matèria estadística amb diferents administracions per tal d'evitar duplicitats d'actuacions i molèsties innecessàries als subministrants d'informació, i contribuir a cohesionar el sistema estadístic de Catalunya. Aquests convenis de col·laboració amb l'Institut són, en el cas de les institucions públiques catalanes de caràcter territorial, l'única via d'entrada al Pla estadístic i, en conseqüència, a l'estadística oficial, per la qual cosa adquireixen una gran importància des del punt de vista institucional.

2. EL CONVENI INTERADMINISTRATIU COM A UN INSTRUMENT JURÍDIC DE COL·LABORACIÓ

Dins de les relacions de cooperació entre les administracions públiques la formalització de convenis de col·laboració constitueix una fórmula tècnico-jurídica que consisteix en un contracte basat en l'autonomia de la voluntat de les administracions, per mitjà del qual les parts concreten i modulen les condicions de l'exercici de les competències respectives per tal d'obtenir un fi d'interès públic. La distribució i possible concurrència de competències en les administracions públiques condueix a la necessitat d'aquests pactes de col·laboració sense renúncia de les competències pròpies.

En relació amb el seu règim jurídic, l'article 1 de la Llei de contractes de l'Estat exclou de la seva regulació tant els convenis interadministratius com els convenis que subscrigui l'Administració amb particulars amb finalitats d'interès públic. No obstant això, aquests convenis de col·laboració són una classe especial de contractes sotmesos al Dret Administratiu que tenen com a límit els principis de bona administració i l'ordenament jurídic i han de formalitzar-se amb l'objecte d'aconseguir un interès públic.

Els criteris generals d'homogeneïtzació de convenis entre l'Administració central i les comunitats autònomes es van establir en les sessions de la Comissió Delegada per a la Política Autonòmica de 13 de desembre de 1984 i de 18 de juny de 1985. Aquesta Comissió Delegada del Govern central és qui autoritza la subscripció de cadascun dels convenis i els criteris adoptats es refereixen a la consideració dels convenis com a instrument de política autonòmica del Govern central, en els quals s'ha de tenir en compte, sempre que hi hagi aportació de recursos pressupostaris, la política d'inversions de cada departament en el marc de les directrius del Govern.

Les primeres regles sobre convenis entre l'Administració central i les comunitats autònomes es van establir en l'Acord del Consell de Ministres de 2 de març de 1990 i es justifiquen amb l'objecte de perfeccionar la qualitat tècnico-jurídica d'aquests intruments de col·laboració. Aquestes regles es refereixen als títols competencials, als objectius, a la definició dels mecanismes d'assistència tècnica, coordinació i seguiment, finançament, vigència i prorrogabilitat.

Actualment l'article 6.2 de la Llei 30/1992, de 26 de novembre de règim jurídic de les administracions públiques i del procediment administratiu comú, regula aquests tipus de convenis i estableix aquells elements que han d'especificar els convenis de col·laboració i que coincideixen essencialment amb el contingut de l'Acord del Consell de Ministres a què s'ha fet referència. Cal remarcar que es presta una major atenció als supòsits d'extinció dels convenis per causa distinta de la finalització del termini de vigència i a les conseqüències que pot comportar per a les actuacions en curs.

Respecte als convenis, concebuts com a tècnica de col·la boració de les entitats locals, tot i respectar l'autonomia local, l'article 10 de la "Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local", regula en l'àmbit d'aquestes entitats locals els principis de col·laboració i coordinació i especifica la pertinença de la coordinació entre els ens locals i les altres administracions públiques quan es produeixi concurrència o complementarietat competencial i en particular l'article 57 estableix que la cooperació econòmica, tècnica i administrativa entre l'Administració local i les administracions de l'Estat i de les comunitats autònomes poden tenir lloc per mitjà de consorcis o convenis administratius.

La Llei 8/1987, de 15 d'abril, municipal i de règim local de Catalunya recull, així mateix, en l'article 129 c) els principis de cooperació, col·laboració i assistència recíproques entre l'Administració de la Generalitat i els ens locals de Catalunya per al millor compliment de les funcions que els correspongui.

3. EL CONVENI DE COL·LABORACIÓ EN LA LEGISLACIÓ ESTADÍSTICA CATALANA

La legislació estadística catalana preveu d'una manera palesa el conveni interadministratiu com a un instrument indispensable per a l'actuació estadística de les administracions públiques catalanes, especialment pel que fa a convenis entre l'òrgan estadístic de la Generalitat, tant amb l'Administració central com amb les entitats locals de Catalunya.

El contingut de les funcions de l'Institut d'Estadística de Catalunya, establertes per l'article 45 de la Llei 14/1987, de 9 de juliol, d'estadística i per l'article 2 del Decret 341/1989, d'11 de desembre, de creació de l'Institut d'Estadística de Catalunya, permeten deduir la necessitat del seu exercici per via de relacions de cooperació amb altres entitats estatals, europees o amb particulars.

L'article 6 de la Llei 30/1991, de 13 de desembre, del Pla estadístic de Catalunya 1992-1995, especifica que l'Institut d'Estadística de Catalunya és l'òrgan responsable de dur a terme el Pla estadístic i els programes anuals d'actuació estadística directament o en col·laboració amb altres entitats públiques o privades i, més concretament, l'article 27 de l'esmentada Llei estableix la figura del Conveni de col·laboració entre l'Institut d'Estadística de Catalunya i les entitats locals de Catalunya pel que fa a la incorporació als programes anuals d'actuació estadística d'estadístiques d'interès especial per a les entitats locals.

Finalment, cal fer esment que les disposicions generals de la Llei 30/1991, de 13 de desembre, del pla estadístic de Catalunya 1992-1995, conceben el Pla estadístic com un instrument marc de col·laboració institucional entre la Generalitat de Catalunya i els seus organismes i empreses amb les entitats públiques catalanes de caràcter territorial, entitats públiques estatals o europees i organitzacions internacionals.

4. TIPOLOGIA JURÍDICA DELS CONVENIS DE DE COL·LABORACIÓ FORMALITZATS PER L'INSTITUT D'ESTADÍSTICA DE CATALUNYA

L'Institut d'Estadística de Catalunya, en el marc de la legislació estadística catalana, ha formalitzat convenis de col·laboració, d'acord amb la regulació aplicable, que es poden anomenar *verticals*, amb l'Administració de l'Estat i amb les entitats locals i els que es poden definir com a *horitzontals*, amb altres departaments o organismes de la Generalitat. També ha signat convenis de cooperació amb les universitats catalanes per a l'actuació conjunta en l'àmbit de la recerca i la formació en el camp estadístic. Aquests convenis es poden dividir en diverses classes:

Segons els subjectes contractants:

- Convenis amb l'Administració de l'Estat
- Convenis amb d'altres departaments o organismes de la Generalitat
- Convenis amb les entitats locals
- Convenis amb les universitats catalanes

Segons l'objecte del conveni:

- Convenis per dur a terme actuacions conjuntes. És el cas de la majoria de convenis signats per l'Institut d'Estadística de Catalunya amb l'Institut Nacional d'Estadística (INE) i amb les universitats catalanes.
- Convenis per a l'assoliment de fins. L'Institut d'Estadística de Catalunya ha signat convenis amb entitats locals territorials amb la finalitat d'incorporar estadístiques d'interès especial per a elles en els programes anuals d'actuació estadística i procedir a la seva realització, dins del marc del Pla estadístic de Catalunya 1992-1995.

Pels seus efectes:

- Convenis marc

  L'Institut d'Estadística de Catalunya ha formalitzat convenis marc que contenien intencions de cooperació o principis de col·laboració relatius a actuacions estadístiques, la major part d'ells amb diputacions, consells comarcals i ajuntaments de Catalunya.

- Convenis contractuals

  Aquests tipus de convenis, que creen vincles d'obligació per a les parts signants de l'acord, constitueixen el conjunt més important signat per l'Institut d'Estadística de Catalunya i contenen objectes i finalitats diversos en l'àmbit d'actuació estadística d'acord amb l'exercici de les competències de l'Institut d'Estadística de Catalunya.

5. CONVENIS DE COL·LABORACIÓ AMB L'ADMINISTRACIÓ DE L'ESTAT

5.1. Objectius

De l'objectiu central del Pla estadístic de Catalunya es poden identificar alguns dels objectius bàsics de la col·laboració entre l'Administració central i la Generalitat de Catalunya. Els camps temàtics objecte d'interès per part d'ambdues administracions són pràcticament coincidents quant a disposar d'informació estadística per a la presa de decisions de les institucions públiques. Tanmateix la desagregació territorial de les dades necessàries és diferent segons l'Administració corresponent; així l'Administració central acostuma a considerar suficient la desagregació provincial i com a màxim tracta aïlladament els grans municipis, mentre que la Generalitat de Catalunya, i en general les diferents administracions autonòmiques de l'Estat, necessiten la informació molt més desagregada territorialment: comarques, municipis i fins i tot àrees inframunicipals: districtes administratius, seccions censals i zones estadístiques.

Així, definits com a pràcticament coincidents els centres d'interès estadístic, múltiples raons determinen la necessitat d'acordar la col·laboració interinstitucional:

- L'aprofitament de les fonts existents. Així es tracta de fer una única operació estadística que permeti atendre les necessitats de cada institució.
- La minimització dels costos, ja que el cost de passar d'un producte estadístic d'utilització estatal a un d'autonòmic freqüentment és un cost merament marginal.
- La minimització de les molèsties als ciutadans, perquè amb una sola demanda d'informació a la població i a les empreses, es pot abastar el diferent ventall de necessitats d'ambdues administracions.
- La comparabilitat de resultats mitjançant la utilització de metodologies idèntiques, ja que interessa a tothom que les dades obtingudes en una operació estadística siguin plenament comparables en l'entorn de cada administració. L'Administració autonòmica necessita emmarcar els seus

resultats amb els de l'entorn estatal i també amb els d'altres comunitats i l'Administració central necessita disposar de dades obtingudes amb idèntica metodologia en tot el territori de l'Estat.

Altres objectius que poden ser tan importants com els anteriors, però que no deriven directament de la legislació estadística són:

- La normalització lingüística, mitjançant la introducció de qüestionaris bilingües en català i castellà en les operacions estadístiques estatals adreçades a la població i a centres i empreses de Catalunya. També l'edició de manuals d'instruccions i fulletons en text català o bilingües. Fins i tot es pot assenyalar que s'han arribat a utilitzar qüestionaris trilingües en el cas del cens de població i de la renovació del padró municipal d'habitants de 1991, ja que els impresos emprats a la Val d'Aran eren exemplars en versió aranesa/catalana/ castellana.

- La presència institucional de la Generalitat de Catalunya i de l'Institut d'Estadística en representació seva, davant la població i les empreses de Catalunya. Aquesta presència institucional pot ser diversa segons el tipus d'intervenció de l'Institut en l'operació estadística tractada.

- La introducció en els qüestionaris de temes d'interès específic per a l'Administració de la Generalitat. En aquest sentit es poden destacar especialment l'estudi del coneixement del català i dels fluxos de desplaçament per motius de treball o estudi, així com el mitjà de transport utilitzat en el desplaçament pel que fa al cens de població i també l'estudi detallat de la ramaderia o els equipaments productius en el cens agrari.

Finalment, però no per això menys destacables, s'ha d'assenyalar la disponibilitat de la informació recollida per part de l'Institut d'Estadística de Catalunya. Aquesta disponibilitat és immediata després de les operacions de tractament estadístic en el cas de les operacions que s'han recollit directament per part de l'Institut, especialment si es tracta d'operacions de tipus censal i que no necessitin operacions addicionals d'elevació de dades obtingudes per mostreig. Com a cas paradigmàtic en aquest sentit es pot assenyalar el cens de població de 1991.

5.2. Continguts

Els continguts dels convenis de col·laboració són força diferents segons el model de col·laboració assolit en cada cas. Així es pot parlar de dos tipus bàsics pel que fa a la distribució institucional de les diferents tasques d'una operació estadística. Aquests tipus els podem identificar com de participació total i de participació parcial:

- *El model de participació total*, es pot considerar aquell que parteix d'una operació estadística definida metodològicament de manera homogènia per a tot l'Estat i és l'organisme autonòmic l'encarregat de desplegar-lo i executar-lo íntegrament en el seu territori. El cas més representatiu és el del cens agrari en el qual l'Institut, a partir de la metodologia bàsica per a tot l'Estat, va executar enterament l'operació estadística a Catalunya, en la totalitat de les etapes de la recerca estadística, des de l'edició dels qüestionaris amb la incorporació d'ítems específics per a Catalunya i dels corresponents manuals d'instrucció, fins a la plena informatització i explotació de les dades censals, passant per totes les fases intermèdies com poden ser les de selecció i formació de personal, treball de camp de recollida o l'enregistrament de dades.

  Altres casos significatius d'aquest model són, en l'àrea econòmica, les enquestes de comerç interior de 1988 i 1992, també l'enquesta de transports i mercaderies per carretera 1992. Aquesta darrera va tenir la particularitat d'implicar, a més de l'INE i l'Institut, el Ministeri d'Obres Públiques i el Departament de Política Territorial i Obres Públiques de la Generalitat. En l'àrea social pertanyen també a aquest model l'enquesta de finançament i despesa de l'ensenyament privat no universitari del curs 1990-91, l'enquesta de biblioteques de 1992 i l'enquesta d'estructura de les explotacions agràries de 1993.

- *El model de participació parcial* admet diferents graus i nivells de distribució institucional de les tasques estadístiques i pot abastar intervencions diverses de cada administració. Aquesta col·laboració es pot limitar a uns continguts mínims com poden ser: la realització de versions catalanes dels qüestionaris, o col·laborar en la identificació de directoris estadístics bàsics, fins a participacions més complexes, en les quals es distribueixen les diferents etapes de la recerca estadística entre les diverses institucions afectades. Un cas típic de participació complexa pot ser el del cens de població i habitatge de 1991, en el qual l'INE va fer el desplegament metodològic del projecte, l'edició dels qüestionaris i el treball de camp per a la recollida de dades, mentre que l'Institut d'Estadística de Catalunya va participar amb la introducció de preguntes específiques, la realització de la versió catalana dels qüestionaris i finalment l'enregistrament, validació i tractament informàtic de les dades, així com les explotacions corresponents.

  Es poden esmentar també per la seva significació en l'àmbit del coneixement estadístic de l'economia catalana, les intervencions en les enquestes industrials dels anys 1991, 19 92 i 1993. Això ha suposat promoure conjuntament l'operació davant dels empresaris industrials de Catalunya, fer l'edició bilingüe dels qüestionaris cosa que implica una important feina de terminologia en català (en els quasi 80 models de qüestionari es troben

5.000 termes referits a diferents entrades i productes de l'activitat industrial i, finalment, suposa fer una elevació i una tabulació específica per a Catalunya. El procés d'elevació resulta particularment rellevant atès que els resultats passen a ser compatibles no ja amb l'enquesta industrial com a operació estadística singular, sinó amb la comptabilitat econòmica de Catalunya. D'aquesta manera s'aprofiten fonts d'informació complementària i s'arriba a uns resultats més realistes que els que proporciona directament l'enquesta.

Un cas particular és el de l'enquesta dels comptes de l'Administració pública. Es tracta d'una col·laboració entre la Intervenció de la Generalitat i la Intervenció General de l'Estat, amb la participació de l'Institut d'Estadística de Catalunya. El treball de camp és fet per les dues intervencions, mentre que l'edició dels qüestionaris, la validació, elevació i tabulació les porta a terme l'Institut d'Estadística de Catalunya.

Amb independència del model de col·laboració acordat en cada cas, els continguts més freqüents en els diferents convenis subscrits amb l'INE es refereixen als següents apartats:

- Edició de qüestionaris, realització de versions bilingües.
- Propaganda i difusió de l'operació estadística entre els destinataris. Presentació al sector a qui va dirigida l'operació estadística.
- Selecció, contractació i formació de personal. S'ha d'assenyalar la importància que el personal de camp sigui enterament bilingüe.
- Recollida, depuració i revisió de qüestionaris.
- Enregistrament de dades, prescripcions tècniques per a efectuar-lo, criteris de validació i controls de qualitat.
- Depuració i actualització de directoris en el cas de les enquestes per mostreig.
- Lliurament de l'arxiu estadístic a l'INE en suport magnètic. Avaluació posterior de l'INE i conformitat definitiva de l'arxiu.
- Lliurament final dels qüestionaris.
- Disponibilitat de la informació per a explotacions estadístiques pròpies de cadascuna de les parts.
- Contraprestacions econòmiques si s'escau.
- Establiment de comissions de seguiment del conveni.

6. CONVENIS DE COL·LABORACIÓ ESTADÍSTICA AMB L'ADMINISTRACIÓ LOCAL

6.1. Finalitat dels convenis

La legislació catalana reconeix la competència de l'Administració local per a la realització d'estadístiques que puguin ser del seu interès per tal de dur a terme, d'una manera més eficaç, les tasques que té encomanades de govern i d'administració. Al mateix temps concep el Pla estadístic de Catalunya, com l'instrument marc de col·laboració institucional entre la Generalitat de Catalunya i els seus organismes i empreses d'una banda, i les entitats públiques catalanes de caràcter territorial de l'altra, per tal d'obtenir informació fiable i útil per a la presa de decisions, evitar la duplicitat d'actuacions estadístiques, la qual cosa comporta sempre un increment innecesari de la despesa pública i de molèsties als ciutadans, i contribuir progressivament a la constitució d'un sistema estadístic integral de Catalunya.

La participació en el Pla estadístic de Catalunya dóna a l'activitat estadística la categoria d'estadística oficial, és a dir, tenen la consideració d'estadística d'interès públic per part del Parlament de Catalunya, i en conseqüència gaudeixen de l'obligatorietat de la col·laboració ciutadana. Al mateix temps, suposa el compromís de fer les estadístiques segons unes normes reguladores que garanteixin la correcció tècnica, la comparabilitat de resultats i tots aquells preceptes establerts per la legislació estadística catalana.

Les consideracions exposades en l'apartat anterior, tenen plena vigència pel que fa a l'estadística local. Així una activitat duta a terme per un ajuntament, consell comarcal o diputació pot estar inclosa en l'esmentat Pla i ser considerada estadística oficial.

La via d'entrada ordinària de l'estadística de les entitats territorials en el Pla és mitjançant conveni amb l'Institut d'Estadística de Catalunya. D'aquesta manera es respecta el caràcter voluntari de la participació, tal com preveu la legislació vigent en matèria estadística a Catalunya.

6.2. Models de participació en el Pla estadístic de Catalunya

La Llei del pla estadístic de Catalunya 1992-95 preveu la participació de les entitats territorials des de dues perspectives diferents, segons el nivell de desenvolupament de l'activitat estadística local. Una que podríem considerar de caràcter passiu i una altra d'actuació directa.

En el primer cas, les entitats territorials de Catalunya poden sol·licitar la col·laboració de l'Institut d'Estadística de Catalunya, mitjançant una memòria justificativa i la signatura del corresponent conveni, per dur a terme les següents activitats:

a) Dur a terme íntegrament les estadístiques que els siguin d'especial interès, sempre que compleixin els objectius del Pla estadístic.

b) Incloure aspectes que els siguin d'especial interès en les estadístiques que es facin en el marc del Pla estadístic.

c) Desagregar les dades o els productes de difusió que els siguin d'especial interès derivats de les estadístiques que es facin en execució del Pla estadístic.

d) Rebre el servei d'assistència tècnica per a la realització, per part d'elles, d'activitats estadístiques del seu interès que s'ajustin als objectius del Pla.

Pel que fa a la segona possibilitat de participació, la de caràcter actiu, l'Institut d'Estadística de Catalunya pot proposar a les entitats territorials, i sempre per la via de conveni, les actuacions que segueixen:

a) Realització d'estadístiques noves o activitats estadístiques complementàries que consideri d'interès per a la progressiva constitució del sistema estadístic català.

b) Proposar la inclusió d'estadístiques que ja porten a terme aquestes entitats, si ho considera interessant per assolir els objectius previstos en el Pla.

6.3. Convenis establerts en l'actualitat

Per tal de fer efectives les possibilitats descrites en el punt anterior, l'Institut d'Estadística de Catalunya ha signat convenis de col·laboració en matèria estadística amb els ajuntaments de Barcelona, l'Hospitalet de Llobregat, Badalona, Sabadell, Terrassa, Santa Coloma de Gramenet, Lleida, Tarragona, Mataró, Manresa, Granollers, Esplugues de Llobregat i Sant Cugat del Vallès; amb els consells comarcals del Barcelonès, Vallès Occidental, Baix Llobregat, Vallès Oriental, Bages, Alt Empordà, Alt Penedès, Garrotxa, Alt Camp i Urgell; i amb les quatre diputacions de Catalunya. Els acords formalitzats són de dos tipus:

a) Convenis marc

Tenen com a objecte establir la màxima cooperació, entre les institucions signatàries, en tots els vessants de l'activitat estadística. Així, l'intercanvi de metodologia, d'informació, la col·laboració recíproca en estudis tècnics, l'assistència tècnica i la possibilitat de proposar la inclusió d'activitats estadístiques d'interès comú en els programes anuals, formen part dels objectius d'aquest acord.

Un actiu molt important, derivat d'aquests convenis, és l'establiment d'una comissió de seguiment formada per representants de les dues administracions. La seva finalitat és promoure la realització de noves activitats, definir en termes tècnics l'abast de les actuacions estadístiques i vetllar per l'acompliment efectiu dels acords. Amb el temps, el contacte entre els tècnics que integren aquesta plataforma s'ha fet molt estret i s'ha establert una relació molt directa que constitueix un canal permanent de comunicació que agilita enormement l'intercanvi d'informació.

b) Convenis específics

Els objectius assenyalats en el conveni marc es desenvolupen mitjançant convenis específics. Aquests concreten l'activitat estadística a desenvolupar i estableixen els compromisos de cadascuna de les parts. Les activitats estadístiques compreses en aquests acords s'incorporen als programes anuals d'actuació estadística i tenen la consideració d'estadística oficial.

Així, s'ha signat un conveni específic amb totes les entitats d'àmbit local indicades anteriorment, en virtut del qual l'Institut d'Estadística de Catalunya durà a terme una tabulació específica dels censos de població i habitatge al nivell màxim de desagregació tècnica i jurídica possible, així com unes tabulacions específiques bàsiques dels censos d'edificis i locals.

Aquesta tabulació específica s'ha definit per a cadascuna de les administracions signatàries en el marc de la comissió de seguiment respectiva i respon a les diferents necessitats plantejades. En el cas dels ajuntaments de volum de població elevat, s'ha dut a terme una desagregació territorial fins a l'àmbit de secció censal. En les institucions signants de caràcter supramunicipal l'especificitat ha estat més aviat de caràcter sectorial. Aquesta activitat de l'Institut ha estat inclosa en els Programes anuals d'actuació estadística de 1993 i 1994.

Cal fer esment al caràcter actiu de la participació de l'Ajuntament de Barcelona en aquests programes anuals. La seva tradició estadística com a productor i difusor de les estadístiques de la ciutat, confirmada en el fet que publica un *Anuari estadístic de la ciutat* des de 1906 i que en el 1923 es crea l'Institut Municipal d'Estadística, ha fet que s'incorporin en els programes anuals onze fitxes d'activitats estadístiques dutes a terme íntegrament per l'esmentat Ajuntament i que són considerades d'especial interès per part de l'Institut d'Estadística de Catalunya.

Aquesta presència activa del món local en l'estadística oficial s'ha vist incrementada amb la incorporació de l'*Anuari estadístic de la ciutat de l'Hospitalet* al Programa anual d'actuació estadística de 1994.

7. CONVENIS AMB LES UNIVERSITATS CATALANES

Una darrera línia de convenis és la que s'adreça a la millora dels mètodes estadístics. En aquest àmbit cal ressenyar el conveni amb la Universitat de Barcelona per al desenvolupament de la modelització economètrica regional (projecte HISPALINK). Aquesta modelització permet conèixer estimacions de l'ocupació i del valor afegit per sectors de l'economia catalana amb molt poc retard i també fer prediccions amb un model que, si bé és català, està construït de forma articulada en altres models d'altres comunitats autònomes. Un segon àmbit de treball ha estat el càlcul d'errors de l'enquesta industrial i de l'enquesta de població activa. Aquest càlcul resultava complex atès que no es pot fer servir una aproximació algebràica, sinó una aproximació iterativa mitjançant el mètode de la semi-mostra reiterada.

Un dels temes més complicats d'estimació per al coneixement de l'economia catalana és el saldo comercial entre Catalunya i la resta d'Espanya. La importància d'aquesta qüestió va determinar la signatura d'un conveni entre la Universitat Autònoma de Barcelona per tal de fer servir una doble metodologia per estimar aquests resultats: el mètode dels saldos aparents (aproximació comptable) i el mètode dels modes de transport (pas de dades físiques a dades econòmiques). Per acabar només apuntar que amb la Universitat Politècnica de Catalunya i amb la Universitat Pompeu Fabra s'han promogut convenis de contingut docent i de promoció de la investigació.

Dins d'aquesta línia de col·laboració amb les universitats s'ha d'assenyalar el conveni signat amb el Centre d'Estudis Demogràfics per avaluar l'exhaustivitat i redreçar les dades estadístiques del moviment natural de la població en el període comprès entre 1975 i 1984.

ENGLISH SUMMARY:

AGREEMENTS OF COLLABORATION WITH THE CATALAN STATISTICAL INSTITUTE

Jordi Bacaria, Joaquim Capellades,
Àlex Costa, Manuel Falguera

The Catalan Statistical Plan has been created to be the instrument for organizing and planning the statistics of the Catalan Government and those of its independent institutions. Also it wants to be the frame tool of institutional collaboration between the Catalan Government and the local public catalan entities, for the progressive constitution of the catalan integral statistical system, and also with the public institutions of the State or internationals.

At the organized structure of the catalan official statistics the Catalan Statistical Institute has the main roll since it is the Generalitat's unit specialized in statistics, responsable for its planification and for the management of the statistical system as a whole. The Catalan Statistical Institute has established a number of agreements of cooperation with differente administrations to avoid doing the same thing twice causing unnecessary trouble to the information suppliers and thus contributing to the integration of the catalan statistical system.

At the Catalan Statistical Plan it is possible to identify some of the main targets of the collaboration between the central Administration and the Catalan Government. Both Administrations share practically the same fields of interest for having statistical information for the decision making of public institutions. Also the disintegration required by the varios self-administrations of the State varies a lot, in general they need that the necessary data be much more territorially splited.

Due to the fact that the statistical points of interest are almost identical there are several reasons why there is a need for agreements of collaboration amont institutions, these are: taken advantage of the existing sources, the lowest cost, less inconvenience to citizens, comparing de results, language normalization, the institutional presence of the Catalan Government, introdution at the questionnaries of themes of specific interest to the Administration of Generalitat and the punctual disponibility of the collected data.

The contents of the agreements are very different depending on the type of collaboration they come to in each case. Thus it is possible to speak about two basic types with reference to the institutional distribution of the different work-pieces of a statistical operation.

The model of "total participation" is that of a statistical proceeding methodologically defined in the same way for the whole of the State, being the self-administration concern the one that has to developed it and do it completely in its territory.

The model of "partial participation" can have different grades and levels of distributing the statistical work. There can be several kinds of participating for instance doing the catalan version of the questionnaries or collaborating at the identification of the basic statistics directory up to more complex ways through the various steps of the statistical process among all the institutions implied. A typical case of complex participation can be the Population and Housing Cens of 1991. Also the participation at the industrial surveys of the years 1991, 1992, and 1993.

For the catalan public institutions of territorial character the agreements of collaboration with the Institute are the only way of entering the Statistical Plan and therefore the official statistics which gives them much significance from the institutional point of view. The catalan legislation aknowledges the competence of the local Administration to elaborate statistics that can be of interest to them in order to accomplish in a more efficient way the works they were assigned by the Government and the Administration. Participating at the Statistical Plan for Catalonia gives official category to the statistical activity. The common entrance of the statistics of the territorial entities into the Plan is on behalf agreement with the Catalan Statitical Institute. The Institute has signned agreements to make effective the possibilities mentioned above.

Frame agreements: aim to the establishment of the maximum cooperation among the signning institutions.

Specific agreements: the targest aimed to at the frame agreements are developed through specific agreements.

A last line of agreements tries to improve the statistical methods. In this area we should underline the agreements signned with several catalan universities with the aim of developing different tools or statistical products.

SECCIÓ DOCENT I PROBLEMES

La introducció de la nova "SECCIÓ DOCENT I PROBLEMES" a la revista QÜESTIIÓ es fa amb l'objectiu d'incloure una secció on es publiquen articles de caire docent, difícilment publicables en revistes de recerca. Alhora es continua amb l'antiga secció de problemes. A cada número de QÜESTIIÓ s'inclourà d'un a tres problemes i les solucions es donaran en el número següent.

Els lectors poden, si ho volen, proposar problemes amb les solucions pertinents i enviar-los a QÜESTIIÓ, que farà una selecció i en publicarà els més adequats, fent la corresponent referència a l'autor.

També seran ben rebudes solucions alternatives a les propostes fetes per l'autor dels problemes; l'editorial es reservarà, però, el dret a publicar-les.

PROBLEMES PROPOSATS

PROBLEMA N° 51

Demostrar que la descomposición tradicional de la varianza poblacional en variación dentro y entre estratos, es válida para cualquier población infinita con varianza finita.

M. Ruiz Espejo

Universidad Complutense de Madrid

PROBLEMA N° 52

Supuesto conocido el tamaño del estrato de no respuesta, N_2, proponer un estimador insesgado de la varianza del estimador de la media poblacional de Hansen y Hurwitz (1946, *JASA*, **41**, 517–529) y del estimador propuesto por Ruiz (1988, *Trab. Estadíst.*, **3**, **1**, 71–80) frente a los problemas de no respuesta y errores de medida.

M. Ruiz Espejo

Universidad Complutense de Madrid

PROBLEMA N° 53

Sea un juego bipersonal de suma nula con matriz de ganancias:

$$\begin{array}{cc} & \text{Jugador B} \\ \text{Jugador A} & \begin{array}{|cc|} \hline a & b \\ c & d \\ \hline \end{array} \end{array}$$

Supuesto que los valores de la matriz de ganancias son diferentes entre sí, y que no se presenta dominio entre líneas:

I) Obtener las estrategias óptimas de ambos jugadores y el valor del juego.

II) ¿Qué fila (columna) deberá elegirse con mayor probabilidad?

III) Estudiar el comportamiento de las estragegias óptimas y el valor del juego cuando $a \to \infty$, manteniéndose fijos el resto de elementos de la matriz de ganancias y la condición de no dominio.

IV) Estudiar el comportamiento particular de los juegos con matrices de ganancias:

IV-1)

n	$n+2$
$n+3$	$n+1$

IV-2)

n	$3n$
$4n$	$2n$

Ramón Alonso Sanz
E.T.S.I. Agrónomos
Madrid

PROBLEMA N° 54

Sea un juego bipersonal de suma nula con matriz de ganancias:

	Jugador B	
Jugador A	a	b
	c	a

Supuesto que $a = d, a \neq b, a \neq c$ y que no se presenta dominio entre líneas:

I) Obtener las estrategias óptimas de ambos jugadores y el valor del juego.

II) ¿Qué fila (columna) deberá elegirse con mayor probabilidad?

III) Estudiar el comportamiento de las estrategias óptimas cuando b crece, manteniéndose fijos los valores de c y a. Estudiar el caso límite $b \longrightarrow \infty$.

IV) Estudiar el comportamiento de las estrategias óptimas cuando a crece manteniéndose fijos los valores de b y c. Estudiar el caso límite $a \longrightarrow \infty$.

Ramón Alonso Sanz
E.T.S.I. Agrónomos
Madrid

REGRESIÓN ORTOGONAL Y COMPONENTES PRINCIPALES

J. ALBERTO MARTÍNEZ ARNAIZ*

Escuela de Empresariales. Bilbao

In this work the Principal Components Analysis is presented, starting from the orthogonal regression plane. On this basis, the data reduction technique is exposed in the three-dimensional case. Finally, the correlation matrix analysis is considerated, as well as its extension to p dimensions.

INTRODUCCIÓN

Parece que el objetivo final al que debe apuntar la enseñanza de la Estadística en una Escuela de Empresariales se podría formular como sigue: capacitación al futuro diplomado para realizar, mediante un Ordenador y el software estadístico adecuado, el análisis de los datos reales existentes en su empresa y en el ámbito económico en el que ésta se encuentra inmersa.

Si se entra en el detalle de objetivos concretos, se daría un consenso amplio en torno a la utilidad del Análisis de Componentes Principales mediante el programa SPAD. En relación con el dilema en torno a si se opta por una presentación rigurosa de tal técnica, o bien se prefiere ofrecer al alumno los conocimientos precisos para saber interpretar las "salidas" que suministra el Ordenador, cabría preguntarse, ¿existe alguna opción intermedia?

Este trabajo pretende defender precisamente una posible solución de compromiso, consistente en:

- exponer al alumno, ya familiarizado con la regresión convencional, el plano de regresión ortogonal.

*J. Alberto Martínez Arnaiz. Doctor en Ciencias Económicas. Profesor de Estadística. Escuela de Empresariales. Bilbao.

– a continuación, presentar la técnica de reducción de rango en el caso tridimensional.

– por último, tratar el caso del análisis de la matriz de correlación, así como su extensión a p dimensiones.

2. PLANO DE REGRESIÓN ORTOGONAL Y VARIANZA RESIDUAL

La regresión convencional exige la separación de las variables en dos clases: la variable a explicar por un lado, y los regresores por otro. La regresión ortogonal, en cambio, estudia el conjunto de variables en un solo bloque: todas las variables son a la vez explicativas y explicadas.

El objetivo de la regresión ortogonal (en $\mathbb{R}^3$) consiste en ajustar un plano a la nube de puntos. El criterio utilizado es el mínimo cuadrático; la diferencia con la regresión convencional radica en que el error de regresión ya no se define como diferencia entre valor predicho y valor observado.

Sea (x, y, z) una variable tridimensional, y $M_0(x_0, y_0, z_0)$ una de las observaciones de la misma. Sea, además,

$$\alpha x + \beta y + \gamma z + \delta = 0$$

la ecuación del plano de regresión ortogonal que deseamos obtener. Impondremos al plano la restricción:

$$\alpha^2 + \beta^2 + \gamma^2 = 1$$

Pues bien, el error de regresión se define aquí como la distancia del punto observado al plano:

$$e_0 = |\alpha x_0 + \beta y_0 + \gamma z_0 + \delta|$$

El criterio de obtención del plano se establecerá del siguiente modo : "el plano de regresión ortogonal es aquel cuya media de cuadrados de distancias a los puntos observados sea mínima".

Si suponemos que hay m observaciones, la función a minimizar será:

$$\frac{1}{m}\sum_{i=1}^{m}(e_i)^2 = \frac{1}{m}\sum(\alpha x_i + \beta y_i + \gamma z_i + \delta)^2, \quad i = (1, 2, \ldots, m)$$

función que al incluir la restricción impuesta se transforma en:

$$L = \frac{1}{m}\sum(\alpha x_i + \beta y_i + \gamma z_i + \delta)^2 - \mu(\alpha^2 + \beta^2 + \gamma^2 - 1)$$

que, a su vez, derivando con respecto a δ da lugar a la condicion necesaria:

$$\begin{aligned}\frac{1}{m}\sum 2(\alpha x_i + \beta y_i + \gamma z_i + \delta) = 0 \Rightarrow \\ \frac{1}{m}\sum(\alpha x_i + \beta y_i + \gamma z_i + \delta) = 0 \Rightarrow \\ \alpha\overline{x} + \beta\overline{y} + \gamma\overline{z} + \delta = 0\end{aligned}$$

Esta última expresión prueba que el plano buscado pasa por el centro de gravedad de la nube:

$$(\overline{x}, \overline{y}, \overline{z})$$

De esta primera conclusión extraemos una recomendación operativa:

- en primer lugar, debemos centrar las variables (restar de cada una de ellas su media)
- en segundo lugar, obtendremos el plano de regresión ortogonal que pasa por el origen de coordenadas.

El resultado que acabamos de obtener nos invita a replantear ligeramente nuestro problema teórico. Supondremos que operamos sobre variables centradas, y nuestro objetivo consistirá en hallar el plano de regresión ortogonal que pase por el origen de coordenadas.

Sean, (x, y, z) una variable centrada tridimensional, y $M_0(x_0, y_0, z_0)$ una de las observaciones de la misma. Sea, además,

$$\alpha x + \beta y + \gamma z = 0$$

la ecuación del plano de regresión ortogonal que deseamos obtener. Impondremos al plano la restricción:

$$\alpha^2 + \beta^2 + \gamma^2 = 1$$

El error de regresión se define como la distancia del punto observado al plano:

$$e_0 = |\alpha x_0 + \beta y_0 + \gamma z_0|$$

La función a minimizar será:

$$\frac{1}{m}\sum_{i=1}^{m}(e_i)^2 = \frac{1}{m}\sum_{i=1}^{m}(\alpha x_i + \beta y_i + \gamma z_i)^2, \quad i = 1, 2, \ldots, m)$$

función que al incluir la restricción impuesta se transforma en:

$$L = \frac{1}{m}\sum_{i=1}^{m}(\alpha x_i + \beta y_i + \gamma z_i)^2 - \mu(\alpha^2 + \beta^2 + \gamma^2 - 1)$$

que, a su vez, da lugar a las condiciones necesarias de mínimo:

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}2(\alpha x_i + \beta y_i + \gamma z_i)x_i - 2\mu\alpha &= 0 \\
\frac{1}{m}\sum_{i=1}^{m}2(\alpha x_i + \beta y_i + \gamma z_i)y_i - 2\mu\beta &= 0 \\
\frac{1}{m}\sum_{i=1}^{m}2(\alpha x_i + \beta y_i + \gamma z_i)z_i - 2\mu\gamma &= 0 \\
-(\alpha^2 + \beta^2 + \gamma^2 - 1) &= 0
\end{aligned}$$

Las tres primeras ecuaciones son equivalentes a:

$$\begin{aligned}
\alpha(s_x)^2 + \beta s_{xy} + \gamma s_{xz} &= \mu\alpha \\
\alpha(s_{xy})^2 + \beta(s_y)^2 + \gamma s_{yz} &= \mu\beta \\
\alpha(s_{xz})^2 + \beta s_{yz}^2 + \gamma(s_z)^2 &= \mu\gamma
\end{aligned}$$

que, en expresión matricial:

$$\begin{bmatrix} s_x^2 & s_{xy} & s_{xz} \\ s_{xy} & s_y^2 & s_{yz} \\ s_{xz} & s_{yz} & s_z^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \mu$$

nos indican que $(\alpha\ \beta\ \gamma)$ forman un vector propio de la matriz de covarianzas. ¿ Cuál de los tres ? Para resolver esta duda es preciso volver a la función media de cuadrados de los errores:

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}(e_i)^2 &= \frac{1}{m}\sum_{i=1}^{m}(\alpha x_i + \beta y_i + \gamma z_i)^2 = \\
&= \alpha^2(s_x)^2 + \beta^2(s_y)^2 + \gamma^2(s_z)^2 + 2\alpha\beta s_{xy} + 2\alpha\gamma s_{xz} + 2\beta\gamma s_{yz} = \\
&= [\alpha \quad \beta \quad \gamma] \begin{bmatrix} s_x^2 & s_{xy} & s_{xz} \\ s_{xy} & s_y^2 & s_{yz} \\ s_{xz} & s_{yz} & s_z^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \mu
\end{aligned}$$

Ahora bien, si la función que se quiere minimizar equivale a μ, de los tres valores propios posibles μ_1 μ_2 μ_3 debemos seleccionar el mínimo. Convengamos en que este sea μ_3. En este punto podemos enunciar la solución al problema que nos hemos planteado: el plano (que pase por el origen de coordenadas) de regresión ortogonal que mejor se ajusta a una nube de puntos (cuyo centro de gravedad coincida con el origen) tiene por coeficientes los elementos del vector propio de la matriz de covarianzas ($\alpha 3$ β_3 γ_3) correspondientes al valor propio mínimo μ_3.

3. COMBINACIÓN LINEAL ÓPTIMA DE (x, y, z). VARIANZA EXPLICADA

Se ha dicho que la regresión ortogonal es un instrumento del análisis de interdependencias. La utilización de esta técnica se realiza en forma de reducción de rango: se trata de sustituir las tres variables (x, y, z) por dos, siendo éstas combinación lineal de aquellas tres.

El vector $(\alpha_3 \quad \beta_3 \quad \gamma_3)$ nos informa de la dirección perpendicular al plano de regresión ortogonal. Los otros dos vectores propios de la matriz de covarianzas (simétrica real), correspondientes a los valores propios $\mu_1 \quad \mu_2$ (supondremos $\mu_1 > \mu_2 > \mu_3$),

$$(\alpha_1 \quad \beta_1 \quad \gamma_1) \qquad (\alpha_2 \quad \beta_2 \quad \gamma_2)$$

conforman con aquel una matriz ortogonal $\mathbf{P}$. Esta matriz $\mathbf{P}$ permite pasar a un nuevo sistema de ejes coordenados, mediante una rotación que pivota sobre el origen. Las nuevas coordenadas se obtendrán del siguiente modo:

$$\begin{aligned} F_1 &= \alpha_1 x + \beta_1 y + \gamma_1 z \\ F_2 &= \alpha_2 x + \beta_2 y + \gamma_2 z \\ e &= \alpha_3 x + \beta_3 y + \gamma_3 z \end{aligned}$$

Esta tercera coordenada mide la distancia de los puntos de la nube al plano de regresión ortogonal. Las dos primeras coordenadas corresponden a las proyecciones de los puntos de la nube sobre el citado plano.

Pues bien, nuestra técnica de reducción de rango propone precisamente a F_1, F_2 como combinaciones lineales simplificadoras de la variable tridimensional, en tanto que la combinación lineal e quedaría como elemento residual.

Teniendo en cuenta que

$$\overline{x} = \overline{y} = \overline{z} = 0$$

se prueba rapidamente que la media de F_1 es cero. Análogamente se tiene que la media de F_2 es cero y que la media de los residuos es igualmente cero.

La varianza de la combinación lineal F_1 (supondremos que todos los vectores propios están normalizados) resulta ser:

$$
\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}(F_{1i})^2 &= \frac{1}{m}\sum_{i=1}^{m}(\alpha_1 x_i + \beta_1 y_i + \gamma_1 z_i)^2 = \\
&= (\alpha_1)^2(s_x)^2 + (\beta_1)^2(s_y)^2 + (\gamma_1)^2(s_z)^2 + \\
&+ 2\alpha_1\beta_1 s_{xy} + 2\alpha_1\gamma_1 s_{xz} + 2\beta_1\gamma_1 s_{yz} = \\
&= \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 \end{bmatrix} \begin{bmatrix} s_x^2 & x_{xy} & x_{xz} \\ s_{xy} & s_y^2 & s_{yz} \\ x_{xz} & s_{yz} & x_z^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \gamma_1 \end{bmatrix} = \mu_1
\end{aligned}
$$

es decir, el mayor valor propio.

De igual modo se puede establecer que la varianza de la combinación lineal F_2 es el valor propio μ_2; y también que la varianza residual es el valor propio μ_3.

Los valores de F_1 están incorrelados con los residuos. En efecto, la covarianza entre ambas variables será:

$$
\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}F_{1i}e_i &= \frac{1}{m}\sum_{i=1}^{m}(\alpha_1 x_i + \beta_1 y_i + \gamma_1 z_i)(\alpha_3 x_i + \beta_3 y_i + \gamma_3 z_i) = \\
&= \alpha_1\alpha_3(s_x)^2 + \beta_1\beta_3(s_y)^2 + \gamma_1\gamma_3(s_z)^2 + (\alpha_1\beta_3 + \alpha_3\beta_1)s_{xy} + \\
&+ (\alpha_1\gamma_3 + \alpha 3\gamma_1)s_{xz} + (\beta_1\gamma_3 + \beta_3\gamma_1)s_{yz} = \\
&= \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 \end{bmatrix} \begin{bmatrix} s_x^2 & x_{xy} & x_{xz} \\ s_{xy} & s_y^2 & s_{yz} \\ x_{xz} & s_{yz} & x_z^2 \end{bmatrix} \begin{bmatrix} \alpha_3 \\ \beta_3 \\ \gamma_3 \end{bmatrix} = \\
&= \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 \end{bmatrix} \begin{bmatrix} \alpha_3 \\ \beta_3 \\ \gamma_3 \end{bmatrix} \mu_3 = 0
\end{aligned}
$$

Del mismo modo se establece la incorrelación entre la variable F_2 y los residuos, o entre las variables F_1 y F_2.

Designaremos combinación lineal óptima a F_1 por ser la de máxima varianza; a F_2 la calificaremos de combinación lineal subóptima; por último, la combinación lineal e se denominará residual.

La parte de varianza explicada por el par (F_1, F_2), es decir, por el plano de regresión ortogonal, se calculará mediante:

$$\frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + \mu_3}$$

Los planos formados por (F_2, e) o por (F_3, e) también contienen información interesante; la parte de varianza explicada será en cada uno de los casos:

$$\frac{\mu_1 + \mu_3}{\mu_1 + \mu_2 + \mu_3} \qquad \frac{\mu_2 + \mu_3}{\mu_1 + \mu_2 + \mu_3}$$

4. UN EJEMPLO

Consideremos la base de datos[1],

i	x	y	z
1	-16.9680	-7.5500	-15.3420
2	-18.3820	-4.7200	-13.3400
3	-18.3820	-8.4920	-12.0080
...	...	...	...
166	16.2610	13.4450	17.0120
167	20.5030	12.0290	13.0100
168	16.2610	12.5020	17.3450
169	20.5030	11.0860	13.3430

conformada por 169 puntos del espacio $\mathbb{R}^3$ que son parte de la estructura de un avión (figura 1).

[1] Los datos están disponibles como fichero "avion.dat" en: anonymous ftp **from** porthos.bio.ub.es **directory**: /pub/multicua.

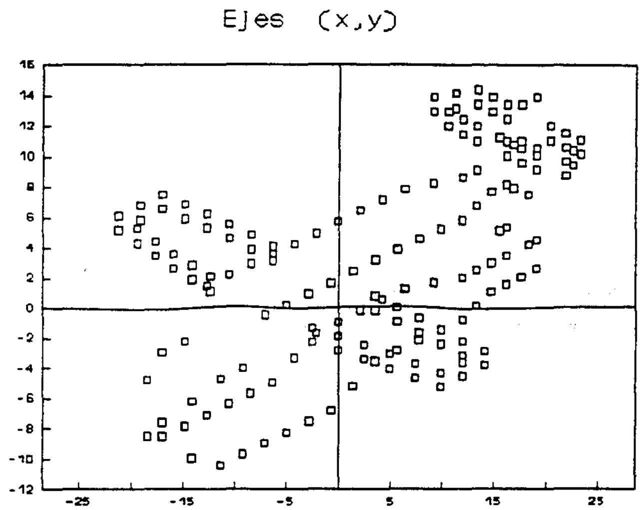

Figura 1

El análisis de regresión ortogonal ofrece los siguientes resultados:

– *vector de medias:*

$$\begin{bmatrix} 3.723 & 3.185 & 2.827 \end{bmatrix}$$

– *matriz de covarianzas:*

$$\begin{bmatrix} 163.7 & 39.4 & 63.5 \\ 39.4 & 40.0 & 59.8 \\ 63.5 & 59.8 & 137.2 \end{bmatrix}$$

– *raíces propias:*

$$\begin{bmatrix} 239.1 & 90.9 & 10.9 \end{bmatrix}$$

– *vectores propios (en filas):*

$$\begin{bmatrix} -.703 & -.329 & -.631 \\ .708 & -.236 & -.666 \\ -.070 & .915 & -.398 \end{bmatrix}$$

– *coordenadas sobre nueva base:*

i	F_1	F_2	e
1	29.533	−.014	−1.133
2	28.334	−3.014	.758
3	28.733	−3.013	−3.223
...	...	...	...
167	−21.132	−2.992	2.857
167	−21.125	3.010	2.858
168	−21.033	−2.991	1.862
169	−21.025	3.010	1.863

El plano principal (F_1, F_2), que pone de manifiesto la estructura básica del avión, es decir, fuselaje y alas (figura 2) explica el 96.8% de la varianza, en efecto:

$$\frac{239.1 + 90.9}{239.1 + 90.9 + 10.9} = 0.968$$

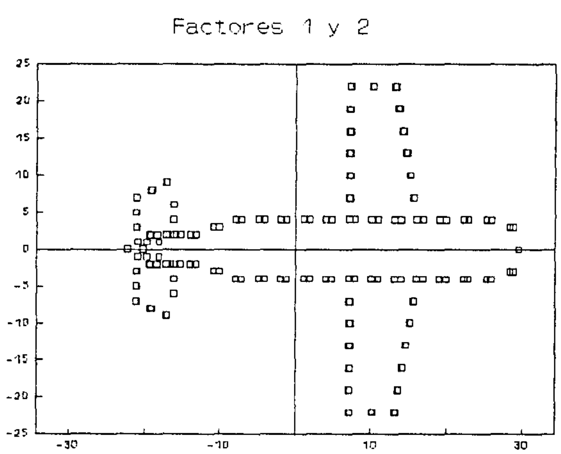

Figura 2

El plano de la figura 3 (F_1, e) muestra el perfil del avión y explica el 73.3% de la varianza.

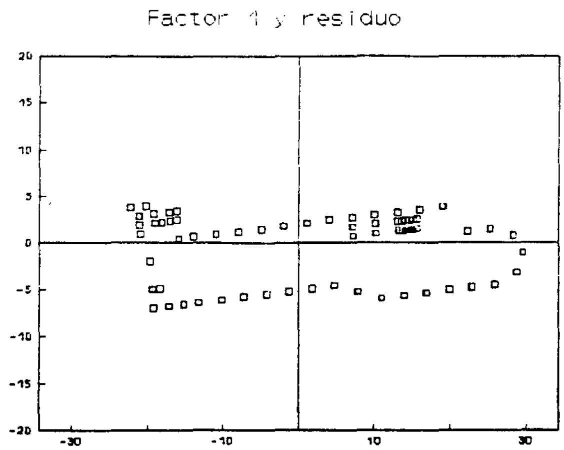

Figura 3

5. REGRESIÓN ORTOGONAL Y ANÁLISIS EN COMPONENTES PRINCIPALES

La técnica expuesta coincide plenamente con el análisis de componentes principales. Conviene, si acaso, aclarar que nuestro trabajo se ha realizado sobre datos centrados y por tanto se ha analizado la matriz de covarianzas, en tanto que el ACP habitual (conocido como normado) se suele realizar por motivos de igualación de escalas, sobre datos tipificados, y consiguientemente se basa en el análisis de la matriz de correlación.

El resumen de operaciones necesarias para realizar un ACP normado sobre una base de datos $p-$dimensional sería:

- tipificación de los datos: $t_1, t_2, \ldots, t_p$
- diagonalización de la matriz de correlación (valores y vectores propios):

$$\mu_1 > \mu_2 > \mu_3 > \cdots > \mu_p$$

$$
\begin{array}{lcl}
v_1 & : & (\alpha_1 \quad \beta_1 \quad \gamma_1 \quad \cdots\cdots) \\
v_2 & : & (\alpha_2 \quad \beta_2 \quad \gamma_2 \quad \cdots\cdots) \\
v_3 & : & (\alpha_3 \quad \beta_3 \quad \gamma_3 \quad \cdots\cdots) \\
 & & \cdots\cdots\cdots\cdots\cdots\cdots \\
v_p & : & (\alpha_p \quad \beta_p \quad \gamma_p \quad \cdots\cdots)
\end{array}
$$

- se puede, entonces, reducir el rango; supongamos que se decide sustituir las p variables por 3 combinaciones lineales de aquellas:

$$
\begin{array}{lcl}
F_1 & = & \alpha_1 t_1 + \beta_1 t_2 + \gamma_1 t_3 + \cdots \\
F_2 & = & \alpha_2 t_1 + \beta_2 t_2 + \gamma_2 t_3 + \cdots \\
F_3 & = & \alpha_3 t_1 + \beta_3 t_2 + \gamma_3 t_3 + \cdots
\end{array}
$$

- la parte de varianza explicada por estas tres combinaciones lineales se calcularía mediante:

$$
\frac{\mu_1 + \mu_2 + \mu_3}{\mu_1 + \mu_2 + \mu_3 + \cdots + \mu_p}
$$

- aunque los vectores propios nos informan sobre la relación entre las variables y las combinaciones lineales, es muy interesante conocer las correlaciones entre unas y otras:

$$
\begin{array}{ccc}
\text{corr }(t_1, F_1) & \text{corr }(t_1, F_2) & \text{corr }(t_1, F_3) \\
\text{corr }(t_2, F_1) & \text{corr }(t_2, F_2) & \text{corr }(t_2, F_3) \\
\cdots\cdots & \cdots\cdots & \cdots\cdots \\
\text{corr }(t_p, F_1) & \text{corr }(t_p, F_2) & \text{corr }(t_p, F_3)
\end{array}
$$

- como las combinaciones lineales están incorreladas entre sí, por simple suma de los cuadrados de los coeficientes de correlación se llega a saber la parte de varianza explicada de cada variable:

$$
\begin{array}{ll}
t_1 : & \text{corr}^2(t_1, F_1) + \text{corr}^2(t_1, F_2) + \text{corr}^2(t_1, F_3) \\
t_2 : & \text{corr}^2(t_2, F_1) + \text{corr}^2(t_2, F_2) + \text{corr}^2(t_2, F_3) \\
\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\
t_p : & \text{corr}^2(t_p, F_1) + \text{corr}_2(t_p, F_2) + \text{corr}^2(t_p, F_3)
\end{array}
$$

BIBLIOGRAFÍA

[1] **Cuadras, C.M.** (1991) "Ejemplos y aplicaciones insólitas en regresión y correlación". *Qüestiió*, **15**, 367–382.

[2] **Martin-Guzman, Martin Pliego** (1987). *Curso básico de Estadística Económica.* Editorial AC.

SOLUCIONS ALS PROBLEMES PROPOSATS AL VOLUM 17. N° 2

PROBLEMA N° 49

1) Si la distribución de X es uniforme en $(0,\beta)$, entonces

$$\begin{aligned} p(x) &= 1/\beta \quad \text{si } 0 < x < \beta \\ p(x) &= 0 \qquad \text{en caso contrario,} \end{aligned}$$

luego

$$\Gamma_\beta = \inf_{x\in(0,\beta)} p(x) = 1/\beta$$

La varianza de X es $\sigma_x^2 = \beta^2/12$, así que

$$\beta^2\Gamma_\beta = \beta^2/\beta = \sqrt{12}\sigma_x.$$

2) Sea $f(x)$ la densidad de otra variable X (que vamos a suponer con distribución absolutamente continua) y $p(x)$ la densidad de la variable U uniforme en $(0,\beta)$. Sea

$$\alpha = \inf_{-\infty<x<+\infty} \left\{ \frac{f(x)}{p(x)} \right\}.$$

α es una constante tal que $0 \leq \alpha \leq 1$. Se verifica por tanto:

(1) $$f(x) \geq \alpha \cdot p(x) \qquad -\infty < x < +\infty$$

Si X_1, X_2 son independientes y con la misma distribución que X, se tiene:

$$\begin{aligned} \mathrm{E}\left[(X_1 - X_2)^2\right] &= \mathrm{E}\left[((X_1-\mu)-(X_2-\mu))^2\right] = \\ &= \mathrm{E}\left[(X_1-\mu)^2\right] + \mathrm{E}\left[(X_2-\mu)^2\right] - \\ &- 2\,\mathrm{E}\left[(X_1-\mu)(X_2-\mu)\right] = 2\sigma_x^2, \end{aligned}$$

siendo $\mu = \mathrm{E}(X)$, $\sigma_x^2 = \mathrm{var}(X)$. De forma análoga

$$\mathrm{E}\left[(U_1 - U_2)\right]^2 = 2\sigma_U^2 = 2\beta^2/12.$$

Teniendo en cuenta (1), obtenemos las desigualdades:

$$
\begin{aligned}
2\sigma_x^2 &= \mathrm{E}\left[(X_1 - X_2)^2\right] = \int\int (x_1 - x_2)^2 f(x_1) f(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \geq \\
&\geq \int\int (x_1 - x_2)^2 \alpha p(x_1) \alpha p(x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 = \alpha^2 \cdot \mathrm{E}\left[(U_1 - U_2)^2\right] = \\
&= 2\alpha^2 \sigma_U^2, \\
\sigma_x^2 &\geq \alpha^2 \beta^2 / 12, \\
\sqrt{12}\sigma_x &\geq \alpha \cdot \beta
\end{aligned}
\tag{2}
$$

Si $p(x) = 1/\beta$ para $0 < x < \beta$, entonces

$$
\alpha = \inf_{-\infty < x < +\infty} \left\{ \frac{f(x)}{p(x)} \right\} = \inf_{x \in (0,\beta)} \{\beta \cdot f(x)\} = \beta \cdot \Gamma_\beta
$$

Sustituyendo en (2) obtenemos finalmente

$$
\sqrt{12}\sigma_x \geq \beta^2 \Gamma_\beta . \tag{3}
$$

La igualdad en (3) sólo se cumple si $X \equiv U$, pues si X no es U, entonces $\alpha < \beta\Gamma_\beta$.

Notas:

a) Una generalización válida para cualquier par de variables X, Y es

$$
0 \leq \inf_{-\infty < x < +\infty} \left\{ \frac{g(x)}{f(x)} \right\} \leq \frac{\sigma_y}{\sigma_x},
$$

donde X tiene densidad $f(x)$ y varianza σ_x^2, Y tiene densidad $g(x)$ y varianza $\sigma_y^2 > 0$.

b) En la resolución del apartado 2), se ha seguido la solución enviada por C.R. Rao.

C.M. Cuadras
Universitat de Barcelona

PROBLEMA N° 50

La probabilidad buscada, p, viene dada por:

$$p = \int_{(\mathbb{R}^+)^3} (2\pi)^{-\frac{3}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}X'\Sigma^{-1}X\Big) \, (\mathrm{d}X)$$

donde X es un vector (3×1). Efectuando el cambio de variable $X = \Sigma^{-1/2}Y$, donde $\Sigma^{-1/2}$ es una matriz (3×3) e Y un vector (3×1), resulta:

$$p = (2\pi)^{-\frac{3}{2}} \int_A \exp\Big(-\frac{1}{2}Y'Y\Big) \, (\mathrm{d}Y)$$

donde $A \subset \mathbb{R}^3$ es el triedro dado por:

$$A = \{y \in \mathbb{R}^3\colon \quad v_i'y \geq 0 \qquad i = 1,2,3\}$$

con

$$\Sigma^{1/2} = (v_1, v_2, v_3) = \begin{pmatrix} v_1' \\ v_2' \\ v_3' \end{pmatrix}$$

es decir los v_i son los vectores columna (3×1) de la matriz $\Sigma^{1/2}$.

Dicha región está pues limitada por tres planos cuyos vectores directores son v_1, v_2 y v_3.

Efectuando un cambio a coordenadas polares obtendremos:

$$\begin{aligned} p &= (2\pi)^{-\frac{3}{2}} \int_0^\infty r^2 \, \mathrm{e}^{-\frac{r^2}{2}} \left(\int_B \mathrm{d}\sigma\right) \mathrm{d}r = \\ &= (2\pi)^{-\frac{3}{2}} \int_0^\infty \sqrt{2t} \, \mathrm{e}^{-t} \, \mathrm{d}t \left(\int_B \mathrm{d}\sigma\right) = \\ &= \frac{1}{2\pi^{\frac{3}{2}}} \Gamma\left(\frac{3}{2}\right) \left(\int_B \mathrm{d}\sigma\right) = \frac{1}{4\pi} \int_B \mathrm{d}\sigma \end{aligned}$$

Donde $\mathrm{d}\sigma$ es el elemento de superficie de una esfera unidad tridimensional y B es la porción de superficie definida por:

$$B = \{z \in \mathbb{R}^3\colon \|z\| = 1, \quad v_i'z \geq 0 \qquad i = 1,2,3\}$$

es decir, se trata de un "triángulo esférico" sobre la superficie de la esfera unidad. La probabilidad p será pues el "tanto por uno" que representa la superficie de B respecto de superficie total de dicha esfera, 4π.

Los ángulos de dicho triángulo esférico α, β, γ dependen de los ángulos entre los vectores directores v_1, v_2, v_3. Obsérvese que estos verifican, al ser las columnas de $\Sigma^{1/2}$,

$$< v_i, v_i >= 1 \qquad < v_1, v_2 >= \sigma_{12} \qquad < v_1, v_3 >= \sigma_{13} \qquad < v_2, v_3 >= \sigma_{23}$$

Por tanto los ángulos serán:

$$\alpha = \pi - \text{ arc cos } \sigma_{13} \qquad \beta = \pi - \text{ arc cos } \sigma_{23} \qquad \gamma = \pi - \text{ arc cos } \sigma_{12}$$

La superficie del triángulo esférico (sobre una esfera unitaria), será:

$$\mathcal{S} = \alpha + \beta + \gamma - \pi$$

por tanto:

$$\begin{aligned} p &= P(X_1 \geq 0, X_2 \geq 0, X_3 \geq 0) = \\ &= \frac{1}{4\pi}(2\pi - \text{ arc cos } \sigma_{12} - \text{ arc cos } \sigma_{13} - \text{ arc cos } \sigma_{23}) = \\ &= \frac{1}{2} - \frac{1}{4\pi}\sum_{i<j=2}^{3} \text{ arc cos } \sigma_{ij} \end{aligned}$$

o en términos de arc sen,

$$p = \frac{1}{8} + \frac{1}{4\pi}\left(\sum_{i<j=2}^{3} \text{ arc sen } \sigma_{ij}\right)$$

Josep M. Oller
Universitat de Barcelona

COMENTARI DE LLIBRES

C. Radhakrishna Rao

ESTADÍSTICA Y VERDAD: Aprovechando el Azar

Traducción de C.M. Cuadras y J.M. Oller

Publicaciones y Promociones Universitarias.
Barcelona, 1994.
XXIX + 219 pp + tablas y figuras.

Se trata de la versión española de "Statistics and Truth" de C.R. Rao, publicado por primera vez en 1989 por CSIR-New Delhi, India. Un amplio comentario de esta primera versión puede encontrarse en *Qüestiió*, Volumen **14**, 1990, 209–213.

Los temas cubiertos en esta versión han sido reorganizados para proporcionar una visión coherente de la Estadística como una disciplina separada, resaltando su significación filosófica y lógica, así como sus aspectos técnicos y aplicados.

Escrito por uno de los más destacados estadísticos con experiencia en diversos campos de aplicación de la estadística, el libro versa sobre los aspectos filosóficos y metodológicos del tratamiento de la información, recogida y análisis de datos, para poder abordar con perspicacia los problemas que surgen en la investigación científica, la política seguida por los gobiernos o, simplemente, en la toma de decisiones en la vida diaria.

El autor disipa las dudas acerca de la opinión de que el azar es el producto de nuestra ignorancia, que hace imposible el poder realizar predicciones acertadas. Ilustra cómo nuestro pensamiento ha cambiado gracias a la cuantificación de la incertidumbre, probando que el azar ya no es un impedimento, sino un modo de expresión de nuestro conocimiento.

En realidad, el azar puede ser creativo y ayudarnos en la búsqueda de la verdad. Basándose en el estudio de las leyes del azar y mediante ejemplos y aplicaciones, el libro demuestra con elocuencia que la estadística es la ciencia, la tecnología y el arte de extraer información de los datos. Se subraya cómo las ideas estadísticas juegan un papel fundamental en la investigación científica y en otras áreas, incluso cuando todavía no era reconocida como una disciplina separada.

El libro pone de manifiesto cómo la estadística se está desarrollando como una versátil, potente e insustituible herramienta, utilizada en diversos campos de la actividad humana, tales como la literatura, el derecho, la industria, la arqueología y la medicina.

Esta versión española puede ser de gran utilidad para los profesores y alumnos de estadística, así como usuarios y público en general, por la cantidad de comentarios metodológicos y prácticos, y la excelente selección de ejemplos que contiene.

M.C.P.

Judith M. Tanur, Frederick Mosteller, William H. Kruskal, Erich L. Lehmann, Richard F. Link, Richard S. Pieters, Gerald R. Rising (editores).

LA ESTADÍSTICA: UNA GUIA DE LO DESCONOCIDO

Traducción a cargo de un grupo de profesores pertenecientes a la Sociedad de Estadística e Investigación Operativa de España, bajo la coordinación de Daniel Peña.

Alianza Editorial, S.A.
Madrid, 1992.
XXVI + 407pp + tablas + gráficos.

El libro es una traducción al castellano de "Statistics: A guide to the unknown /3r.Ed.", publicada en 1989 por Brooks/Cole.

Como en la propia obra se indica, el libro está dirigido a lectores sin conocimientos especiales de estadística, probabilidad o matemáticas, describiendo con claridad las aplicaciones de estas herramientas en campos profesionales tan diferentes como la medicina, biología, economía, ciencias políticas y sociología, ingeniería, administración de empresas, astronomía,etc. Apoyándose en casos reales concretos, los artículos ilustran paso a paso cómo contribuye la estadística a resolver importantes problemas en estos ámbitos. Exponen el diseño de los experimentos, tratan las dificultades que conlleva hacer inferencias a partir de datos imperfectos y proponen material de trabajo al lector interesado.

La Estadística: una Guia de lo Desconocido es fruto del esfuerzo conjunto de la Asociación Americana de Estadística y el Consejo Nacional de Profesores de Matemáticas de Estados Unidos. Estructurado en cuatro partes (Nuestro Mundo Biológico, Nuestro Mundo Político, Nuestro Mundo Social y Nuestro Mundo Físico), la obra contiene 29 capítulos sobre temas tan diferentes como la distribución al azar de una vacuna en un experimento de sanidad a gran escala, los efectos de aumentar los impuestos sobre el consumo del tabaco, la discriminación en el empleo, el significado de las palabras, estadística en el deporte, estimar la probabilidad de un terremoto, etc. Una colección de problemas o cuestiones dirigidas al lector aumentan el interés formativo de la obra.

Contiene una clasificación de los artículos por el origen de los datos, y otra por los procedimentos estadísticos utilizados. Por sus numerosos ejemplos comentados en un estilo llano y ameno, ilustrados con tablas y figuras, es una obra muy recomendable para profesores, alumnos, usuarios y profesionales de la Estadística.

M.C.P.

NOVETATS DE SOFTWARE

La introducció de la nova secció de "NOVETATS DE SOFTWARE" a la revista QÜESTIIÓ es fa amb la finalitat de promoure l'intercanvi d'informació relacionada amb programes d'ordinador disponibles, destinats a la implementació de metodologia estadística, d'informàtica o d'investigació operativa.

A causa de l'important creixement que ha experimentat darrerament la utilització dels ordinadors a totes les àrees científiques i tècniques i, a les esmentades més amunt, en particular, hi ha un bon nombre d'investigadors que han desenvolupat un software propi, l'existència del qual és desconeguda, de vegades, per a molts lectors que el podrien aprofitar. Per això, creiem que és convenient i útil fer-lo conèixer mitjançant aquesta revista, amb el benentès que només actuaria com a mitjà de difusió.

Per tal d'uniformitzar la descripció del software, adjuntem una butlleta que ha de ser omplenada i tramesa a l'editorial de QÜESTIIÓ.

Amb tota certesa, la vostra col·laboració serà d'utilitat per a molts lectors als qui facilitarà el treball i que, alhora, podran ajudar els autors del programes suggerint-los possibles millores.

Nom del programa: Programas en batch en el entorno μTSP

Area/àrees d'aplicació (Estadística, Sistemes, etc.): Econometría, Estadística, Regresión, Series de Tiempo.

Descripció del software:

- Llenguatge: μTSP
- Ordinador/s: IBM PC y compatibles.
- Sistema operatiu: PC DOS 3.0, MS DOS 3.0 y posteriores.

Està disponible en els suports següents:

Floppy disk/diskette. Assenyaleu:

Mida: 3.5 Densitat: DD una dues cares

Distribuït per:

Micro TSP: Quantitative Micro Software. 4521 Campus Drive, Suite 336
Irvine, California 92715. TFNO (714) 856-3368 FAX (714) 856-2044
Programas para ordenar submuestras:
Rafaela Dios Palomares. E.T.S.I. Agrónomos y Montes.
Universidad de Córdoba.

Configuració mínima de hardware requerida:

- De 512K a 640K de Memoria de Acceso (memoria DOS). La memoria expandida adicional (EMS) será usada para la gestión de gráficos.
- Cualquier monitor permite la salida de Textos. Para la salida de gráficos son válidos CGA, EGA o VGA compatible IBM o un Hércules compatible monocromo.

Requereix l'ensinistrament de l'usuari: Cualquier usuario conocedor de los métodos que se aplican en el paquete puede trabajar fácilmente con μTSP, ayudándose del manual y del sistema de menús en modo interactivo. El diseño de macros requiere el conocimiento de la técnica elemental de programación. La utilización de los programas en batch que se presentan, es inmediata con la documentación que se adjunta.

Documentació

- Micro TSP USER'S MANUAL. Q.M.S. (Ver distribución).
- Artículo que se presenta por Rafaela Dios.

Llistat, font disponible: Sólo están disponibles los listados de las macros que se presentan.

Grau de desenvolupament: Producto acabado y verificado.

Es fa servir aquest software normalment? Si para el paquete μTSP.

En cas afirmatiu

des de quan? 1982
a quants llocs? diversas universidades y empresas.

L'autor d'aquest software està disponible per atendre les preguntes dels usuaris? La autora de los programas en batch, SI.

Descripció del que fa l'esmentat software: (200 paraules aproximadament).

El paquete μTSP, está dedicado al análisis econométrico de series tanto de corte transversal como de tipo cronológico, aplicando la metodología más reciente en el campo de la especificación, estimación y verificación del modelo, así como de la predicción. La opción de trabajar en batch permite la creación de programas y otorga la posibilidad al usuario de ejecutar su propio software para la realización de ciertas tareas de docencia, investigación y gestión que serían muy laboriosas e incluso imposibles en modo interactivo. Además los programas pueden incluir no sólo comandos válidos en el entorno interactivo, sino algunos específicos que únicamente se utilizan dentro de una macro. Así, podemos comentar, por ejemplo, la posibilidad de generar bucles que iteran determinada operación cambiando el valor de los parámetros oportunos.

La metodología del trabajo consiste en editar el programa en un fichero texto y rodarlo después desde el entorno μTSP, utilizando el comando RUN.

Como aplicaciones podemos comentar la posibilidad de realizar transformaciones de datos, salidas no contempladas convencionalmente en el paquete, contrastes especiales e incluso pruebas o métodos que efectúan las últimas versiones pero que podemos montar mediante macros para ampliar una versión más antigua. El trabajo en batch ofrece también la importante opción de ejecutar programas sin requerir la intervención ni la presencia del operador.

Presentamos aquí ocho macros creadas para ordenar submuestras. Se han diseñado y puesto a punto con el fin de soslayar el problema de que el paquete μTSP, ejecuta el comando de ordenación SORT, sin atender a la estructura muestral del momento establecida por SMPL. Se contemplan los distintos tipos de datos del espacio de trabajo y la ordenación en los dos sentidos, dando también como salida los extremos de la submuestra.

Possibles usuaris: Docentes, Investigadores, Economistas, Gerentes, Empresarios...

Camps d'interès: Economía, Ciencias Biológicas, Ciencias Sociales...

Nom de l'autor/s:

De μTSP	Q.M.S. (Ver distribución)
De las macros	Rafaela Dios Palomares

Institució:

Escuela Técnica Superior de Ingenieros Agrónomos y Montes.
Departamento de Estadística. Universidad de Córdoba. España.

Adreça: Alameda del Obispo. Apdo. de Correos 3048. 14080 Córdoba.

Número de telèfon: (957) 218568 y 218479

Butlleta de subscripció a la revista **Qüestiió**

Nom i cognoms ____

____

Empresa/Institució ____

____

Adreça ____

____

Codi postal ______________ ciutat ______________________________

Tel. ____________________ Fax ____________________

Data d'expedició ______________

Signatura: DNI o NIF ________________

Desitjo subscriure'm a **Qüestiió** per a l'any 1994.
El preu de la subscripció és de 2.700 PTA.

Forma de pagament

- ☐ Transferència al compte de la Caixa de Catalunya número 6985/77, Agència 100, Comte d'Urgell 162, 08036 Barcelona
- ☐ Domiciliació bancària
- ☐ Taló nominatiu a l'Institut d'Estadística de Catalunya
- ☐ Gir postal
- ☐ En efectiu

Retornar aquesta butlleta (o una fotocòpia) a:

Qüestiió:
Institut d'Estadística de Catalunya
Via Laietana, 58
08003 Barcelona

Autorització de domiciliació bancària per al pagament de les subscripcions anuals de la revista **Qüestiió**

El sotasignat ____

autoritza el Banc/Caixa ______________________________________

Agència núm. ____________ adreça ______________________________

Codi postal ______________ ciutat _______________________________

a abonar a la revista **Qüestiió** amb càrrec al meu compte número __________ ____________________________________, les subscripcions a **Qüestiió**.

______________________ , a _____ d________________ de 19 _____

(Signatura)

El sotasignat ____

autoritza la revista **Qüestiió** a carregar al meu compte número ___________ ______________________, al Banc/Caixa ___________________________ ____

Agència núm. ____________ adreça ______________________________

Codi postal ______________ ciutat _______________________________

l'import de les tarifes vigents de les subscripcions a la revista **Qüestiió.**

______________________ , a _____ d________________ de 19 _____

(Signatura)