

# A BAYESIAN APPROACH TO DETECT INFORMATIVE OBSERVATIONS IN A REGRESSION EXPERIMENT BASED ON GENERALIZED ENTROPY MEASURES

D. MORALES AND L. PARDO

**Abstract.** In this paper we identify subsets of the data that appear to have a disproportionate influence on the estimated normal regression model in a Bayesian context. Generalized entropy measures are used to detect a set of most informative observations in a given design.

## 1. Introduction

Let  $\mathcal{C}$  be a class of statistical experiments  $Y$ . In this paper a statistical experiment is understood by the following three conditions:

- (i) a listing of possible outcomes, i.e., a measurable sample space  $(Y, \mathcal{B})$ ;
- (ii) a listing of possible elementary hypotheses (explaining theories or states of nature), i.e., a nonempty set  $\Theta$ ; and
- (iii) a correspondence assigning the random outcome, i.e., a statistical family  $P = \{P_\theta, \theta \in \Theta\}$  consisting of probability measures on  $(Y, \mathcal{B})$ .

We shall suppose that the family of probability measures is dominated by a  $\sigma$ -finite measure, so that they may be described through their density functions,

---

Received December 30, 1991; revised February 10, 1993.

(1980) A.M.S. Subject Classification: Primary 94A15, 62B10.

*Key words and phrases:* Normal regression model, Informative Observations, generalized entropy measures.

This work was partially supported by the Direccin General de Investigaci3n Cientifica y T3cnica (DGICYT) under the contract PB91-0387.

$f(y/\theta)$ , with respect to a measure  $\mu(y)$ . We shall also suppose that  $\Theta$  is endowed with a  $\sigma$ -field of subsets, usually, it is a subset of an  $n$ -dimensional Euclidean space. A prior distribution for  $\Theta$  is a probability density function  $p(\theta)$  with respect to a measure denoted by  $\lambda(\theta)$ . In this paper the dominating measure will be either Lebesgue or counting measure. Thus, in accordance with these conventional notations, we have,

$$f(y) = \int_{\Theta} f(y/\theta)p(\theta)d\lambda(\theta),$$

and by Bayes's theorem we have  $p(\theta/y) = (f(y/\theta)p(\theta))/f(y)$ .

For a prior distribution  $p(\cdot)$ , the  $(r, s)$ -entropy measure is given by

$$\mathcal{E}_r^s(p(\cdot)) = \begin{cases} H_r^s(p(\cdot)) = (2^{1-s} - 1)^{-1} \left\{ \left( \int_{\Theta} p(\theta)^r d\lambda(\theta) \right)^{\frac{s-1}{r-1}} - 1 \right\}, & r \neq 1, s \neq 1 \\ H_1^s(p(\cdot)) = (2^{1-s} - 1)^{-1} (2^{(1-s)H(p(\cdot))} - 1), & r = 1, s \neq 1 \\ H_r^1(p(\cdot)) = \frac{1}{1-r} \log_2 \left( \int_{\Theta} p(\theta)^r d\lambda(\theta) \right), & r \neq 1, s = 1 \\ H(p(\cdot)) = - \int_{\Theta} p(\theta) \log_2 p(\theta) d\lambda(\theta), & r = 1, s = 1 \end{cases}$$

for all  $r \in (0, \infty)$  and  $s \in (-\infty, \infty)$ . It is understood that all the integrals involved exist.

For any probability density function  $f(x)$  we have

$$H_1^s(f) = \lim_{r \rightarrow 1} H_r^s(f), \quad H_r^1(f) = \lim_{s \rightarrow 1} H_r^s(f) \text{ and } H(f) = \lim_{s \rightarrow 1} H_1^s(f) = \lim_{r \rightarrow 1} H_r^1(f).$$

For operational purposes, a systematic attempt to develop a generalization of Shannon's entropy (1948) was carried out by Rényi (1961),  $H_r^1(f)$ , first. Since then many other authors have studied new generalizations: Havrda and Charvát (1967),  $H_s^s(f)$ , Arimoto (1971),  $H_{1/t}^t(f)$ , and Sharma and Mittal (1975),  $H_1^s(f)$  and  $H_r^s(f)$ . Finally, a review on  $(r, s)$ -entropy measures, their importance and their applications is given in Taneja et al (1989).

After the experiment has been performed and a value  $y$  is observed, the  $(r, s)$ -entropy measure on the posterior distribution is given by  $\mathcal{E}_r^s(p(\cdot/y))$ .

The information about  $\theta$  given by the observation  $y$  of  $Y$  is defined as follows:

$$\mathcal{G}_r^s(Y, p(\cdot), y) = \mathcal{E}_r^s(p(\cdot)) - \mathcal{E}_r^s(p(\cdot/y)),$$

for all  $r \in (0, \infty)$  and  $s \in (-\infty, \infty)$ .

The average amount of  $(r, s)$ -information over  $\Theta$  provided by the experiment  $Y$ , when the prior knowledge is  $p(\theta)$ , is given by

$$\mathcal{G}_r^s(Y, p(\cdot)) = \mathcal{E}_r^s(p(\cdot)) - E_Y(\mathcal{E}_r^s(p(\cdot/y))),$$

We can easily find examples, when the measures  $\mathcal{E}_r^s(p(\cdot))$  and  $\mathcal{E}_r^s(p(\cdot/x))$  become negative. In Taneja et al. (1989) it is established that the measure  $\mathcal{G}_r^s(Y, p(\cdot))$  is nonnegative for all  $(r, s) \in \Gamma$ , where

$$\Gamma = \left\{ (r, s) / r > 0, s \geq 2 - \frac{1}{r} \right\}.$$

Furthermore,  $\mathcal{E}_r^s(p(\cdot))$  is concave for all  $(r, s) \in \Gamma$ .

In this paper we suppose that the class of all possible experiments  $\mathcal{C}$  is composed of

$$Y = \mathbf{R}^n, \Theta = \mathbf{R}^k, y_i = \sum_{j=1}^k \theta_j a_{ij} + e_i, \quad i = 1, 2, \dots, n,$$

where  $e = (e_1, e_2, \dots, e_n)$  is normally distributed with mean vector  $(0, \dots, 0)$  and precision matrix  $\sigma^{-2}I_n$ . Also suppose that  $\theta$  is normally distributed with mean vector  $\theta_0$  and precision matrix  $\sigma^{-2}P_0$ . The  $n \times k$  matrix  $A = (a_{ij})$ , with  $\text{rank}(A) = k < n$ , is called the allocation or design matrix and the rows of  $A$  are called the allocation vectors. If in an experiment a set of  $t$  observations is unavailable, where  $t$  is a positive integer such that  $(n - t) \geq k$ , the problem of measuring the corresponding loss of information arises. In the case of a limited budget for an experiment it may be wise not to collect the least informative set of observations. The number of available measures in the literature for identifying influential points is quite large. A partial list, with very readable explanations

is given in Chatterjee and Hadi (1986). Further discussions can be find in Cook and Weisberg (1980) and (1982), Cook (1986), Lawrence (1988), Ghosh (1982, 1983) and Ali (1990).

In this paper we use the  $(r, s)$ -information measure,  $\mathcal{G}_r^s(Y, p(\cdot))$ , to analyze the loss of information due to the unavailability of a set of observations. This measure can be used to identify the least informative set of observations as well as to identify subsets of the data that appear to have a disproportionate influence on the estimated parameter  $\theta$ .

**2. Unified  $(r, s)$ -entropy measures in the regression experiment model**

$$Y = A\theta + e$$

In this section we calculate the  $(r, s)$ -information measure provided by a regression experiment about the vector  $\theta$ , when the initial opinions about  $\theta$  are described by a multivariate normal density.

**Theorem 1.** *Let us consider the regression experiment  $Y = A\theta + e$ , where  $e$  has a multivariate normal density with mean vector  $(0, \dots, 0)$  and precision matrix  $\sigma^{-2}I_n$ , and where  $A$  is a known  $n \times k$  matrix. Suppose the prior knowledge on  $\theta$  is expressed by a multivariate normal with mean vector  $\theta_0$  and precision matrix  $\sigma^{-2}P_0$ . Then*

$$\mathcal{G}_r^s(Y, p(\cdot)) = \begin{cases} \frac{|P_0|^{\frac{s-1}{2}} \left(1 - |P_0^{-1}A'A + I|^{\frac{s-1}{2}}\right)}{r^{\frac{k(s-1)}{2(r-1)}} (2^{1-s} - 1) (2\pi)^{\frac{(s-1)k}{2}} \sigma^{s-1}}, & r \neq 1, s \neq 1 \\ \frac{|P_0|^{\frac{s-1}{2}} \left(|P_0^{-1}A'A + I|^{\frac{s-1}{2}} - 1\right)}{(2\pi e)^{\frac{k(s-1)}{2}} (2^{s-1} - 1) \sigma^{s-1}}, & s \neq 1, r = 1 \\ \frac{1}{2} \log_2 |P_0^{-1}A'A + I|, & s = 1 \end{cases}$$

where  $(r, s) \in \Gamma$ .

**Proof.** Given a multivariate normal distribution with mean vector  $T(k \times 1)$

and precision matrix  $C$  ( $k \times k$ ), it follows

$$\begin{aligned} &H_r^s(N(T, C)) \\ &= (2^{1-s} - 1)^{-1} \left\{ \left[ \int_{\mathbb{R}^k} ((2\pi)^{-(k/2)} |C|^{1/2} \exp(-\frac{1}{2}(x-T)'C(x-T)))^r dx \right]^{\frac{s-1}{r-1}} - 1 \right\} \\ &= (2^{1-s} - 1)^{-1} \left( |C|^{\frac{s-1}{2}} r^{-\frac{k(s-1)}{2(r-1)}} (2\pi)^{-\frac{(s-1)k}{2}} - 1 \right), \quad r \neq 1, s \neq 1 \end{aligned}$$

Since the posterior distribution (see DeGroot (1970)) is multivariate normal with precision matrix  $(A'A + P_0)\sigma^{-2}$ , we have

$$H_r^s(p(\cdot)) = (2^{1-s} - 1)^{-1} \left( |P_0\sigma^{-2}|^{\frac{s-1}{2}} r^{-\frac{k(s-1)}{2(r-1)}} (2\pi)^{-\frac{(s-1)k}{2}} - 1 \right), \quad r \neq 1, s \neq 1$$

and

$$\mathcal{G}_r^s(Y, p(\cdot)) = \frac{|P_0|^{\frac{s-1}{2}} \left( 1 - |P_0^{-1}A'A + I|^{\frac{s-1}{2}} \right)}{r^{\frac{k(s-1)}{2(r-1)}} (2^{1-s} - 1) (2\pi)^{\frac{(s-1)k}{2}} \sigma^{s-1}}, \quad r \neq 1, s \neq 1, (r, s) \in \Gamma$$

and by continuity of  $\mathcal{E}_r^s$  in  $r$  and  $s$  we obtain the proposed result after applying the L'Hôpital rule.

**Remark 1.** If  $\phi = M\theta$  is a orthogonal transformation, then  $\mathcal{G}_r^s(Y, p(\cdot))$  remains the same, whether we consider information about  $\theta$  or about  $\phi$ .

### 3. Loss of Information due to unavailable observations

Let  $A_1((n-t) \times k)$  be the allocation matrix corresponding to the situation when any set of  $t$  observations is unavailable in the experiment. Clearly, there are  $\binom{n}{t}$  possible allocation matrices. Let  $Y_1((n-t) \times 1)$  be the vector of observations corresponding to the allocation matrix  $A_1$ . Further, let  $A_2 t \times k$  be the rows in  $A$  corresponding to the vector of unavailable observations  $Y_2 (t \times 1)$ . We partition  $Y$  and  $A$  as follows

$$Y' = (Y_1', Y_2') \text{ and } A' = (A_1', A_2')$$

Since the distribution of  $Y_1$  given  $\theta$  is a multivariate normal with mean vector  $A_1\theta$  and precision matrix  $\sigma^{-2}I_{(n-t)}$ , and the prior distribution of  $\theta$  is multivariate normal with mean  $\theta_0$  and precision matrix  $\sigma^{-2}P_0$ , the distribution of  $\theta$  given  $Y_1$  is multivariate normal with parameters  $(\mu_1, \sigma^{-2}P_1)$ , where

$$\mu_1 = (P_0 + A_1' I_{(n-t)} A_1)^{-1} (P_0 \theta_0 + A_1' I_{(n-t)} Y_1)$$

and

$$P_1 = (P_0 + A_1' I_{(n-t)} A_1).$$

Note also that the marginal distribution of  $Y_1$  is multivariate normal with mean vector  $A_1\theta_0$  and covariance matrix  $V_1 = I_{n-t} + A_1 P_0^{-1} A_1'$ .

Now, we give the following definition:

**Definition 1.** A measure of loss of information due to the unavailability of the vector  $Y_2$  in the allocation matrix  $A$  based on the unified  $(r, s)$ -entropy measure is given by

$$\mathcal{L}_r^s(A_2; A_1) = \mathcal{G}_r^s(Y, p(\cdot)) - \mathcal{G}_r^s(Y_1, p(\cdot))$$

The next theorem establishes the expression of  $\mathcal{L}_r^s(A_2; A_1)$ .

**Theorem 2.** *The amount of information loss due to  $Y_2$ , the vector of unavailable observation in the given allocation matrix  $A$ , is*

$$\mathcal{L}_r^s(A_2, A_1) = \begin{cases} \frac{|V_1|^{\frac{s-1}{2}} \left(1 - |A_2 S A_2' + I_t|^{\frac{s-1}{2}}\right) |P_0|^{\frac{s-1}{2}}}{r^{\frac{k(s-1)}{2(r-1)}} (2^{1-s} - 1) (2\pi)^{\frac{(s-1)k}{2}} \sigma^{s-1}}, & r \neq 1, s \neq 1 \\ \frac{|V_1|^{\frac{s-1}{2}} \left(|A_2 S A_2' + I_t|^{\frac{s-1}{2}} - 1\right) |P_0|^{\frac{s-1}{2}}}{(2\pi e)^{\frac{k(s-1)}{2}} (2^{s-1} - 1) \sigma^{s-1}}, & s \neq 1, r = 1 \\ \frac{1}{2} \log_2 |A_2 S A_2' + I_t|, & s = 1 \end{cases}$$

where

$$S = P_1^{-1} = P_0^{-1} - P_0^{-1} A_1' V_1^{-1} A_1 P_0^{-1} \text{ and } V_1 = I_{n-t} + A_1 P_0^{-1} A_1'$$

**Proof.** First, we consider  $r \neq 1, s \neq 1$ . Then

$$\begin{aligned} \mathcal{L}_r^s(A_2; A_1) &= \mathcal{G}_r^s(Y, p(\cdot)) - \mathcal{G}_r^s(Y_1, p(\cdot)) \\ &= \rho \left[ |P_0^{-1} A_1' A_1 + I_k|^{\frac{s-1}{2}} - |P_0^{-1} A' A + I_k|^{\frac{s-1}{2}} \right] \\ &= \rho \left[ |I_{n-t} + A_1 P_0^{-1} A_1'|^{\frac{s-1}{2}} - |I_n + A P_0^{-1} A'|^{\frac{s-1}{2}} \right], \end{aligned}$$

where

$$\rho = \frac{|P_0|^{\frac{s-1}{2}}}{r^{\frac{k(s-1)}{2(r-1)}} (2^{1-s} - 1) (2\pi)^{\frac{(s-1)k}{2}} \sigma^{s-1}}.$$

Now in accordance with the partition of  $A$ , we partition  $I_n + A P_0^{-1} A'$  as follows:

$$\begin{aligned} |I_n + A P_0^{-1} A'| &= |V_1| |I_t + A_2 P_0^{-1} A_2' - A_2 P_0^{-1} A_1' V_1^{-1} A_1 P_0^{-1} A_2'| \\ &= |V_1| |I_t + A_2 S A_2'|, \end{aligned}$$

where

$$S = P_0^{-1} - P_0^{-1} A_1' V_1^{-1} A_1 P_0^{-1}.$$

Therefore for  $r \neq 1$  and  $s \neq 1$ , we get

$$\mathcal{L}_r^s(A_2; A_1) = \rho |V_1|^{\frac{s-1}{2}} \left[ 1 - |I_t + A_2 S A_2'|^{\frac{s-1}{2}} \right].$$

Now the results follows by continuity of  $\mathcal{L}_r^s$  with respect to  $r$  and  $s$ .

**Remark 2.** To compare the influence of two sets of  $t$  observations Ghosh and Namini (1989) proposed several measures. One of their measures is

$$G(A_2; A_1) = |I_t + A_2 (A_1' A_1)^{-1} A_2'|$$

If we consider  $P_0 \simeq 0$ ; i.e., if we suppose a low level of prior knowledge and independence between the  $\beta_i$ 's, we obtain

$$\mathcal{L}_r^s(A_2; A_1) = aG(A_2; A_1) + b$$

For a set of unavailable observations with minimum  $\mathcal{L}_r^s(A_2, A_1)$  the resulting design  $A_1^*$  will be the best among all possible designs  $A_1$ . The observations,  $Y_1$ ,

corresponding to the best design  $A_1^*$  will be called most informative. We consider the case  $t = 1$  and we write  $A_2(1 \times k) = \alpha'$ . Here  $\alpha (k \times 1)$  is an arbitrary non-null vector with fixed euclidean norm  $a = \|\alpha\|$ . Therefore

$$\mathcal{L}_r^s(A_2, A_1) = K \frac{1 - (1 + \alpha' S \alpha)^{(s-1)/2}}{2^{1-s} - 1}, \quad (1)$$

where

$$K = \frac{|V_1|^{\frac{s-1}{2}} |P_0|^{\frac{s-1}{2}}}{r^{\frac{k(s-1)}{2(r-1)}} (2\pi)^{\frac{(s-1)k}{2}} \sigma^{s-1}}$$

is a positive number for all  $r \in (0, \infty)$  and  $s \in (-\infty, +\infty)$ .

**Theorem 3.** *Among all observations vectors  $\alpha (k \times 1)$  with fixed euclidean norm  $a$ , the loss of information (1) will be minimum if the observation is taken in the direction of the eigenvector associated to the smallest eigenvalue of  $S$ . Furthermore the unavailability of an observation in the direction of the eigenvector associated to the largest eigenvalue of  $S$  maximizes the loss of information (1).*

**Proof.** For all  $r \in (0, \infty)$  and any  $s \in (-\infty, +\infty)$ ,  $\mathcal{L}_s^r(A_2, A_1)$  is an increasing function of  $\alpha' S \alpha$ , so the problem

$$\begin{aligned} & \text{minimize } \mathcal{L}_s^r(A_2, A_1) \\ & \alpha / \|\alpha\| = a \end{aligned}$$

reduces to

$$\begin{aligned} & \text{minimize } \alpha' S \alpha. \\ & \alpha / \|\alpha\| = a \end{aligned}$$

As  $S$  is a symmetric and positive definite matrix, we can write  $S = CDC'$ , where  $D (k \times k)$  is a diagonal matrix whose elements are the eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$  of  $S$  and  $C (n \times k)$  is a matrix whose columns are the corresponding orthonormal eigenvectors. Accordingly, it follows that

$$\alpha' S \alpha = \alpha' CDC' \alpha = \sum_{i=1}^k \lambda_i (\alpha' T_i)^2 = a^2 \sum_{i=1}^k \lambda_i \cos^2 \theta_i.$$



As  $(T_1, T_2, \dots, T_k)$  is an orthonormal basis of  $\mathbb{R}^k$ ,  $1 = \sum_{i=1}^k \cos^2 \theta_i$ , so the problem reduces to

$$\text{minimize}_{\pi_1, \dots, \pi_k} \sum_{i=1}^k \lambda_i \pi_i.$$

under the constrain  $1 = \pi_1 + \pi_2 + \dots + \pi_k$ , where  $\pi_i = \cos^2 \theta_i$ . Clearly, a minimum is attained when  $\pi_k = 1$ , i.e., when  $\alpha$  is proportional to  $T_k$ , and a maximum is attained when  $\pi_1 = 1$ , i.e., when  $\alpha$  is proportional to  $T_1$ .

**Remark 3.** The result of theorem 3 should be extended to the care of unavailable  $t$  observations; however no clear statement can be given. To show the complexity of the problem, we analyze the case  $t = 2$ . Let us write  $A_2(2 \times k) = [\alpha', \beta']'$ , where  $\alpha(k \times 1)$  and  $\beta(k \times 1)$  are arbitrary non-null vectors with fixed euclidean norms  $a$  and  $b$  respectively. Therefore, the problem to be solved is:

$$\begin{aligned} \text{minimize } \mathcal{L}_s^T(A_2, A_1) &= \text{minimize } |I_t + A_2 S A_2'| \\ \alpha / \|\alpha\| &= a & \alpha / \|\alpha\| &= a \\ \beta / \|\beta\| &= b & \beta / \|\beta\| &= b \end{aligned}$$

$$\begin{aligned} &= \text{minimize } \{(\alpha' S \alpha + 1)(\beta' S \beta + 1) - (\alpha' S \beta)^2\} \\ \alpha / \|\alpha\| &= a \\ \beta / \|\beta\| &= b \end{aligned}$$

$$\begin{aligned} &= \text{minimize}_{p_1, \dots, p_k, q_1, \dots, q_k} \left\{ \left(1 + a^2 \sum_{i=1}^k \lambda_i p_i\right) \left(1 + b^2 \sum_{i=1}^k \lambda_i q_i\right) - \left(ab \sum_{i=1}^k \lambda_i (p_i q_i)^{1/2}\right) \right\}, \end{aligned}$$

under the constrains  $p_1 + p_2 + \dots + p_k = 1$ ,  $q_1 + q_2 + \dots + q_k = 1$ ,  $p_i \geq 0$ ,  $q_i \geq 0$ ,  $i = 1, \dots, k$ , where  $p_i = \cos^2 \theta_i$ ,  $q_i = \cos^2 \nu_i$ ,  $\theta_i = \text{Angle}(\alpha, T_i)$ ,  $\nu_i = \text{Angle}(\beta, T_i)$  and  $\lambda_i$  and  $T_i$  are the  $i$ th eigenvalue and the  $i$ th normalized eigenvector of matrix  $S$  respectively.

### References

[1] M. A. Ali, "A Bayesian Approach to detect Informative Observations in an Experiment", *Commun. Statist. Theory and Meth.*, 19 (7) (1990), 2567-2575.

- [2] S. Arimoto, "Information-Theoretic Consideration on Estimation Problems", *Information and Control*, 19 (1971), 181-194.
- [3] S. Chatterjee and A. S. Hadi, "Influential Observations, High Leverage Points, and Outliers in Linear Regression", *Statistical Science*, 1 (1986), 379-416.
- [4] R. D. Cook and S. Weisberg, "Characterization of an empirical Influence Function for Detecting Influential Cases in Regression", *Technometrics*, 22 (1980), 495-508.
- [5] R. D. Cook and S. Weisberg, "Residuals and Influence in Regression", *Biometrika*, 70 (1982), 1-10.
- [6] R. D. Cook, "Assessment of Local Influence", *Journal of the Royal Statistical Society B*, 48 (1986), 133-169.
- [7] M. H. DeGroot, "Optimal Statistical Decisions", *McGraw-Hill*, 1970.
- [8] S. Ghosh, "Information in an observation in robust Designs", *Comm. Statist. Theor. Meth.*, 11 (1982), 1173-1184.
- [9] S. Ghosh, "Influential Observations in view of Design and Inference", *Comm. Statist. Theor. Meth.*, 12 (1983), 1675-1683.
- [10] S. Gosh and H. Namini, "Influential Observations under robust designs", *Proceeding of Design and Coding Theory Conference, Institute of Mathematics and its Applications, University of Minnesota, Minneapolis, U. S. A.* (1989).
- [11] I. Havrda and F. Charvat, "Quantification method of classification processes: concept of structural  $\alpha$ -entropy", *Kybernetika*, 3 (1967), 30-35.
- [12] A. J. Lawrence, "Regression Transformation Diagnostics Using Local Influence", *Journal of the American Statistical Association*, 83 (1988), 1067-1072.
- [13] A. Renyi, "On Measures of Entropy and Information", *Proced. 4th Berkeley Symp. Math. Statist. and Prob.*, 1 (1961), 547-561.
- [14] C. Shannon, "A Mathematical Theory of Communications", *Bell. System Tech. J.*, 27 (1948), 379-423.
- [15] B. D. Sharma and D. P. Mittal, "New Nonadditive measures of Entropy for Discrete Probability Distributions", *J. Math. Sci.*, 10 (1975), 28-40.
- [16] I. J. Taneja, L. Pardo, D. Morales and M. L. Menéndez, "On Generalized Information and Divergence Measures and their Applications: A Brief Review", *Qüestiió*, Vol. 13, N. 1,2,3, (1989), 47-73.

Departamento de Estadística e I. O., Facultad de Matemáticas, Universidad Complutense de Madrid 28040-Madrid, Spain.