

# **Qüestió**

**Quaderns d'Estadística  
i Investigació Operativa**

Any 2000, volum 24, núm. 3  
Segona època

Entitats patrocinadores:

Universitat Politècnica de Catalunya  
Universitat de Barcelona  
Universitat de Girona  
Universitat Autònoma de Barcelona  
Institut d'Estadística de Catalunya

Entitat col·laboradora:  
International Biometric Society



Generalitat de Catalunya  
**Institut d'Estadística  
de Catalunya**

Any 2000, volum 24, núm. 3

## SUMARI

### *Editorial*

### *Estadística*

A note on the matrix Haffian .....	419
<b>H. Neudecker</b>	
Variables finitas condicionalmente especificadas .....	425
<b>R. Pérez-Villalta</b>	
Análisis de detección de raíces unitarias en series de tiempo. Un enfoque metodológico con tests no similares .....	449
<b>J. Á. Roldán Casas y R. Dios Palomares</b>	

### *Investigació Operativa*

Del poliedro del agente viajero gráfico al de rutas de vehículos con demanda compartida	495
<b>C. Martínez y E. Mota</b>	

### *Estadística Oficial*

Avaluació dels mercats dels serveis d'informació electrònica .....	531
<b>T. Baiget</b>	

### *Biometria*

Cálculo de probabilidades en el análisis de perfiles genéticos compatibles .....	555
<b>M. Sánchez García y P. Cuesta Alvaro</b>	

<i>Secció docent i problemes</i> .....	577
--	-----

### *Comentaris de llibres*

### *Ressenyes d'activitats institucionals*

### *Informació per als autors i lectors*



## EDITORIAL

El darrer número del volum 24, corresponent a l'any 2000, recull un total de sis articles repartits entre les quatre seccions temàtiques, així com dues recensions a l'apartat dedicat a comentaris de llibres. Amb la publicació del tercer número, la producció editorial d'enguany abasta 21 articles que, amb la informació impresa d'altres apartats de la revista, ha representat un total de 620 pàgines. Les xifres de l'exercici confirmen un augment dels continguts impresos respecte de l'any 1999 i una consolidació del nombre d'articles editats que, per tant, permeten millorar la mitjana de la 2a època de la revista, que ja se situa en 19.8 articles i 499 pàgines impreses per volum o any.

Aquesta retrospectiva també permet valorar algunes de les millores que s'han produït en el decurs de l'any 2000: la incorporació de la Universitat Autònoma de Barcelona com a nou patrocinador –amb l'aportació material i científica que això suposa per a la revista–, l'adopció de la nova classificació («2000 Mathematical Subject Classification») dels articles publicats des del número 2 del present volum 24 a partir de l'actualització de la MSC 1991, o bé la ressenya de novetats editorials en matèria estadística publicades per l'Idescat i la resta de la Generalitat de Catalunya al llarg d'aquest any.

Al mateix temps, la difusió electrònica de *Qüestió* mitjançant Internet també ha experimentat ampliacions importants i una projecció creixent: d'una banda, s'ha incorporat la consulta interactiva dels enunciats de problemes proposats i les seves solucions, de moment amb els publicats en els darrers dos volums, així com l'enllaç per a l'adscripció dels descriptors classificadors mitjançant la cerca de les paraules clau de l'article i la consulta de les correspondències amb la MSC 1991 a través del web de l'Idescat. D'altra banda, les 55.800 consultes al web en el decurs de l'any 2000, que han evolucionat des de les 4.500 fins a les 8.500 peticions d'accisos, evidencien un augment del 200% respecte de les 18.300 consultes del 1999.

Per últim, com ja és costum, en el darrer número de *Qüestió* es fa balanç de la presentació i avaluació d'originals que han enregistrat les diferents seccions. D'acord amb la informació divulgada fins el número 3 del volum anterior, les dades següents es refereixen al període gener-desembre de l'any 2000.

- articles sotmesos: 29
- articles en procés d'avaluació: 29
- articles acceptats: 19 (18 publicats i 1 en espera de publicació)
- articles rebutjats: 10

És interessant constatar que el volum d'articles sotmesos a la revista es manté entre les magnituds més elevades de la 2a època de *Qüestió*. D'altra banda, les xifres anteriors es complementen amb un lleuger augment del percentatge de rebuig d'originals (proper al 40% dels articles sotmesos).

C.M. Cuadras i E. Ripoll, editors executius

**Comentari de les seccions  
«Estadística», «Investigació Operativa» i «Biometria»**

En aquestes seccions s'hi publiquen cinc articles: tres d'Estadística, un d'Investigació Operativa i un de Biometria. El primer article, *A note on the matrix Haffian*, de H. Neudecker, és un estudi d'una funció matricial basada en les derivades d'una funció d'una matriu simètrica, obtenint-se una versió en la línia de Magnus i Neudecker. Pel que fa al segon article, *Variables finitas condicionalmente especificadas*, de R. Pérez-Villalta, és una contribució a la teoria de les distribucions bivariants discretes amb marginals donades, emprant el càlcul matricial i la descomposició singular d'una matriu, la qual cosa permet estudiar condicions d'unicitat i existència de la distribució. El darrer article de la secció «Estadística», titulat *Análisis de la detección de raíces unitarias en series de tiempo. Un enfoque metodológico con tests no similares*, de J.A. Roldán i R. Dios, presenta els contrastos d'arrel unitària en una sèrie temporal, estudiant els diferents contrastos per detectar la presència d'arrels unitàries així com les situacions que hom pot trobar a fi de minimitzar l'error, estudi que els autors completen amb una simulació.

A la secció «Investigació Operativa» es publica *Del poliedro del agente viajero gráfico al de rutas de vehículos con demanda compartida*, de M.C. Martínez i E. Mota, els quals estudien el problema clàssic del repartiment de mercaderies als clients quan hi ha una flota fixa disponible, minimitzant la distància del recorregut, però considerant que es pot aprovisionar a cada client amb més d'un vehicle; els autors aporten uns resultats teòrics que amplien resultats ja coneguts i destaquen el seu interès pràctic.

Per últim, l'article de M. Sánchez i P. Cuesta que acull la secció «Biometria», *Cálculo de probabilidades en el análisis de perfiles genéticos compatibles*, planteja els diversos càlculs probabilístics relatius als perfils genètics dels possibles autors d'un delicte determinat en relació a un col·lectiu, n'estudia les condicions de compatibilitat i presenta diverses situacions pràctiques d'interès.

Carles M. Cuadras, editor executiu

### **Comentari de la secció «Estadística Oficial» i altres apartats**

La secció «Estadística Oficial» presenta un únic article que prové de l’Idescat, *Avaluació dels mercats dels serveis d’informació electrònica*, de T. Baiget, en el qual es caracteritzen els reptes metodològics i operatius que ha suposat l’estadística oficial (encara experimental) sobre el sector de la informació electrònica, en el marc del projecte comunitari MSStudyII. El tractament de l’oferta, la demanda i l’estudi de casos i tendències per a cada país/estat participant ha representat, sobretot, innovacions rellevants, tant pel que fa a la delimitació del sector com al tractament de les dades estadístiques de fenòmens emergents.

A continuació, la «Secció docent i problemes» inclou la presentació successiva d’enunciats de problemes i la resolució dels publicats en el número anterior, tots els quals ja es poden consultar també a la pàgina dedicada a la revista en el web de l’Idescat. Seguidament, la secció «Comentaris de llibres» acull dues recensions, la primera de J.M. Oller i la segona de C.M. Cuadras. En el primer cas, la ressenya dels dos volums del llibre *Problemas de probabilidad y Estadística*, de C.M. Cuadras, destaca l’esforç d’actualització i ampliació de l’obra editada el 1971 i que ha esdevingut una referència en l’ensenyament de l’estadística al nostre entorn. En el cas de *Las Matemáticas del siglo XX. Una mirada en 101 artículos*, edició coordinada per A. Martinón, destaca la visió exhaustiva de la progressió de l’instrumental matemàtic al servei de múltiples disciplines, amb una detallada referència als aspectes teòrics, docents, aplicats i, inclús, biogràfics dels principals científics que ho han procurat.

El darrer apartat, dedicat a «Ressenyes d’activitats institucionals», inclou, com ja és habitual, una revisió actualitzada d’activitats de la Sociedad Española de Biometría, amb l’anunci de la VIII Conferència Espanyola de Biometria que tindrà lloc del 28 al 30 de març de 2001 a Pamplona. En segon lloc, es reprenen les recensions d’activitats del «Training for European Statisticians Institute», amb la relació dels 26 cursos corresponents al *Core Programme 2001* que s’impartiran fins el mes de novembre, adreçats, principalment, als membres d’organismes europeus d’estadística oficial. Més endavant, s’ofereix el darrer anunci de les 2es jornades internacionals sobre *Data Mining* organitzades per la Xarxa Temàtica «Enquestes i qualitat de la informació estadística» (Barcelona, 14-15 desembre 2000), amb el suport de l’Idescat i l’ambaixada francesa. Les darreres pàgines es dediquen a ressenyar totes les novetats editorials de l’Idescat i de la resta de la Generalitat de Catalunya que han estat publicades en el decurs de l’any 2000 en l’àmbit de l’estadística.

Enric Ripoll, editor executiu



# **Estadística**



QÜESTIIÓ, vol. 24, 3, p. 419-424, 2000

## A NOTE ON THE MATRIX HAFFIAN

HEINZ NEUDECKER

Cesaro\*

*This note contains a transparent presentation of the matrix Haffian. A basic theorem links this matrix and the differential of the matrix function under investigation, viz  $\nabla F(X)$  and  $dF(X)$ .*

*Frequent use is being made of matrix derivatives as developed by Magnus and Neudecker.*

**Keywords:** Matrix differentials, matrix functions

**AMS Classification (MSC 2000):** primary 62F10, secondary 62C99

---

\*Oosterstraat, 13. 1741 GH Schagen. The Netherlands.  
E-mail: heinz@fee.uva.nl

-Received February 2000.  
-Accepted September 2000.

## 1. INTRODUCTION

Haff (1981, 1982) introduced a matrix function based on the derivatives of the elements of a **square** matrix function  $F(X)$  of a **symmetric** matrix argument  $X$ .

We shall name it «the matrix Haffian». It was used by Haff in various multivariate statistical problems.

In this note we shall attempt to give a transparent presentation of the matrix Haffian, and give some applications.

Basic is a differentiable square matrix function  $F(X)$ , shortly  $F$ , which depends on a **symmetric** matrix  $X$ . Both matrices have the same dimension. A strategic rôle is being played by a square matrix  $\nabla = (d_{ij})$  of differential operators  $d_{ij} := 1/2(1 + \delta_{ij}) \frac{\partial}{\partial x_{ij}}$ , where  $\delta_{ij}$  is the Kronecker delta ( $\delta_{ii} = 1$ ,  $\delta_{ij} = 0$  for  $i \neq j$ ). Haff used the symbol  $D$ , and not  $\nabla$ . In earlier work on the scalar Haffian (Neudecker 2000) the symbol  $\nabla$  was introduced. This was done to avoid confusion with the duplication matrix which was extensively used then. The matrix  $\nabla$  will be applied to  $F$  and produce  $\nabla F$ , the matrix Haffian.

In the article frequent use will be made of matrix differentials and derivatives as proposed by Magnus and Neudecker (1999).

## 2. THE MATRIX HAFFIAN

Consider a differentiable square matrix function  $F(X)$  with **symmetric** matrix argument  $X$ , both of dimension  $m$ . The application of  $\nabla = (d_{ij})$ , a (square) matrix of differential operators  $d_{ij} := 1/2(1 + \delta_{ij}) \frac{\partial}{\partial x_{ij}}$  to  $F$  yields  $\nabla F$ . Its  $ik^h$  typical element is

$$\sum_j d_{ij} f_{jk} = 1/2 \sum_j (1 + \delta_{ij}) \frac{\partial f_{jk}}{\partial x_{ij}} = \frac{\partial f_{ik}}{\partial x_{ii}} + 1/2 \sum_{i \neq j} \frac{\partial f_{jk}}{\partial x_{ij}}.$$

We shall prove a basic result.

### *Theorem*

When  $dF = P(dX)Q'$ , then  $\nabla F = 1/2P'Q' + 1/2(\text{tr}P)Q'$ .

*Proof*

Using

$$dX = \Sigma_{ij}(dx_{ij})E_{ij},$$

where

$$X = \Sigma_{ij}x_{ij}E_{ij}$$

and  $E_{ij}$  is the  $ij^{th}$  basis matrix, we get

$$\begin{aligned} dF &= P(dX)Q' = \Sigma_{ij}(dx_{ij})PE_{ij}Q' = \\ &= \Sigma_i(dx_{ii})PE_{ii}Q' + \sum_{i \neq j}(dx_{ij})PE_{ij}Q'. \quad (i, j = 1 \dots m) \end{aligned}$$

Hence

$$(1) \quad \frac{\partial F}{\partial x_{ii}} = PE_{ii}Q'$$

$$(2) \quad \frac{\partial F}{\partial x_{ij}} = P(E_{ij} + E_{ji})Q'. \quad (i \neq j)$$

The second expression follows from the symmetry of  $X$ , where  $x_{ij} = x_{ji}$  ( $i \neq j$ ).

Consider the  $ik^{th}$  typical element of  $\nabla F$ , viz

$$\begin{aligned} \Sigma_j d_{ij} f_{jk} &= \frac{\partial f_{ik}}{\partial x_{ii}} + 1/2 \sum_{j \neq i} \frac{\partial f_{jk}}{\partial x_{ij}} \\ &= e'_i \frac{\partial F}{\partial x_{ii}} e_k + 1/2 \sum_{j \neq i} e'_j \frac{\partial F}{\partial x_{ij}} e_k \\ &= e'_i P E_{ii} Q' e_k + 1/2 \sum_{j \neq i} e'_j P (E_{ij} + E_{ji}) Q' e_k \\ &= e'_i P e_i e'_i Q' e_k + 1/2 \sum_{j \neq i} e'_j P (e_i e'_j + e_j e'_i) Q' e_k \\ &= p_{ii} q_{ki} + 1/2 \sum_{j \neq i} (p_{ji} q_{kj} + p_{jj} q_{ki}) \\ &= 1/2 \sum_j p_{ji} q_{kj} + 1/2 \sum_j p_{jj} q_{ki} \\ &= 1/2 (QP)_{ki} + 1/2 (\text{tr}P) q_{ki} \\ &= 1/2 (P' Q')_{ik} + 1/2 (\text{tr}P) (Q')_{ik}. \end{aligned}$$

Hence  $(\nabla F)_{ik} = 1/2[P'Q' + (\text{tr}P)Q']_{ik}$ ,

$$\nabla F = 1/2P'Q' + 1/2(\text{tr}P)Q'.$$

We used  $E_{ij} = e_i e_j'$ , where  $e_i$  and  $e_j$  are basis vectors. Further  $(A)_{ik} = a_{ik} = e_i' A e_k$  was applied.

□

### 3. SOME MATRIX HAFFIANS

We consider four matrix Haffians. They are of a simple nature. For one we quote the literature.

$$(1) \quad \nabla P X^{-1} Q' = -1/2 X^{-1} P' X^{-1} Q' - 1/2(\text{tr} P X^{-1}) X^{-1} Q'.$$

*Proof*

Now

$$dF = dP X^{-1} Q' = P(dX^{-1})Q' = -P X^{-1}(dX)X^{-1}Q',$$

by virtue of  $dX^{-1} = -X^{-1}(dX)X^{-1}$ .

Using the Theorem with the substitutions  $P \rightarrow -P X^{-1}$  and  $Q' \rightarrow X^{-1} Q'$  we get the result immediately.

□

**Note.** Haff (1982, Lemma 6i) gave  $\nabla Q X^{-1}$  by using a long-winded procedure. We have  $\nabla Q X^{-1} = -1/2 X^{-1} Q' X^{-1} - 1/2(\text{tr} Q X^{-1}) X^{-1}$ .

$$(2) \quad \nabla P X Q X R' = 1/2 P' Q X R' + 1/2(\text{tr} P) Q X R' + 1/2 Q' X P' R' + 1/2(\text{tr} P X Q) R'.$$

*Proof*

$$dP X Q X R' = P(dX) Q X R' + P X Q(dX) R',$$

which gives the substitutions

$$P \rightarrow P, \quad Q' \rightarrow Q X R' \quad \text{and} \quad P \rightarrow P X Q, \quad Q' \rightarrow R'.$$

The Theorem completes the story.

□

$$(3) \quad \nabla P X^{-2} Q' = -1/2 X^{-1} P' X^{-2} Q' - 1/2(\text{tr} P X^{-1}) X^{-2} Q' - 1/2 X^{-2} P' X^{-1} Q' - 1/2(\text{tr} P X^{-2}) X^{-1} Q'.$$

*Proof*

We have

$$\begin{aligned} dPX^{-2}Q' &= P(dX^{-2})Q' = P(dX^{-1})X^{-1}Q' + PX^{-1}(dX^{-1})Q' \\ &= -PX^{-1}(dX)X^{-2}Q' - PX^{-2}(dX)X^{-2}Q'. \end{aligned}$$

Application of the Theorem with the substitutions  $P \rightarrow -PX^{-1}$ ,  $Q' \rightarrow X^{-2}Q'$  and  $P \rightarrow -PX^{-2}$ ,  $Q' \rightarrow X^{-1}Q'$  finishes the derivation.  $\square$

$$(4) \quad \nabla PX^3Q' = 1/2P'X^2Q' + 1/2(\text{tr}P)X^2Q' + 1/2XP'XQ' + \\ + 1/2(\text{tr}PX)XQ' + 1/2X^2P'Q' + 1/2(\text{tr}PX^2)Q'.$$

*Proof*

$$dPX^3Q' = P(dX^3)Q = P(dX)X^2Q' + PX(dX)XQ' + PX^2(dX)Q'.$$

There are three sets of substitutions, viz.

$$P \rightarrow P, Q' \rightarrow X^2Q'; P \rightarrow PX, Q' \rightarrow XQ' \text{ and } P \rightarrow PX, Q' \rightarrow Q'.$$

This yields the result.  $\square$

#### 4. SOME REMARKS ON SCALAR HAFFIANS

The scalar Haffian is defined as  $\text{tr } \nabla F$ .

See Neudecker (2000) for a thorough discussion and some applications. In that article the identity

$$\text{tr } \nabla F = \text{tr} \frac{\partial g}{\partial x'}$$

was established, where  $x := v(X)$ ,  $g := v(G)$ ,  $G := 1/2(F + F')$  and  $\frac{\partial g}{\partial x'}$  is the Magnus-Neudecker derivative matrix for the vector function  $g(x)$ .

The scalar Haffian can, of course, immediately be derived from the corresponding matrix Haffian as established in the present article.

## REFERENCES

- Haff, L.R. (1981). «Further identities for the Wishart distribution with applications in regression». *Canad. J. Statist.*, 9, 215-224.
- Haff, L.R. (1982). «Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination». *Sankhyā B*, 44, 245-258.
- Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edition. John Wiley, Chichester.
- Neudecker, H. (2000). *A note on the scalar Haffian*. *Quæstio*, 24 (2), 243-249.

## VARIABLES FINITAS CONDICIONALMENTE ESPECIFICADAS

ROMÁN PÉREZ-VILLALTA

Universidad de Sevilla\*

*En este trabajo se estudia la existencia y unicidad de vectores bidimensionales de variables discreta con recorrido finito, cuando se fijan sus distribuciones condicionadas. Para ello, tras repasar la literatura existente sobre el tema, proporcionamos diversos resultados que relacionan diversos temas de álgebra matricial, especialmente la descomposición singular, con el problema que nos ocupa.*

**Finite conditionally specified variables**

**Palabras clave:** Distribuciones condicionadas, especificación condicional, rango de una matriz, descomposición singular, cadenas de Markov

**Clasificación AMS (MSC 2000):** 60E05, 62E10

---

\*Facultad de Económicas y Empresariales. Dpto. de Economía Aplicada I. Avda. Ramón y Cajal, 1.  
41018 Sevilla. España. villalta@cica.es

-Recibido en diciembre de 1999.  
-Aceptado en septiembre de 2000.

## 1. INTRODUCCIÓN

Dado un vector aleatorio bidimensional  $(X, Y)$  podemos especificar su función de cuantía o densidad de diversas formas: dar directamente esa función, dar una marginal y una condicionada (en la misma o distinta componente), dar dos marginales... Dependiendo del caso la densidad o cuantía del vector existirá o no y, además, será única o no.

El caso en el que se dan dos condicionadas, objeto de este trabajo, ha sido estudiado extensamente por Arnold y Press (1989) y Arnold, Castillo y Sarabia (1992) aunque se pueden nombrar trabajos pioneros como Gourieroux y Monfort (1979) y Abrahams y Thomas (1984). Señalemos antes de continuar que la lista anterior no cubre toda la literatura al respecto.

Respecto a las aplicaciones, cabe destacar la construcción de modelos bivariantes (p.e. Arnold, Castillo y Sarabia, 1992), estudios relativos a la convergencia del algoritmo del muestreo de Gibbs (Casella y George, 1992 y Casella, 1996), en Estadística Bayesiana (Arnold, Castillo y Sarabia, 1996 y Basulto, 1995) entre otras.

Este trabajo repasa los resultados existentes para el caso en que las variables aleatorias son finitas y aporta algún resultado adicional utilizando elementos de cálculo matricial y ejemplos que clarifican los distintos métodos expuestos

## 2. PLANTEAMIENTO DEL PROBLEMA Y NOTACIÓN

Sean dos matrices  $\mathbf{A} = (a_{i,j})$  y  $\mathbf{B} = (b_{i,j})$  de orden  $L \times M$  que verifican  $a_{i,j} \geq 0, b_{i,j} \geq 0$   $\forall i = 1, \dots, L; \forall j = 1, \dots, M$  y además

$$\sum_{i=1}^L a_{i,j} = 1, \sum_{j=1}^M b_{i,j} = 1.$$

Esto es,  $\mathbf{A}$  y  $\mathbf{B}'$  son matrices estocásticas por columnas. La cuestión a dilucidar es, bajo qué condiciones dos matrices  $\mathbf{A}$  y  $\mathbf{B}$ , de dimensiones  $L \times M$ , en las condiciones anteriores, pueden ser consideradas distribuciones condicionadas de cierto vector aleatorio  $(X, Y)$  de variables discretas con recorrido finito. Esto es, si  $x_1, x_2, \dots, x_L$  son los posibles valores de  $X$  e  $y_1, y_2, \dots, y_M$  los de  $Y$ , cuando

$$a_{i,j} = P[X = x_i | Y = y_j] \quad i = 1 \dots L$$

$$b_{i,j} = P[Y = y_j | X = x_i] \quad j = 1 \dots M$$

Equivalentemente, bajo qué condiciones existen dos vectores  $\tau$  y  $\eta$ , de dimensiones  $L$  y  $M$  respectivamente, no negativos y de forma que la versión normalizada (la suma de coordenadas unidad) de  $\tau$  y  $\eta$  sean las distribuciones marginales correspondientes.

Nótese que podemos suponer  $\tau_i b_{i,j} = \eta_j a_{i,j}$   $i = 1, \dots, L$   $j = 1, \dots, M$ , ya que si un  $\tau_i$  o un  $\eta_j$  es nulo, el valor  $x_i$  o  $y_j$  tiene todas sus probabilidades nulas y puede ser eliminado del problema.

Esta primera cuestión recibe el nombre de problema de compatibilidad y, en ese contexto, las matrices  $\mathbf{A}$  y  $\mathbf{B}$  se denominan candidatas a distribuciones condicionadas, y además, si la respuesta al problema es afirmativa se dice que  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles.

Un segundo problema a tratar es, supuesto que existan, la unicidad de estas variables  $X$  e  $Y$  o vectores  $\tau$  y  $\eta$ .

Para las matrices  $\mathbf{A}$  y  $\mathbf{B}$  definimos los conjuntos

$$N_A = \{(i, j) : a_{i,j} > 0\}, N_B = \{(i, j) : b_{i,j} > 0\}$$

Cuando se tenga  $N_A = N_B$  los denotaremos, indistintamente, por  $N$ .

### 3. COMPATIBILIDAD

Un primer resultado sobre la compatibilidad de distribuciones puede encontrarse en Arnold y Press (1989) cuya traducción a matrices es el siguiente.

**Teorema 1.** (Arnold y Press, 1989) *Dos matrices  $\mathbf{A}$  y  $\mathbf{B}$  en las condiciones de la sección anterior son compatibles si y sólo si  $N_A = N_B = N$  y además existen dos vectores,  $\mathbf{u} = (u_1, u_2, \dots, u_L)$  y  $\mathbf{v} = (v_1, v_2, \dots, v_M)$  tales que*

$$c_{i,j} = \frac{a_{i,j}}{b_{i,j}} = u_i v_j; \quad \forall (i, j) \in N$$

La primera condición refleja el hecho de que las cuantías condicionadas se anulan en los mismos puntos que coinciden con los puntos en los que se anula la función de cuantía conjunta. Por su parte, la segunda condición hace referencia a la relación entre marginales y condicionadas  $\tau_i b_{i,j} = \eta_j a_{i,j}$   $i = 1, \dots, L$ ;  $j = 1, \dots, M$  de esta forma  $u_i = \tau_i$  y  $v_j = 1/\eta_j$ .

En el caso que nos ocupa, variables finitas, es posible utilizar resultados de álgebra matricial para analizar la compatibilidad de las matrices  $\mathbf{A}$  y  $\mathbf{B}$ . En efecto, supongamos que  $N = \{1, \dots, L\} \times \{1, \dots, M\}$  y denotemos por  $\mathbf{C}$  la matriz de elementos  $c_{i,j} = \frac{a_{i,j}}{b_{i,j}}$ , en esas condiciones se tiene el siguiente

**Teorema 2.** *En las condiciones anteriores  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles si y sólo si el rango de  $\mathbf{C}$  es uno.*

### Demostración

Si rango de  $\mathbf{C}$  es uno, existen dos matrices ortogonales  $\mathbf{U}$  y  $\mathbf{V}$  de órdenes  $L$  y  $M$  respectivamente de forma que  $\mathbf{C} = \mathbf{U}\Sigma\mathbf{V}^t$  donde  $\Sigma$  es una matriz  $L \times M$  cuyo elemento  $(1,1)$  es no nulo y el resto nulos. Obsérvese que esto no es más que la descomposición en valores singulares de la matriz  $\mathbf{C}$  y el elemento no nulo de  $\Sigma$  el único valor singular de  $\mathbf{C}$  (ver p.e. Schott, 1997). Dada la forma de la matriz  $\Sigma$  podemos escribir  $\mathbf{C} = \mathbf{u}\sigma\mathbf{v}^t$  donde  $\mathbf{u}$  y  $\mathbf{v}$  son las primeras columnas de  $\mathbf{U}$  y  $\mathbf{V}$  respectivamente. Así  $c_{i,j} = u_i\sigma v_j, \forall (i,j) \in N$ , y por tanto, incluyendo el valor singular en uno de los vectores  $\mathbf{u}$  o  $\mathbf{v}$ , se tiene la factorización  $c_{i,j} = u_i v_j, \forall (i,j) \in N$ .

Por otra parte, puesto que  $u$  es autovector de  $\mathbf{CC}^t$ , que es no negativa, y está asociado al autovalor  $\sigma > 0$ , puede ser elegido con coordenadas no negativas y puesto que  $\mathbf{v} \propto \mathbf{C}^t \mathbf{u}$  también es no negativo, las matrices  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles y, en este caso, una adecuada elección de las distribuciones marginales es

$$P[X = x_i] = \tau_i = \frac{u_i}{\sum_{i=1}^L u_i} \quad i = 1, \dots, L$$

$$P[Y = y_j] = \eta_j = \frac{v_j^{-1}}{\sum_{j=1}^M v_j^{-1}} \quad j = 1, \dots, M$$

Recíprocamente, si las matrices  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles, el rango de  $\mathbf{C}$  es la unidad pues, si tomamos un menor de dimensión dos cualquiera,

$$\begin{vmatrix} c_{i,j} & c_{i,k} \\ c_{l,j} & c_{l,k} \end{vmatrix} = \begin{vmatrix} \frac{a_{i,j}}{b_{i,j}} & \frac{a_{i,k}}{b_{i,k}} \\ \frac{a_{l,j}}{b_{l,j}} & \frac{a_{l,k}}{b_{l,k}} \end{vmatrix} = \begin{vmatrix} u_i v_j & u_i v_k \\ u_l v_j & u_l v_k \end{vmatrix} = u_i v_j u_l v_k - u_i v_k u_l v_j = 0$$

$\forall i, l = 1, \dots, L$  y  $\forall j, k = 1, \dots, M$ .

□

El resultado anterior puede encontrarse, bajo otra perspectiva en Arnold y Gokhale (1994) y Arnold, Castillo y Sarabia (1999).

Observemos que, en este caso, la condición de compatibilidad vía descomposición singular se simplifica notablemente pues, como observan Arnold y Press (1989) y Arnold, Castillo y Sarabia (1992), se verifica que:

$$c_{i,j} c_{..} = c_i c_{..j} \quad \forall (i,j) \in N$$

donde

$$c_{..} = \sum_{i,j} c_{i,j}, \quad c_{,j} = \sum_i c_{i,j} \quad \text{y} \quad c_{i.} = \sum_j c_{i,j}$$

y en consecuencia:  $u_i \propto c_{i.}$  y  $v_j \propto c_{,j}, \quad \forall (i, j) \in N$

**Ejemplo 1.** Consideremos las matrices:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{4}{5} & \frac{1}{5} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

La matriz de cocientes es

$$\mathbf{C} = \begin{bmatrix} \frac{15}{10} & \frac{15}{20} \\ \frac{24}{20} & \frac{6}{10} \end{bmatrix} = \frac{3}{20} \begin{bmatrix} 10 & 5 \\ 8 & 4 \end{bmatrix},$$

tiene rango unidad y por tanto  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles. Luego  $v_1 \propto \frac{15}{10} + \frac{24}{20} = \frac{54}{20}$  y  $v_2 \propto \frac{15}{20} + \frac{6}{10} = \frac{27}{20}$ . Por otra parte,  $u_1 \propto \frac{15}{10} + \frac{15}{20} = \frac{45}{20}$  y  $u_2 \propto \frac{24}{20} + \frac{6}{10} = \frac{36}{20}$ . Podemos tomar por tanto  $(v_1, v_2) \propto (2, 1)$  y  $(u_1, u_2) \propto (5, 4)$

El producto  $(5, 4)^t(2, 1)$  debe ser una matriz que, elemento a elemento, es proporcional a  $\mathbf{C}$ .

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix} [2, 1] = \begin{bmatrix} 10 & 5 \\ 8 & 4 \end{bmatrix}$$

Multiplicando esta matriz por  $\frac{3}{20}$  se obtiene  $\mathbf{C}$ .

**Ejemplo 2.** Las matrices

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{3}{5} & \frac{2}{5} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

son incompatibles pues la matriz de cocientes  $\mathbf{C}$  tiene, claramente, rango dos:

$$\mathbf{C} = \begin{bmatrix} \frac{15}{10} & \frac{15}{20} \\ \frac{18}{20} & \frac{12}{10} \end{bmatrix} = \frac{3}{20} \begin{bmatrix} 10 & 5 \\ 6 & 8 \end{bmatrix}$$

Estudiamos el caso general. Supongamos que  $N_{\mathbf{A}} = N_{\mathbf{B}} = N$  (lo contrario indica incompatibilidad), pero  $N$  distinto del producto cartesiano  $\{1, \dots, L\} \times \{1, \dots, M\}$ .

Observemos que, en este caso, la matriz de los cocientes  $\mathbf{C}$  no está completamente determinada por las matrices candidatas a distribuciones condicionadas  $\mathbf{A}$  y  $\mathbf{B}$ . Así si, por ejemplo, (Arnold Castillo y Sarabia 1992)

$$\mathbf{A} = \begin{bmatrix} 1/6 & 0 & 3/14 \\ 0 & 1/4 & 4/14 \\ 5/6 & 3/4 & 7/14 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1/4 & 0 & 3/4 \\ 0 & 1/3 & 2/3 \\ 5/18 & 6/18 & 7/18 \end{bmatrix}$$

Los elementos  $c_{12}$  y  $c_{21}$  de  $\mathbf{C}$  están indeterminados, al ser cociente de dos elementos nulos. La observación fundamental es que, bajo la hipótesis de compatibilidad, esos elementos indeterminados de  $\mathbf{C}$  están definidos, pues son cocientes de probabilidades marginales no nulas (pues, en otro caso el elemento se elimina), y pueden ser determinados, bajo ciertas condiciones, a partir de  $\mathbf{A}$  y  $\mathbf{B}$ .

**Teorema 3.** Si  $\mathbf{A}$  y  $\mathbf{B}$  determinan la distribución conjunta de un vector  $(X, Y)$  discreto y finito entonces, la matriz  $\mathbf{C}$  puede completarse de forma que tenga rango unidad.

#### Demostración

Para cada  $(i, j) \notin N$  nos basta tomar

$$c_{i,j} = \frac{P[X = x_i]}{P[Y = y_j]} = \frac{\tau_i}{\eta_j}$$

Obsérvese que estos  $c_{i,j}$  están definidos pues  $P[Y = y_j] \neq 0$  pues, en otro caso,  $y_j$  se habría eliminado del problema.

□

La utilidad del resultado que acabamos de obtener queda patente en el siguiente ejemplo

**Ejemplo 3.** Considerese

$$\mathbf{A} = 10^{-1} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 4 & 2 & 2 \\ 5 & 0 & 2 & 4 \\ 4 & 4 & 5 & 3 \end{bmatrix} \quad \text{y} \quad \mathbf{B} = 10^{-1} \begin{bmatrix} 2 & 1 & 6 & 1 \\ 0 & 4 & 2 & 4 \\ 4 & 0 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

entonces la matriz  $\mathbf{C}$  viene dada (parcialmente) por

$$\mathbf{C} = \begin{bmatrix} 1/2 & 2 & 1/6 & 1 \\ * & 1 & 1 & 1/2 \\ 5/4 & * & 2/3 & 4/3 \\ 4 & 2 & 5/3 & 3/4 \end{bmatrix}$$

$\mathbf{A}$  y  $\mathbf{B}$  son incompatibles pues, por ejemplo

$$\begin{vmatrix} 2 & 1/6 \\ 1 & 1 \end{vmatrix} \neq 0$$

Nótese que si  $\mathbf{C}$  puede completarse de forma que tenga rango unidad podremos aplicarle el Teorema 2 y tenemos el siguiente

**Corolario 1.** En las condiciones anteriores si  $\mathbf{C}$  puede completarse de forma que tenga rango unidad entonces,  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles.

Utilizaremos datos de Arnold, Castillo y Sarabia, (1992) para ilustrar el uso del corolario que acabamos de obtener.

**Ejemplo 4.** Si tomamos las matrices

$$\mathbf{A} = \begin{bmatrix} \frac{1}{6} & 0 & \frac{3}{14} \\ 0 & \frac{1}{4} & \frac{4}{14} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{14} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{5}{18} & \frac{6}{18} & \frac{7}{18} \end{bmatrix}$$

Entonces

$$\mathbf{C} = \begin{bmatrix} \frac{4}{6} & * & \frac{4}{14} \\ * & \frac{3}{4} & \frac{6}{14} \\ \frac{18}{6} & \frac{9}{4} & \frac{18}{14} \end{bmatrix}$$

Haciendo

$$\begin{vmatrix} \frac{4}{6} & c_{12} \\ \frac{18}{6} & \frac{9}{4} \end{vmatrix} = 0 \quad \begin{vmatrix} c_{21} & \frac{3}{4} \\ \frac{18}{6} & \frac{9}{4} \end{vmatrix} = 0$$

Obtenemos  $c_{12} = \frac{1}{2}$  y  $c_{21} = 1$  y así la matriz  $\mathbf{C}$  es de rango unidad.

$$\mathbf{C} = \begin{bmatrix} \frac{4}{6} & \frac{1}{2} & \frac{4}{14} \\ 1 & \frac{3}{4} & \frac{6}{14} \\ \frac{18}{6} & \frac{9}{4} & \frac{18}{14} \end{bmatrix}$$

y aplicando resultados anteriores se tiene  $\mathbf{u} \propto \left(\frac{61}{42}, \frac{61}{28}, \frac{183}{28}\right)$  y  $\mathbf{v} \propto \left(\frac{28}{6}, \frac{14}{4}, \frac{28}{14}\right)$ .

En ocasiones, no es posible completar  $\mathbf{C}$  de manera única y de forma que tenga rango unidad.

**Ejemplo 5.** Consideramos las matrices

$$\mathbf{A} = \begin{bmatrix} 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{3}{4} & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 & \frac{4}{5} & 0 & \frac{1}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \end{bmatrix}$$

La matriz (parcial) de cocientes es

$$\mathbf{C} = \begin{bmatrix} * & \frac{5}{6} & * & \frac{5}{3} \\ 1 & * & \frac{1}{2} & * \\ * & \frac{2}{3} & * & \frac{4}{3} \\ 2 & * & 3 & * \end{bmatrix}$$

Si hacemos  $c_{11} = \lambda \neq 0$  obtenemos

$$\mathbf{C} = \begin{bmatrix} \lambda & \frac{5}{6} & \frac{\lambda}{2} & \frac{5}{3} \\ 1 & \frac{5}{6\lambda} & \frac{1}{2} & \frac{5}{3\lambda} \\ \frac{4}{5}\lambda & \frac{2}{3} & \frac{2}{5}\lambda & \frac{4}{3} \\ 2 & \frac{5}{3\lambda} & 1 & \frac{10}{3\lambda} \end{bmatrix}$$

sin más que imponer que tenga rango 1. Para cada valor de  $\lambda$  se obtienen distintas  $\mathbf{C}$  y por tanto esta matriz no se puede completar de manera única de forma que tenga rango unidad. Además, para cada  $\lambda$  se tendrán distintos vectores  $\mathbf{u}$  y  $\mathbf{v}$ . En este caso,

$$\mathbf{u} \propto \left( \frac{5}{2} + \frac{3}{2}\lambda, \frac{3}{2} + \frac{5}{2\lambda}, 2 + \frac{6}{5}\lambda, 3 + \frac{5}{\lambda} \right) \quad \mathbf{v} \propto \left( 3 + \frac{9}{5}\lambda, \frac{3}{2} + \frac{15}{\lambda}, \frac{3}{2} + \frac{9}{10}\lambda, 3 + \frac{5}{\lambda} \right)$$

Otro resultado de interés, debido a Gupta y Vargas (1990), que da una condición necesaria y suficiente para el caso en que las matrices  $\mathbf{A}$  y  $\mathbf{B}$  tienen una fila (pongamos  $i_0$ ) y una columna (por ejemplo  $j_0$ ) de elementos no nulos. Formalmente podemos enunciarlo en los siguientes términos:

**Teorema 4.** (Gupta y Vargas, 1990) Sean  $\mathbf{A}$  y  $\mathbf{B}$  dos matrices en las condiciones habituales  $N_A = N_B = N$  y supongamos que  $\exists (i_0, j_0) \in N$  tal que, para cada  $j = 1, \dots, M$   $(i_0, j) \in N$  y que para cada  $i = 1, \dots, L$   $(i, j_0) \in N$ . Entonces,  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles si y sólo si

$$b_{ij} \frac{a_{i,j_0}}{b_{i,j_0}} = \kappa a_{i,j} \frac{b_{i_0,j}}{a_{i_0,j}} \quad \forall (i, j) \in N$$

donde  $\kappa$  es una constante que depende únicamente de  $i_0$  y de  $j_0$  y es distinta para cada elección de estos índices.

Puede comprobarse que si  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles la igualdad se tiene tomando

$$\kappa = \left[ \frac{\sum_{j=1}^M c_{i_0,j}^{-1}}{\sum_{i=1}^L c_{i,j_0}} \right]^{-1}$$

mientras que el recíproco se obtiene tomando la distribución conjunta

$$\begin{aligned} p_{i,j} &= \kappa' b_{i,j} \frac{a_{i,j_0}}{b_{i,j_0}} && \text{para } (i, j) \in N \\ p_{i,j} &= 0 && \text{para } (i, j) \notin N \end{aligned}$$

siendo

$$\kappa' = \left[ \sum_{i=1}^L c_{i,j_0} \right]^{-1}$$

**Ejemplo 6.** Consideramos las matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{1}{4} \\ 1 & 1 & \frac{1}{3} & \frac{1}{4} \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \end{bmatrix}$$

Entonces  $N_A = N_B = N = \{(1,3), (1,4), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)\}$

La elección de  $i_0$  y  $j_0$  en este caso ofrece dos posibilidades.  $(i_0, j_0) = (3,3)$  o bien  $(i_0, j_0) = (3,4)$ ; tomemos la primera posibilidad

$$\kappa = \left[ \frac{\sum_{j=1}^4 c_{3j}^{-1}}{\sum_{i=1}^3 c_{i,3}} \right]$$

Comprobar las igualdades es trivial.

La distribución conjunta será:  $p_{i,j} = \kappa' b_{i,j} \frac{a_{i,3}}{p_{i,3}}$  para  $(i,j) \in N$  con  $\kappa' = \left[ \sum_{i=1}^3 c_{i,3} \right] = \frac{3}{10}$ . Obteniéndose, entonces,

$$(p_{i,j}) = \frac{1}{10} \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & 2 & 1 & 1 \end{bmatrix}$$

Obsérvese que en el ejemplo 5 no es aplicable este resultado dada la estructura de las matrices  $\mathbf{A}$  y  $\mathbf{B}$  que contienen elementos nulos en todas sus filas y columnas.

#### 4. UNICIDAD

En toda esta sección supondremos que las matrices  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles, lo que implica, en particular, que sus elementos nulos están en la misma posición. Un primer resultado relativo a la unicidad es el siguiente:

**Teorema 5.** (Gupta y Varga 1990) Si  $\mathbf{A}$  y  $\mathbf{B}$  tienen al menos una fila y una columna de elementos no nulos, la distribución conjunta que determinan  $\mathbf{A}$  y  $\mathbf{B}$  es única.

Así las matrices del ejemplo 5 de la sección anterior son compatibles como ya vimos, además, al tener una fila y una columna formada con elementos no nulos, podemos concluir que la distribución conjunta que definen es única.

Otro enfoque de este problema es el siguiente. Supóngase que  $\mathbf{C}$  puede ser completada de manera única y de forma que tenga rango unidad, entonces aplicando la descomposición singular obtenemos

$$\mathbf{C} = \begin{pmatrix} u_1 \\ \vdots \\ u_L \end{pmatrix} (v_1 \dots v_M)$$

donde el valor singular lo hemos incluido en uno de los vectores  $\mathbf{u}$  o  $\mathbf{v}$ . Puesto que el vector  $\mathbf{u}$  es autovector asociado al único autovalor no nulo de  $\mathbf{C}'\mathbf{C}$ , y puesto que este autovector es simple,  $\mathbf{u}$  es único salvo proporcionalidad (recordar que se toma no negativo) pero, la marginal que corresponde a  $\mathbf{u}$  está obviamente normalizada y por consiguiente es única, de ahí que la distribución conjunta también. En definitiva se ha establecido el siguiente

**Teorema 6.** *Sean  $\mathbf{A}$  y  $\mathbf{B}$  dos matrices compatibles. Si la matriz de cocientes puede ser completada de manera única de forma que tenga rango unidad, entonces, la distribución que determinan  $\mathbf{A}$  y  $\mathbf{B}$  es única.*

El siguiente ejemplo, que utiliza matrices de Arnold, Castillo y Sarabia (1992) aclara este resultado.

**Ejemplo 7.** Consideremos las matrices

$$\mathbf{A} = \begin{bmatrix} \frac{1}{6} & 0 & \frac{3}{14} \\ 0 & \frac{1}{4} & \frac{4}{14} \\ \frac{5}{6} & \frac{3}{4} & \frac{7}{14} \end{bmatrix} \quad \text{y} \quad \mathbf{B} = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{5}{18} & \frac{6}{18} & \frac{7}{18} \end{bmatrix}$$

que tienen como matriz de cocientes

$$\mathbf{C} = \begin{bmatrix} \frac{4}{6} & * & \frac{4}{14} \\ * & \frac{3}{4} & \frac{6}{14} \\ \frac{18}{6} & \frac{9}{4} & \frac{18}{14} \end{bmatrix}$$

que puede ser completada de manera única de forma que tenga rango unidad:

$$\mathbf{C} = \begin{bmatrix} \frac{4}{6} & \frac{1}{2} & \frac{4}{14} \\ 1 & \frac{3}{4} & \frac{6}{14} \\ \frac{18}{6} & \frac{9}{4} & \frac{18}{14} \end{bmatrix}$$

y vimos que es compatible. Ahora podemos afirmar, además, que la distribución conjunta es única.

Como ya vimos en la sección precedente no siempre es posible completar  $\mathbf{C}$  de manera única con la condición  $rg\mathbf{C} = 1$ . Así pues, llegados a este punto, parece evidente la necesidad de caracterizar dichas matrices. Para ello, y con el fin de aligerar la notación, supondremos que mediante cambios de filas y columnas se han llevado el mayor número posible de elementos desconocidos de  $\mathbf{C}$  a una submatriz de ella situada en la esquina superior izquierda. Obsérvese que esto no conlleva ninguna pérdida de generalidad, supone únicamente una reordenación de los valores del vector aleatorio.

Supondremos, además, que todos aquellos valores desconocidos de  $\mathbf{C}$  que se puedan calcular unívocamente vía la condición  $rg\mathbf{C} = 1$ , han sido calculados. Esto implica que esos elementos han sido determinados mediante menores de dimensión dos con un único elemento desconocido. Estos elementos serán denominados *determinables o calculables*. En definitiva, la forma de la matriz  $\mathbf{C}$  es

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

donde  $\mathbf{C}_{11}$  es una matriz de orden  $r \times s$ , con todos sus elementos indeterminados. Además de los elementos de  $\mathbf{C}_{11}$ , es posible que existan otros elementos indeterminados en  $\mathbf{C}_{12}$ ,  $\mathbf{C}_{21}$  y  $\mathbf{C}_{22}$ . En estas condiciones se tiene

**Teorema 7.** Si  $\mathbf{C}_{22}$  tiene todos sus elementos indeterminados y  $\mathbf{C}_{12}$ ,  $\mathbf{C}_{21}$  tienen todos sus elementos determinados, entonces, la distribución que determinan  $\mathbf{A}$  y  $\mathbf{B}$  no es única.

#### Demostración

Utilizando la descomposición singular para la matriz  $\mathbf{C}_{12}$  ( $rg\mathbf{C}_{12} = 1$ ) se puede determinar un par de vectores  $(u_1, u_2, \dots, u_r)^t$  y  $(v_{s+1}, v_{s+2}, \dots, v_M)^t$  de forma que

$$\mathbf{C}_{12} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{pmatrix} (v_{s+1}, v_{s+2}, \dots, v_M)$$

pero también, si  $\lambda > 0$ :

$$\mathbf{C}_{12} = \begin{pmatrix} \lambda u_1 \\ \lambda u_2 \\ \vdots \\ \lambda u_r \end{pmatrix} \left( \frac{1}{\lambda} v_{s+1}, \frac{1}{\lambda} v_{s+2}, \dots, \frac{1}{\lambda} v_M \right)$$

Análogamente utilizando  $\mathbf{C}_{21}$ :

$$\mathbf{C}_{21} = \begin{pmatrix} u_{r+1} \\ u_{r+2} \\ \vdots \\ u_L \end{pmatrix} (v_1, v_2, \dots, v_s)$$

Pero también si  $\mu > 0$ :

$$\mathbf{C}_{21} = \begin{pmatrix} \mu u_{r+1} \\ \mu u_{r+2} \\ \vdots \\ \mu u_M \end{pmatrix} \left( \frac{1}{\mu} v_1, \frac{1}{\mu} v_2, \dots, \frac{1}{\mu} v_s \right)$$

Ahora los vectores

$$\mathbf{u} = (\lambda u_1, \lambda u_2, \dots, \lambda u_r, \mu u_{r+1}, \mu u_{r+2}, \dots, \mu u_L)^t$$

$$\mathbf{v} = \left( \frac{1}{\mu} v_1, \frac{1}{\mu} v_2, \dots, \frac{1}{\mu} v_s, \frac{1}{\lambda} v_{s+1}, \frac{1}{\lambda} v_{s+2}, \dots, \frac{1}{\lambda} v_M \right)$$

son proporcionales a las distribuciones marginales y si  $\lambda \neq \mu$  es evidente que no son únicos aún normalizándolos. En consecuencia la distribución conjunta no es única.

□

**Teorema 8.** Si la distribución conjunta que determinan  $\mathbf{A}$  y  $\mathbf{B}$  no es única, la matriz de cocientes  $\mathbf{C}$  puede escribirse, permutando filas y columnas, como

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

con  $\mathbf{C}_{11}$  y  $\mathbf{C}_{22}$  matrices de elementos indeterminados.

*Demostración*

Supongamos que por permutación de filas y columnas la matriz  $\mathbf{C}$  se ha escrito en la forma

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

con  $\mathbf{C}_{11}$  de elementos desconocidos. Esta forma de escribir  $\mathbf{C}$  no es única, nosotros seleccionaremos la siguiente:  $\mathbf{C}_{11}$  tiene el mayor número posible de columnas de entre todas las particiones de  $\mathbf{C}$  del tipo anterior y, para ese número de columnas, el mayor número de filas posibles. Supongamos que las dimensiones resultantes son  $r \times s$ . Probaremos que entonces,  $\mathbf{C}_{22}$  está formada por elementos desconocidos. En efecto, supongamos, sin pérdida de generalidad, que el elemento  $c_{r+1,s+1}$  es conocido (o ha sido calculado en algún paso anterior) entonces, algún elemento de su columna  $c_{ij}$ ,  $i = 1, \dots, r; j = s + 1, \dots, M$  es conocido pues, si no, las dimensiones de  $\mathbf{C}_{11}$  no serían máximas. Supongamos, sin pérdida de generalidad, que es  $c_{r+1,s+1}$ . Análogamente trabajando con la fila  $r + 1$  tendremos que podemos suponer que  $c_{r+1,s}$  es conocido. Así pues son conocidos  $c_{r+1,s}, c_{r,s+1}, c_{r+1,s+1}$ , pero entonces como el determinante

$$\begin{vmatrix} c_{r,s} & c_{r,s+1} \\ c_{r+1,s} & c_{r+1,s+1} \end{vmatrix} = 0$$

tenemos que  $c_{rs}$  es conocido, lo que es absurdo. Obsérvese que  $\mathbf{C}_{11}$  existe realmente (i.e.  $r > 0$  y  $s > 0$ ) pues la distribución conjunta no es única y en consecuencia existe, al menos, un elemento no determinable ya que de no ser así, todos los elementos de  $\mathbf{C}$  se pueden determinar y, en consecuencia, aplicando resultados anteriores la distribución conjunta es única.

□

Veamos algunos ejemplos:

**Ejemplo 8.** Retomando el ejemplo 4 de la sección anterior, donde ningún elemento desconocido es determinable, y permutando filas y columnas

$$\mathbf{C} = \left[ \begin{array}{cccc} * & \frac{5}{6} & * & \frac{5}{3} \\ 1 & * & \frac{1}{2} & * \\ * & \frac{2}{3} & * & \frac{4}{3} \\ 2 & * & 3 & * \end{array} \right] \longrightarrow \left[ \begin{array}{cccc} * & * & \frac{5}{6} & \frac{5}{3} \\ 1 & \frac{1}{2} & * & * \\ * & * & \frac{2}{3} & \frac{4}{3} \\ 2 & 3 & * & * \end{array} \right] \longrightarrow \left[ \begin{array}{cccc} * & * & \frac{5}{6} & \frac{5}{3} \\ * & * & \frac{2}{3} & \frac{4}{3} \\ 1 & \frac{1}{2} & * & * \\ 2 & 3 & * & * \end{array} \right]$$

con lo que, aplicando resultados anteriores, se concluye que, aún siendo compatibles, la distribución conjunta no es única.

El siguiente ejemplo muestra la utilización de la descomposición singular en submatrices para obtener diversas distribuciones marginales.

**Ejemplo 9.** Las siguientes matrices pueden considerarse como distribuciones condicionadas de un vector aleatorio finito:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{bmatrix}$$

La matriz de cocientes viene dada por:

$$\mathbf{C} = \begin{bmatrix} * & * & * & * & \frac{1}{2} & \frac{1}{2} & * \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} & * & * & * & * \\ * & * & * & \frac{1}{2} & * & * & \frac{1}{2} \\ 1 & * & 1 & * & * & * & * \end{bmatrix}$$

Intercambiando la 2<sup>a</sup> y 3<sup>a</sup> fila y las columnas 5<sup>a</sup> y 7<sup>a</sup> obtenemos:

$$\mathbf{C} \rightarrow \begin{bmatrix} * & * & * & * & * & \frac{1}{2} & \frac{1}{2} \\ * & * & * & \frac{1}{2} & \frac{1}{2} & * & * \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} & * & * & * & * \\ 1 & * & 1 & * & * & * & * \end{bmatrix}$$

El elemento  $c_{42}$  es determinable mediante la condición  $\text{rg } \mathbf{C} = 1$ , obteniéndose que  $c_{42} = 1/2$ . Los elementos conocidos forman menores de orden dos nulos y, en consecuencia, las matrices  $\mathbf{A}$  y  $\mathbf{B}$  se pueden considerar distribuciones condicionadas si  $\mathbf{C}$  se puede completar de forma que tenga rango unidad. La matriz  $\mathbf{C}_{12}$  reproduce la estructura por cajas que estudiamos en el teorema 8 y además  $\mathbf{C}$  también la reproduce.

Completaremos en primer lugar las submatrices de la submatriz  $\mathbf{C}_{1,2}$ : la submatriz que ocupa el lugar 1,2 en  $\mathbf{C}_{12}$  es  $(\frac{1}{2}, \frac{1}{2})$ . Aplicando la descomposición en valores singulares (como siempre incluimos el valor singular en los vectores) obtenemos:

$$\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}^t = \mathbf{u}' \mathbf{v}'^t$$

donde podemos tomar  $\mathbf{u}' = \lambda$ ,  $\mathbf{v}' = \begin{bmatrix} \frac{1}{2\lambda} \\ \frac{1}{2\lambda} \end{bmatrix}$  para cualquier número real  $\lambda > 0$ . Observamos que este  $\mathbf{u}'$  es la componente  $u_1$  del vector  $\mathbf{u}$  que resuelve el problema en la

matriz  $\mathbf{C}$  y  $\mathbf{v}'$  contiene los componentes  $v_6 v_7$  del vector  $\mathbf{v}$  que resuelve el problema en la matriz  $\mathbf{C}$ , así se tiene:

$$u_1 = \lambda \quad v_6 = \frac{1}{2\lambda} \quad v_7 = \frac{1}{2\lambda}$$

Trabajando ahora con la submatriz que ocupa el lugar 2,1 en  $\mathbf{C}_{1,2}$  obtenemos

$$\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}^t = \mu \begin{bmatrix} \frac{1}{2\mu} \\ \frac{1}{2\mu} \end{bmatrix}^t$$

y como antes se tiene

$$u_2 = \mu \quad v_4 = \frac{1}{2\mu} \quad v_5 = \frac{1}{2\mu}$$

con estos vectores es posible reconstruir  $\mathbf{C}_{1,2}$

$$\mathbf{C}_{1,2} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} (v_4 v_5 v_6 v_7) = \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \left( \frac{1}{2\mu}, \frac{1}{2\mu}, \frac{1}{2\lambda}, \frac{1}{2\lambda} \right) = \begin{bmatrix} \frac{\lambda}{2\mu} & \frac{\lambda}{2\mu} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{\mu}{2\lambda} & \frac{\mu}{2\lambda} \end{bmatrix}$$

En consecuencia, la matriz  $\mathbf{C}$  adopta la forma

$$\mathbf{C} = \begin{bmatrix} * & * & * & \frac{\lambda}{2\mu} & \frac{\lambda}{2\mu} & \frac{1}{2} & \frac{1}{2} \\ * & * & * & \frac{1}{2} & \frac{1}{2} & \frac{\mu}{2\lambda} & \frac{\mu}{2\lambda} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} & * & * & * & * \\ 1 & \frac{1}{2} & 1 & * & * & * & * \end{bmatrix}$$

Para completar  $\mathbf{u}$  y  $\mathbf{v}$  necesitamos  $u_3, u_4, v_1, v_2$ , y  $v_3$ , los obtendremos de la descomposición en valores singulares de la matriz  $\mathbf{C}_{2,1}$

$$\begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ 1 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/2 \end{bmatrix} (2, 1, 2) = \begin{bmatrix} \frac{1}{3}\omega \\ \frac{1}{2}\omega \end{bmatrix} \left( \frac{2}{\omega}, \frac{1}{\omega}, \frac{2}{\omega} \right)$$

en consecuencia

$$u_3 = \frac{1}{3}\omega, u_4 = \frac{1}{2}\omega, v_1 = \frac{2}{\omega}, v_2 = \frac{1}{\omega}, y v_3 = \frac{2}{\omega}$$

$$\mathbf{v}' = \left( \frac{2}{\omega}, \frac{1}{\omega}, \frac{2}{\omega}, \frac{1}{2\mu}, \frac{1}{2\mu}, \frac{1}{2\lambda}, \frac{1}{2\lambda} \right)$$

Así se obtiene

$$\mathbf{C} = \begin{bmatrix} \lambda \\ \mu \\ \frac{1}{3}\omega \\ \frac{1}{2}\omega \end{bmatrix} \left( \frac{2}{\omega}, \frac{1}{\omega}, \frac{2}{\omega}, \frac{1}{2\mu}, \frac{1}{2\mu}, \frac{2}{2\lambda}, \frac{2}{2\lambda} \right) = \begin{bmatrix} \frac{2\lambda}{\omega} & \frac{\lambda}{\omega} & \frac{2\lambda}{\omega} & \frac{\lambda}{2\mu} & \frac{\lambda}{2\mu} & \frac{1}{2} & \frac{1}{2} \\ \frac{2\mu}{\omega} & \frac{\mu}{\omega} & \frac{2\mu}{\omega} & \frac{1}{2} & \frac{1}{2} & \frac{\mu}{2\lambda} & \frac{\mu}{2\lambda} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} & \frac{\omega}{6\mu} & \frac{\omega}{6\mu} & \frac{\omega}{6\lambda} & \frac{\omega}{6\lambda} \\ 1 & \frac{1}{2} & 1 & \frac{\omega}{4\mu} & \frac{\omega}{4\mu} & \frac{\omega}{4\lambda} & \frac{\omega}{4\lambda} \end{bmatrix}$$

Para obtener distribuciones marginales, bastará con normalizar los vectores  $\mathbf{u}$  y  $\mathbf{v}$ , sin embargo, aún en ese caso, dando valores positivos a  $\lambda$ ,  $\mu$  y  $\omega$  obtenemos distintas de estas distribuciones con lo que se tiene la no unicidad.

Una vía alternativa en el estudio de la unicidad es el uso de las cadenas de Markov tal y como proponen Arnold y Press (1989). Este enfoque surge de la siguiente observación:

Si  $\mathbf{A}$  y  $\mathbf{B}$  son compatibles existe un vector bidimensional  $(X, Y)$  de variables aleatorias finitas tal que para cada  $i = 1, 2, \dots, L$  y para cada  $j = 1, 2, \dots, M$ .

$$a_{1,j} = P[X = x_i | Y = y_j] \quad b_{i,j} = P[Y = y_j | X = x_i]$$

denotamos, como en la sección precedente:

$$\tau_i = P[X = x_i] \quad \eta_j = P[Y = y_j]$$

que son las distribuciones marginales, entonces se tienen los siguientes resultados por aplicación de teorema de la probabilidad total.

$$P[X = x_i] = \sum_{j=1}^M P[X = x_i | Y = y_j] P[Y = y_j] = \sum_{j=1}^M a_{i,j} \eta_j = (a_{i,1} a_{i,2} \dots a_{i,M}) \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_M \end{bmatrix}$$

es decir la fila  $i$ -ésima de  $\mathbf{A}$ , por el vector de probabilidad marginal, en consecuencia

$$\tau = \mathbf{A}\eta$$

Por otra parte

$$P[Y = y_j] = \sum_{i=1}^L P[Y = y_j | X = x_i] P[X = x_i] = \sum_{i=1}^L b_{i,j} \tau_i = (b_{1,j} b_{2,j} \dots b_{L,j}) \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_L \end{bmatrix}$$

esto es la columna  $j$ -ésima de  $\mathbf{B}$  por la distribución marginal de  $X$ ; así

$$\eta = \mathbf{B}^t \tau$$

Combinando lo anterior

$$\tau = A\eta = A\mathbf{B}^t \tau$$

puesto que  $A\mathbf{B}^t$  es una matriz estocástica por columnas, al serlo  $A$  y  $\mathbf{B}^t$ ,  $A\mathbf{B}^t$  puede ser considerada como la matriz de transición de una cadena de Markov de  $L$  estados y, consecuentemente,  $\tau$  se interpreta como una distribución invariante o estacionaria de la cadena.

Arnold, y Press (1989) observan que si  $A$  y por tanto  $\mathbf{B}$  tienen sus elementos no nulos se tiene la unicidad, resultado que obtuvimos más arriba.

Los resultados anteriores se pueden generalizar al tratamiento de distribuciones discretas de recorrido infinito con un número finito de elementos (ver Pérez-Villalta, 1997).

## 5. REFERENCIAS

- Abrahams J. y Thomas J.B. (1984). «A note on the characterization of bivariate densities by conditional densities». *Communications in Statistics: Theory and Methods*, 13(3), 395-400
- Arnold B.C., Castillo E. y Sarabia J.M. (1992). *Conditionally specified distributions*. Lectures Notes in Statistics, Volume 73, Springer Verlang. Berlin.
- (1996) «Bayesian analysis for classical distributions using conditionally specified priors». *Sankhyā, Ser. B*, 60, 228-245.
- (1999). *Conditional specification of statistical model*. Springer Series in Statistics. Springer. New York.
- Arnold B.C. y Gokhale, D.V. (1994). «On uniform marginal representations of contingency tables». *Statistics and Probability Letter*, 21, 311-316
- Arnold B.C. y Press S.J. (1989). «Compatible conditional distributions». *Journal of the American Statistical Association*, 84, 152-156
- Basulto J. (1995). «Una propuesta para obtener distribuciones a priori públicas». *Publicación del Departamento de Economía Aplicada I de la Universidad de Sevilla*.
- Casella G. y George E.I. (1992). «Explaining the Gibbs sampling». *The American Statistician*, 46(3), 167-174.
- Casella G. (1996). «Statistical inference and MonteCarlo algorithms». *Test*, 5, 249-344.
- Gourieroux Ch. y Monfort A. (1979). «On the characterization of a joint probability distribution by conditional distributions». *Journal of Econometrics*, 10 115-118.

- Gupta A.K. y Varga T. (1990). «Characterization of joint density by conditional densities». *Communications in Statistics: Theory and Methods*, 19(12), 4643-4659.
- Pérez-Villalta, R. (1997). *Distribuciones bivariantes especificadas con condicionadas o Marginales*. Trabajo de Investigación. Departamento de Economía Aplicada I Universidad de Sevilla.
- Schott, J.R. (1997). *Matrix analysis for statistics*. John Wiley and sons. New York.

## **ENGLISH SUMMARY**

### **FINITE CONDITIONALLY SPECIFIED VARIABLES**

ROMÁN PÉREZ-VILLALTA

Universidad de Sevilla\*

*In this research, the existence and uniqueness of bidimensional vectors of discrete variables with finite rank is studied when their condition distribution are set. To do that and after revising the literature about the subject, we provide several results that relate some issues of matricial algebra, specially the singular descomposition, to the problem that we are interested in.*

**Keywords:** Conditional distribution, conditional specification, matrix range, singular descomposition, Markov chain

**AMS Classification (MSC 2000):** 60E05, 62E10

---

\*Facultad de Económicas y Empresariales, Dpto. de Economía Aplicada I. Avda. Ramón y Cajal, 1.  
41018 Sevilla. España. villalta@cica.es

-Received December 1999.

-Accepted September 2000.

## 1. INTRODUCTION

Given a bivariate random vector  $(X, Y)$  we can specify its density function in several ways: providing either that function directly, a marginal and a conditional function (in the same or different component), two marginal densities... Depending of the case, the density function of the vector will exist or not, and moreover, it will be the unique one or not.

The case where two conditional functions are given, objective of this research, has been studied extensively by Arnold and Press(1989), and Arnold, Castillo and Sarabia(1979), although it can be mentioned pioneering research such as Gourieroux and Monfort (1979), Abrahams and Thomas (1984). Let's note before continuing that, the previous list of researches do not cover the extent literature in the matter (with regard to the subject under discussion).

This research reviews the existing results in the case where the random variables are finite and contributes(illustrates) some additional result using elements of matricial calculus (calculos matriciales) and examples that clarify the different methods exposed.

## 2. POSING THE PROBLEM AND NOTATION

Let two matrices be  $\mathbf{A} = (a_{i,j})$  y  $\mathbf{B} = (b_{i,j})$  (de orden)  $L \times M$  that verifies  $a_{i,j} \geq 0, b_{i,j} \geq 0 \forall i = 1, \dots, L; \forall j = 1, \dots, M$  and moreover,

$$\sum_{i=1}^L a_{i,j} = 1, \sum_{j=1}^M b_{i,j} = 1.$$

The question to elucidate is, under what conditions two matrices  $\mathbf{A}$  y  $\mathbf{B}$ , with dimensions  $L \times M$ , in the conditions before, can be considered conditional distributions of a certain random vector  $(X, Y)$  of discrete variables with finite range. That is, if  $x_1, x_2, \dots, x_L$  are the possible values of  $X$  and  $y_1, y_2, \dots, y_M$  those of  $Y$ , when can the following expressions be verify?

$$a_{i,j} = P[X = x_i | Y = y_j] \quad i = 1 \dots L$$

$$b_{i,j} = P[Y = y_j | X = x_i] \quad j = 1 \dots M$$

In the same way, we are interested in knowing under what conditions can exist two vectors  $\tau$  and  $\eta$ , with dimensions  $L$  and  $M$  respectively, non negative y so that the normalized version (the sum of the coordinates are the unity)of  $\tau$  and  $\eta$  constitutes the respective marginal distributions. Note that we can suppose  $\tau_i b_{i,j} = \eta_j a_{i,j} \quad i = 1, \dots, L \quad j = 1, \dots, M$ , since if a  $\tau_i$  or a  $\eta_j$  is null, the value  $x_i$  or  $y_j$  has all of its probabilities null and can be eliminated of the problem.

This first question is called compatibility problem and, in that context, the matrices  $\mathbf{A}$  and  $\mathbf{B}$  are denominated candidates to conditional distributions, and moreover, if the answer the problem is affirmative it can be said that  $\mathbf{A}$  and  $\mathbf{B}$  are compatible.

A second problem to be considered is the uniqueness of these variables  $X$  and  $Y$  or vectores  $\tau$  y  $\eta$ , when it exists.

For the matrices  $\mathbf{A}$  and  $\mathbf{B}$ , we define the sets

$$N_A = \{(i, j) : a_{i,j} > 0\}, N_B = \{(i, j) : b_{i,j} > 0\}$$

when  $N_A = N_B$  are obtained, it will be denoted by  $N$  indistinctly

### 3. COMPATIBILITY

In this section we study the following results:

**Theorem 1.** (Arnold y Press, 1989) *Two matrices  $\mathbf{A}$  y  $\mathbf{B}$  under the conditions of the section before are compatible if and only if  $N_A = N_B = N$  an moreover there exist two vectors,  $\mathbf{u} = (u_1, u_2, \dots, u_L)$  y  $\mathbf{v} = (v_1, v_2, \dots, v_M)$  so that*

$$c_{i,j} = \frac{a_{i,j}}{b_{i,j}} = u_i v_j; \quad \forall (i, j) \in N$$

**Theorem 2.** *Under the previous conditions  $\mathbf{A}$  and  $\mathbf{B}$  are compatibles if and only if the rank of  $\mathbf{C}$  is equal to one.*

The result before can be found, under other perspective in Arnold and Gokhale (1994), and Arnold, Castillo and Sarabia (1999). When there exist null elements in  $\mathbf{A}$  y  $\mathbf{B}$ , the matrix  $\mathbf{C}$  contains undetermined (indeterminados) elementos, however these can be calculated:

**Theorem 3.** *If  $\mathbf{A}$  and  $\mathbf{B}$  determin the joint distribution of a discrete and finite vector  $(X, Y)$  then, the matrix  $\mathbf{C}$  can be completed so that it has rank equal to one (rango unidad).*

**Corollary 1.** *Under the previous conditions if  $\mathbf{C}$  can be completed so that it has rank equal to one then,  $\mathbf{A}$  and  $\mathbf{B}$  are compatibles.*

**Theorem 4.** (Gupta y Vargas, 1990) Let  $\mathbf{A}$  and  $\mathbf{B}$  be two matrices in the usual conditions  $N_A = N_B = N$  and let's suppose that  $\exists(i_0, j_0) \in N$  such that, for every  $j = 1, \dots, M$   $(i_0, j) \in N$  and for every  $i = 1, \dots, L$   $(i, j_0) \in N$ . Then,  $\mathbf{A}$  and  $\mathbf{B}$  are compatible iff (if and only if)

$$b_{ij} \frac{a_{i,j_0}}{b_{i,j_0}} = \kappa a_{i,j} \frac{b_{i_0,j}}{a_{i_0,j}} \quad \forall (i, j) \in N$$

where  $\kappa$  is a constant that depends only of  $i_0$  and  $j_0$  and it is different for every choice of these indexes.

#### 4. UNIQUENESS

**Theorem 5.** (Gupta y Varga 1990) If  $\mathbf{A}$  and  $\mathbf{B}$  have at least a row and a column of no null elements, the joint distribution that  $\mathbf{A}$  and  $\mathbf{B}$  determine is unique.

**Theorem 6.** Let  $\mathbf{A}$  and  $\mathbf{B}$  be two compatible matrices. If the matrix of the quotients can be completed in a unique way so that it has rank equal to one, then the distribution that  $\mathbf{A}$  and  $\mathbf{B}$  determine is unique.

It is not always possible to complete  $\mathbf{C}$  in a unique way with the condition  $rg \mathbf{C} = 1$ . So, at this point, the necessity of typifying such matrices seems obvious. To get that and with the objective of making the notation easier, we will suppose that by changes of rows and columns, the most possible number of elements of  $\mathbf{C}$  has been moved to a sub-matrix of it, set at the top left corner.

Moreover, we will suppose that all those unknown values of  $\mathbf{C}$  that can be calculated uniquely, according to the condition  $rg \mathbf{C} = 1$ , have been calculated. These elements will be denominated *determinable or calculatable*. In summary, the form of the matrix  $\mathbf{C}$  is

$$\mathbf{C} = \left[ \begin{array}{cc} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{array} \right]$$

where  $\mathbf{C}_{11}$  is a matrix with dimension  $r \times s$ , with all of its elements undetermined. Besides the elements of  $\mathbf{C}_{11}$ , it is possible that other undetermined elements exist in  $\mathbf{C}_{12}$ ,  $\mathbf{C}_{21}$  y  $\mathbf{C}_{22}$ .

**Theorem 7.** If  $\mathbf{C}_{22}$  has all of their elements undetermined and  $\mathbf{C}_{12}$ ,  $\mathbf{C}_{21}$  have all of its elements determined, then the distribution that determines  $\mathbf{A}$  and  $\mathbf{B}$  is not unique.

**Theorem 8.** If the joint distribution that determines  $\mathbf{A}$  and  $\mathbf{B}$  is not unique, the matrix of quotients  $\mathbf{C}$  can be written permutating rows and columns, such as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

with  $\mathbf{C}_{11}$  and  $\mathbf{C}_{22}$  matrices of undetermined elements.

## ANÁLISIS DE DETECCIÓN DE RAÍCES UNITARIAS EN SERIES DE TIEMPO. UN ENFOQUE METODOLÓGICO CON TESTS NO SIMILARES

JOSÉ ÁNGEL ROLDÁN CASAS

RAFAELA DIOS PALOMARES

Universidad de Córdoba\*

*El presente artículo recoge los resultados de una investigación llevada a cabo con el fin de analizar, desde la perspectiva de la no similaridad, las distribuciones de los distintos estadísticos planteados por Dickey y Fuller para contrastar la presencia de raíz unitaria. Asimismo, se definen zonas de rechazo y aceptación de las hipótesis nulas para cada estadístico, considerando las distintas distribuciones del mismo, y se estudian las situaciones con las que nos podemos encontrar de cara a deducir una pauta de comportamiento que minimice el riesgo de error. Se presenta un ejemplo en el que se pone de manifiesto que la no similaridad de los contrastes tradicionales de raíz unitaria nos puede llevar a tomar una decisión equivocada. Teniendo en cuenta dicha no similaridad, se propone una estrategia de contraste secuencial para resolver situaciones de indecisión acerca del rechazo o no de la existencia de raíz unitaria. Finalmente, se lleva a cabo un experimento Monte Carlo para determinar empíricamente la probabilidad de que el valor calculado de los estadísticos implicados en el contraste caigan en las diferentes zonas planteadas.*

**Testing for unit roots in time series: An analysis with non similar tests**

**Palabras clave:** Raíz unitaria, tendencia estocástica, test no similar, test de Dickey-Fuller, experimento Monte Carlo

**Clasificación AMS (MSC 2000):** 62M10

\* Universidad de Córdoba. Dpto. de Estadística e Investigación Operativa. E.S.I.A y de Montes.  
Apartado 3048. 14080 Córdoba.

–Recibido en marzo de 1998.  
–Aceptado en marzo de 2000.

## 1. INTRODUCCIÓN

Una de las prácticas econométricas más usadas en Modelización y Predicción Económica es el Análisis de Series de Tiempo, tanto desde el punto de vista clásico como desde el enfoque Box-Jenkins. Es muy frecuente, además, que una serie cronológica presente tendencia que puede ser de dos tipos: estocástica y/o determinista.

La tendencia estocástica hace que la serie permanezca largos períodos de tiempo por encima o por debajo de su valor central, reflejando así la no estacionariedad en media de la misma. Por ejemplo, una serie con tendencia estocástica podría ser la generada por el proceso  $Y_t = Y_{t-1} + e_t$ , donde  $e_t$  es una serie estacionaria. Este proceso es no estacionario debido a la presencia de una raíz unitaria en su parte autorregresiva. De esta forma, si diferenciamos  $Y_t$ , resulta una serie, que ya sería estacionaria y, por tanto, podría ser modelizada mediante la metodología Box-Jenkins. Es por este motivo que se establece como práctica necesaria en el Análisis de Series de Tiempo, el contraste sobre la presencia de raíces unitarias en la serie original. Una práctica habitual es utilizar el correlograma de la serie para determinar la posible presencia de raíz unitaria. Una caída lineal de las autocorrelaciones estimadas sería, en principio, un indicio de existencia de raíz unitaria.

La tendencia determinista, en cambio, es una cadencia temporal, lineal o no, que por sus características se puede estimar con más o menos precisión. Si se sustrae dicha tendencia de la serie original se obtiene la componente puramente estocástica de la misma. Una serie con tendencia determinista generada con el modelo  $Y_t = \mu + \beta t + e_t$  (con  $e_t$  ruido blanco), en principio, debe presentar un correlograma con una caída exponencial. No obstante, en ocasiones, el dominio de la tendencia determinista lineal sobre la componente estocástica es tal, que el correlograma presenta una caída lineal, lo que podría llevarnos a la idea equivocada de existencia de raíz unitaria.

Por tanto, la simple observación del gráfico o correlograma de una serie, es una buena forma para detectar la no estacionariedad de dicha serie, pero no siempre es válido para distinguir si la tendencia que muestra la serie es de naturaleza determinista o estocástica. Por ello, se hace necesaria la aplicación de un test que nos permita discernir entre una tendencia y otra.

Los contrastes de raíz unitaria pioneros en la literatura, se deben a Fuller (1976), y Dickey y Fuller (1979). Estos autores propusieron un método simple para contrastar la existencia de una raíz unitaria en una serie temporal  $Y_t$  que puede venir generada por tres mecanismos autorregresivos de orden 1 diferentes:

$$\begin{aligned}
 (1) \quad Y_t &= \rho Y_{t-1} + e_t & t = 1, 2, \dots \\
 (2) \quad Y_t &= \mu + \rho Y_{t-1} + e_t & t = 1, 2, \dots \\
 (3) \quad Y_t &= \mu + \beta t + \rho Y_{t-1} + e_t & t = 1, 2, \dots
 \end{aligned}$$

donde  $Y_0 = 0$  y  $e_t$  es una secuencia de variables aleatorias independientes y normalmente distribuidas con media cero y varianza  $\sigma^2$  ( $\text{NID}(0, \sigma^2)$  en lo sucesivo). El factor diferenciador en los modelos anteriores es la presencia o ausencia de elementos deterministas. En los tres casos, el test que proponen permite contrastar la hipótesis de que  $\rho = 1$ , o lo que es igual, la hipótesis de existencia de raíz unitaria en la parte autorregresiva del modelo en cuestión. El contraste se basa en la estimación por mínimos cuadrados ordinarios (MCO) del parámetro  $\rho$  en (1), (2) o (3). Esta estimación convenientemente normalizada y su estadístico  $t$  asociado para  $\rho = 1$ , son los estadísticos que Dickey y Fuller (1979) proponen para llevar a cabo el contraste. Estos dos estadísticos no tienen una distribución conocida bajo la hipótesis nula ( $\rho = 1$ ), por lo que Dickey y Fuller dedujeron las distribuciones empíricas de los mismos, mediante métodos de simulación.

La independencia y homocedasticidad del término de error son supuestos muy restrictivos y no muy acordes con la realidad, pues, en la práctica, muchas series económicas presentan una estructura algo más compleja. En este sentido, Dickey y Fuller (1981) desarrollaron el contraste de raíz unitaria asumiendo que el proceso generador de datos era un proceso AR( $p$ ) (es lo que se conoce como Dickey-Fuller Aumentado o DFA).

Posteriormente, Said y Dickey (1984) extendieron el DFA al caso de un proceso ARMA( $p, q$ ) con  $p$  y  $q$  desconocidos, que, a lo sumo, contiene una raíz unitaria en su parte autorregresiva. También, para un ARMA( $p, q$ ), Solo (1984) desarrolla un procedimiento de contraste basado en el test de los multiplicadores de Lagrange. En esta misma línea, Phillips (1987) propone un procedimiento alternativo al DFA, consistente en una corrección no paramétrica de los estadísticos de Dickey y Fuller, que permite llevar a cabo el contraste de raíz unitaria en (1) teniendo en cuenta un cierto grado de dependencia y heterocedasticidad en el término de error. Phillips y Perron (1988) extendieron este procedimiento a los procesos (2) y (3). No obstante, todos estos contrastes presentan considerables distorsiones en el tamaño debido a la presencia de términos de medias móviles en el modelo a estimar (ver Schwert (1989)). Como alternativa, Hall (1989, 1992) y Pantula y Hall (1991) proponen estimar el modelo correspondiente, por variables instrumentales.

Otra crítica que se suele hacer a los tests de raíz unitaria es la falta de consistencia entre la hipótesis nula y la alternativa de dichos contrastes, en relación al tipo de tendencia determinista que presenta el proceso bajo cada una de las hipótesis. Los parámetros de los diferentes modelos que se consideran como posibles mecanismos generadores de la serie objeto de análisis, tienen un significado diferente según sea cierta o no la hipótesis

nula. Para alcanzar esta consistencia entre las hipótesis, autores como Dickey (1984) y Barghava (1986), entre otros, utilizan una parametrización diferente a (1), (2) y (3) para representar el proceso generador de datos:

$$(4) \quad \begin{aligned} Y_t &= \mu^* + \beta^* t + u_t \\ u_t &= \rho u_{t-1} + e_t \end{aligned} \quad t = 1, 2, \dots$$

donde  $e_t \sim NID(0, \sigma^2)$  y  $u_0$  es una variable aleatoria independiente de  $e_t$ . Esta parametrización permitió a Bhargava (1986) desarrollar los contrastes de raíz unitaria invariantes más potentes frente a determinadas alternativas, tomando como base al trabajo de Sargan y Barghava (1983). Asimismo, Schmidt y Phillips (1992) proponen contrastar la hipótesis nula de raíz unitaria aplicando en (4) el test de los multiplicadores de Lagrange, procedimiento que Ahn (1993) generalizó para el caso en que  $u_t$  es un proceso autorregresivo de medias móviles. Hwang y Schmidt (1996) desarrollan un test de raíz unitaria estimando (4) por mínimos cuadrados generalizados.

Otro de los inconvenientes que presentan algunos de estos contrastes, es que se basan en estadísticos cuyas distribuciones bajo la hipótesis nula (raíz unitaria) dependen de los valores que tomen otros parámetros (molestanos) que intervienen en el modelo, es decir, son tests no similares<sup>1</sup>. Esta no similaridad se refleja en la baja potencia que presentan en algunas de las alternativas, provocando graves errores en la toma de decisión sobre la presencia de raíz unitaria en la serie. Evidencia de este hecho es el trabajo de Nelson y Plosser (1982) en el que estudiaban las raíces unitarias presentes en diversas series económicas de los Estados Unidos y que posteriormente Choi (1990) revisó, encontrando que se habían detectado erróneamente raíces unitarias en la mayoría de estas series. En el campo de la similaridad encontramos diversos trabajos. Por un lado, Evans y Savin (1984) y Nankervis y Savin (1985), tras constatar la no similaridad de los estadísticos en los que se basan los tests de Dickey-Fuller, desarrollan varios tests no similares para contrastar la hipótesis de raíz unitaria en un modelo autorregresivo de primer orden con constante. Por otro lado, Kiviet y Phillips (1990, 1992) desarrollan un procedimiento para contrastar de forma similar y exacta si el coeficiente de la variable dependiente retardada en un modelo autorregresivo de primer orden que puede contener variables exógenas, es igual a la unidad (se le conoce con el nombre de test KPh). Concretamente, el test similar de la hipótesis de raíz unitaria alrededor de una función del tiempo determinista de grado  $d$ , se consigue introduciendo en el proceso generador de datos el regresor redundante adicional  $t^{d+1}$ . Este regresor hace que la estimación del coeficiente de la variable dependiente retardada y su estadístico  $t$  asociado sean invariantes con respecto a los parámetros molestanos que intervienen en el proceso generador. Por ejemplo, si en el modelo (3) consideramos a la variable  $t$  como un regresor adicional redundante, el test basado en la estimación MCO de (3) es un contraste similar de

---

<sup>1</sup> Ver evidencia teórica en Nankervis y Savin (1985, 1987).

la hipótesis de raíz unitaria alrededor de una posible constante. En realidad, el test KPh es una generalización de los contrastes de Dickey-Fuller.

Por otro lado, algunos autores como Fukushige et al. (1990) y Ogaki y Park (1990) señalan que, en muchas aplicaciones (especialmente en el contexto de la cointegración), es más atractiva una hipótesis nula de estacionariedad frente a una alternativa de presencia de raíces unitarias. En este sentido, autores como Kwiatkowski et al. (1992), Khan y Ogaki (1992), Bierens y Guo (1993) y Leybourne y McCabe (1994) han desarrollado diferentes contrastes de hipótesis nula estacionaria frente a la alternativa de raíz unitaria.

Finalmente, entre las últimas aportaciones para contrastar raíces unitarias destacan, por un lado, el trabajo de Ferretti y Romo (1996) en el que proponen contrastar la existencia de raíz unitaria aplicando la técnica bootstrap; el trabajo de Shin y So (1997), quienes desarrollan tests semiparamétricos de raíz unitaria basados en estimadores simétricos; y, por último, el de Hasan y Koenker (1997), que proponen un contraste de raíz unitaria basado en el test de rangos.

En definitiva, el objetivo de todos estos contrastes es establecer una metodología que permita determinar si una serie temporal tiene tendencia estocástica (raíz unitaria) o no, con la mayor potencia posible e intentando diferenciarla de la tendencia determinista. El acierto en dicha decisión es fundamental en la determinación del grado de integración de las series, siendo ésta básica tanto para la especificación y estimación del modelo, como para el análisis de cointegración que tanto auge está tomando últimamente.

En nuestro trabajo nos centramos, principalmente, en el carácter no similar de los tests de raíz unitaria, pues, como expondremos, es una de las principales causas de la baja potencia de estos contrastes. Consideramos, por tanto, del máximo interés realizar un estudio que desarrolle toda esta problemática.

Así, el presente trabajo se ha desarrollado con el objetivo principal de:

- Analizar desde la perspectiva de la no similaridad las distribuciones de los distintos estadísticos que habitualmente se emplean para contrastar la presencia de raíz unitaria.

Planteamos para ello como objetivos secundarios:

- Definir zonas de rechazo y aceptación de las hipótesis nulas para cada estadístico considerando las distintas distribuciones del mismo.
- Estudiar las distintas situaciones con las que nos podemos encontrar con el fin de deducir una pauta de comportamiento que minimice el riesgo de error.

- Estudiar la incidencia de los estadísticos en cada una de las zonas definidas mediante un experimento Monte Carlo.

Así, en la sección 2.1 analizamos las distribuciones de los estadísticos  $t$  asociados a los estimadores MCO de los parámetros que intervienen en el modelo con constante y tendencia determinista lineal. En la sección 2.2 se estudia el modelo autorregresivo con constante. En la sección 2.3 se considera un proceso autorregresivo puro de orden 1. En la sección 3, se presenta un ejemplo en el que se pone de manifiesto que la no similaridad de los contrastes tradicionales de raíz unitaria nos puede llevar a tomar una decisión equivocada. Teniendo en cuenta dicha no similaridad, se propone una estrategia de contraste secuencial para resolver situaciones de indecisión acerca del rechazo o no de la existencia de raíz unitaria. Finalmente, se presentan los resultados de un experimento Monte Carlo que recoge la probabilidad de que el valor calculado de cada estadístico caiga en las diferentes zonas establecidas en la sección 2.

## 2. ANÁLISIS DE LAS DISTRIBUCIONES DE LOS ESTADÍSTICOS $t$ Y DE SUS CONTRASTES ASOCIADOS

En esta sección se analizan las distribuciones de los estadísticos  $t$  asociados a los estimadores MCO de los parámetros que intervienen en los modelos (1), (2) y (3). Es decir, se toman los tres procesos generadores de datos que consideraron Dickey y Fuller para desarrollar sus tests de raíz unitaria, asumiendo en cada uno que el término de error es independiente e idénticamente distribuido con media cero y varianza constante, y que la condición inicial  $Y_0$  es constante e igual a cero.

### 2.1. Planteamiento y estimación del modelo completo

En primer lugar analizamos el proceso más general (3), es decir, aquel en el que aparece, además del término autorregresivo de orden 1, un término constante y una tendencia determinista lineal. Dicho proceso se representa mediante el modelo

$$(5) \quad Y_t = \mu + \beta t + \rho Y_{t-1} + e_t \quad t = 1, 2, \dots, T$$

donde  $\mu$  es una constante,  $t$  la tendencia determinista lineal, y para el que se asume  $Y_0 = 0$  y  $e_t \sim NID(0, \sigma^2)$ .

De la estimación MCO del modelo (5) se obtienen los siguientes estimadores y estadísticos:

$\hat{\rho}_T$ : Estimador MCO del parámetro  $\rho$

$\hat{\beta}_T$ : Estimador MCO del parámetro  $\beta$

$\hat{\mu}_\tau$ : Estimador MCO del parámetro  $\mu$

$t_{\rho_\tau}$ : Estadístico  $t$  asociado a la estimación MCO de  $\rho$

$t_{\beta_\tau}$ : Estadístico  $t$  asociado a la estimación MCO de  $\beta$

$t_{\mu_\tau}$ : Estadístico  $t$  asociado a la estimación MCO de  $\mu$

Cuando  $\beta = 0$  y  $\rho = 1$ , al estadístico  $t_{\mu_\tau}$  para  $\mu = 0$  se le suele dar la notación específica  $\hat{\tau}_{\alpha_\tau}$ . Asimismo, cuando  $\mu = 0$  y  $\rho = 1$ , al estadístico  $t_{\beta_\tau}$  para  $\beta = 0$  se le denota  $\hat{\tau}_{\beta_\tau}$ . Finalmente, cuando  $\beta = 0$ , siendo  $\mu$  cualquier valor real, al estadístico  $t_{\rho_\tau}$  para  $\rho = 1$  se le denota  $\hat{\tau}_\tau$ .

## A) CONTRASTE SOBRE $\rho$

### A.1) Distribución de $t_{\rho_\tau}$

En principio, se puede utilizar el estadístico  $t_{\rho_\tau}$  para llevar a cabo el contraste unilateral de presencia de raíz unitaria

$$\begin{aligned} H_0 : \rho &= 1 \\ H_1 : \rho &< 1 \end{aligned}$$

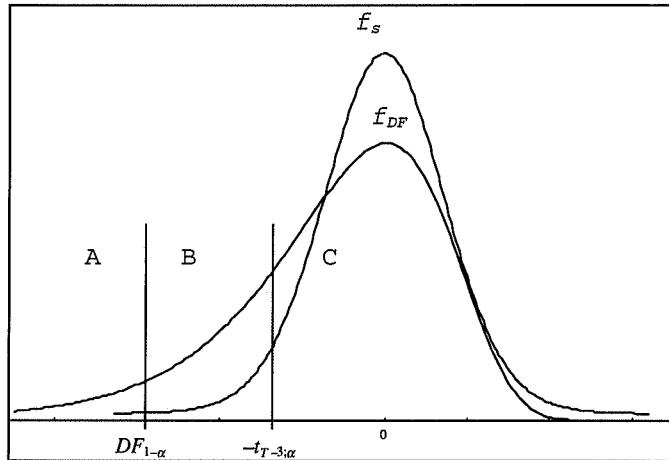
Dickey y Fuller (1979) obtuvieron la representación de la distribución límite de  $\hat{\tau}_\tau$  bajo  $Y_0 = 0$  y  $e_t \sim \text{NID}(0, \sigma^2)$ , es decir, la distribución límite de  $t_{\rho_\tau}$  para  $\rho = 1$  asumiendo que el coeficiente  $\beta$  de la variable  $t$  es cero. Esta distribución es independiente del valor de  $\mu$  (ver Dickey (1976)), y se mantiene en el caso de que  $e_t \sim \text{iid}(0, \sigma^2)$ . En cualquier caso, se trata de una distribución no conocida, y Dickey (1976) utilizó métodos Monte Carlo para estimar los cuantiles de la distribución de  $\hat{\tau}_\tau$ , tanto para tamaños de muestra finitos como para la distribución límite. Algunos de estos cuantiles empíricos aparecen tabulados para  $T = 25, 50, 100, 250$  y  $500$  en Fuller (1996, p. 642)<sup>2</sup>.

Por otro lado, Dickey y Fuller (1979) demuestran que si  $\beta \neq 0$  la distribución límite de  $t_{\rho_\tau}$  para  $\rho = 1$  es normal estándar.

La influencia de  $\beta$  en la distribución de  $t_{\rho_\tau}$  para  $\rho = 1$  y tamaños de muestra finitos, fue estudiada por Nakervis y Savin (1987). Estos autores, asumiendo  $Y_0 = 0$  sin pérdida de generalidad, y  $e_t \sim N(0, 1)$ , prueban que, para cada  $T$  fijo, la distribución de  $t_{\rho_\tau}$  (para  $\rho = 1$ ) converge a una  $t$  de Student con  $T - 3$  grados de libertad cuando  $|\beta| \rightarrow \infty$  (Nankervis y Savin, 1987, p. 396).

---

<sup>2</sup>Guilkey y Schmidt (1989) proporcionan más cuantiles para más tamaños de muestra.



**Figura 1.** Densidad de  $t_{\rho_T}$  bajo  $\rho = 1$ , según el valor de  $\beta$

Fuente: Elaboración propia.

La distribución empírica de  $\hat{\tau}_T$  que obtuvieron Dickey y Fuller presenta asimetría positiva. Además, para cada  $T$  y un mismo nivel de significación, los valores críticos son menores que los correspondientes a una  $t$  de Student con  $T - 3$  grados de libertad. Por tanto, el resultado de Nankervis y Savin (1987) implica que, a medida que aumenta el valor de  $\beta$ , la distribución de  $\hat{\tau}_T$  se va desplazando a la derecha y haciéndose simétrica, hasta convertirse en una  $t_{T-3}$ .

Obviamente, en la práctica  $\beta$  no será infinito, por lo que la  $t_{T-3}$  es una aproximación a la distribución de  $t_{\rho_T}$  para  $\rho = 1$ ,  $\beta \neq 0$  y  $T$  finito que, según un estudio Monte Carlo de Nankervis y Savin (1987), es buena para valores de  $\beta$  pequeños<sup>3</sup>.

### A.2) Contraste de la hipótesis $\rho = 1$

Según lo anterior, el contraste basado en el estadístico  $t_{\rho_T}$  constituye un test no similar, ya que su distribución bajo la hipótesis nula de existencia de raíz unitaria depende del valor que tome el parámetro  $\beta$ , siendo independiente del valor que tome  $\mu$ . Por tanto, para llevar a cabo en (5) el contraste unilateral planteado anteriormente, hay que tener en cuenta las dos distribuciones extremas del estadístico  $t_{\rho_T}$ , pudiéndose establecer, como veremos seguidamente, zonas de aceptación, duda y rechazo según el nivel de

---

<sup>3</sup>Las aproximaciones son satisfactorias para  $\rho = 1$  y  $T = 100$  cuando  $\beta \geq 0.1$ .

significación  $\alpha$  fijado. En la figura 1,  $f_{DF}$  representa la densidad empírica de  $\hat{\tau}_\tau$  (siendo  $\rho = 1$  y  $\beta = 0$ ) para el valor de  $T$  considerado, mientras que  $f_S$  es la densidad de la  $t_{T-3}$ , también bajo  $\rho = 1$ , tomada como aproximación al caso en el que  $\beta \neq 0$ .

El valor  $DF_{1-\alpha}$  representa el cuantil de orden  $1 - \alpha$  de la distribución empírica de  $\hat{\tau}_\tau$ . Así, el intervalo  $(DF_{1-\alpha}; +\infty)$  constituye la región de aceptación del contraste unilateral

$$H_0 : \rho = 1 \quad (\text{asumiendo } \beta = 0)$$

$$H_1 : \rho < 1$$

basado en  $\hat{\tau}_\tau$ .

Por otro lado, el valor  $t_{T-3;\alpha}$  denota el cuantil de orden  $\alpha$  de la  $t$  de Student con  $T - 3$  grados de libertad. Por tanto, el intervalo  $(-t_{T-3;\alpha}; +\infty)$  determinaría la región de aceptación del contraste unilateral

$$H_0 : \rho = 1 \quad (\text{asumiendo } \beta \neq 0)$$

$$H_1 : \rho < 1$$

que realizaríamos en el caso de que se conociera que  $\beta \neq 0$ .

En la práctica, lo habitual es que no se conozca el valor del parámetro  $\beta$ , por lo que desarrollamos el contraste considerando las dos distribuciones simultáneamente. Así, la región de aceptación de tamaño  $\alpha$  del contraste unilateral sería  $(-t_{T-3;\alpha}; +\infty)$ , que corresponde a la **zona C** de la figura 1. De esta forma, si el valor calculado del estadístico  $t_{\rho_\tau} (\hat{t}_{\rho_\tau}$  en lo sucesivo) cae en la zona C no se rechaza la hipótesis nula ( $\rho = 1$ ). Además, al ser la zona C la intersección de las regiones de aceptación  $(DF_{1-\alpha/2}; +\infty)$  y  $(-t_{T-3;\alpha}; +\infty)$ , el valor  $\hat{t}_{\rho_\tau}$  llevaría a aceptar la hipótesis nula del contraste basado solamente en  $\hat{\tau}_\tau$ , así como la del contraste basado únicamente en la  $t_{T-3}$ . Por tanto, cuando  $\hat{t}_{\rho_\tau}$  caiga en la zona C, se puede concluir que  $\rho = 1$ , pudiendo  $\beta$ , y por supuesto  $\mu$ , tomar cualquier valor.

Cuando el valor calculado del estadístico  $t_{\rho_\tau}$  cae en la **zona A**, constituida por el intervalo  $(-\infty; DF_{1-\alpha})$ , en principio, la conclusión debe ser que  $\rho < 1$ . No obstante, no debemos caer en la tentación de añadir a la decisión sobre  $\rho$  que  $\mu = 0$  por el hecho de haber aplicado el contraste basado en una distribución (la de  $\hat{\tau}_\tau$ ) obtenida bajo el supuesto de  $\mu = 0$ . Si tenemos en cuenta que  $DF_{1-\alpha} < -t_{T-3;\alpha}$ , entonces, si se ha rechazado la hipótesis nula ( $\rho = 1$ ) para la distribución empírica de  $\hat{\tau}_\tau$  seguro que también se rechaza para la  $t_{T-3}$ . Esto significa que cuando  $\hat{t}_{\rho_\tau}$  cae en la zona A, cabe esperar que se deba bien a que  $\rho < 1$  y  $\beta = 0$ , o a que  $\rho < 1$  y  $\beta \neq 0$ . Por tanto, para un valor de  $\hat{t}_{\rho_\tau}$  en la zona A, la conclusión debe ser que  $\rho < 1$ , pudiendo  $\beta$  tomar cualquier valor<sup>4</sup>.

---

<sup>4</sup> Esta es la razón por la que en el planteamiento de las hipótesis del contraste basado en  $t_{\rho_\tau}$  no se dice nada acerca de  $\beta$  en las alternativas, pues, bajo estas hipótesis, el parámetro  $\beta$  puede tomar cualquier valor.

En el caso de que  $\hat{t}_{\rho_\tau}$  caiga en el intervalo  $(DF_{1-\alpha}; -t_{T-3;\alpha})$  o **zona B** de la figura, se pueden plantear dos situaciones diferentes:

- a) El valor calculado de  $t_{\rho_\tau}$  pertenece al intervalo  $(DF_{1-\alpha}; +\infty)$ , por lo que no se rechaza la hipótesis nula del contraste basado en  $\hat{\tau}_\tau$ . De esta forma, la conclusión debe ser  $\rho = 1$  asumiendo que  $\beta = 0$ .
- b) El valor  $\hat{t}_{\rho_\tau}$  comparado con el valor crítico  $-t_{T-3;\alpha}$  de la  $t_{T-3}$ , es tal que  $\hat{t}_{\rho_\tau} < -t_{T-3;\alpha}$ , rechazándose la hipótesis nula del contraste basado en la  $t_{T-3}$ . Por tanto, lo más probable es que  $\rho < 1$ , siendo  $\beta \neq 0$ , de donde se excluye el caso  $\rho < 1$  y  $\beta = 0$ , correspondiente al error de tipo II del contraste basado en  $\hat{\tau}_\tau$ .

Por tanto, si el valor  $\hat{t}_{\rho_\tau}$  está en la zona B lo más probable es que sea porque o bien  $\rho = 1$  y  $\beta = 0$ , o bien  $\rho < 1$  y  $\beta \neq 0$ . Se observa, pues, que el utilizar unos valores críticos u otros para el contraste de raíz unitaria en (5) nos lleva a tomar decisiones diferentes sobre  $\rho$ , lo que supone la posibilidad de cometer un error adicional al inherente a todo problema inferencial. Por todo esto llamaremos a la zona B, zona de duda.

En este caso, la duda podría despejarse si, de alguna forma, se pudiese determinar si  $\beta = 0$  o  $\beta \neq 0$ . Por ello, en la sección siguiente analizamos las características de la distribución del estadístico  $t$  asociado a la estimación MCO de  $\beta$  en (5).

Los valores críticos de  $t_{\rho_\tau}$  que permiten la definición de las zonas de aceptación, rechazo y duda descritas para este estadístico se presentan en la tabla A del anexo, para distintos tamaños muestrales. Por otro lado, la tabla 1 resume las consecuencias que se derivan según la zona en la que caiga el valor calculado de  $t_{\rho_\tau}$ .

**Tabla 1.** Consecuencia según la zona en la que cae el valor  $\hat{t}_{\rho_\tau}$

Zona	Extremos	Consecuencia
A	$(-\infty; DF_{1-\alpha})$	$\rho < 1, \beta$ cualquiera
B	$(DF_{1-\alpha}; -t_{T-3;\alpha})$	$\rho = 1 \text{ y } \beta = 0$ ó $\rho < 1 \text{ y } \beta \neq 0$
C	$(-t_{T-3;\alpha}; +\infty)$	$\rho = 1, \beta$ cualquiera

**Fuente:** Elaboración propia.

## B) CONTRASTE SOBRE $\beta$

### B.1) Distribución de $t_{\beta_\tau}$

El estadístico  $t$  asociado a la estimación MCO de  $\beta$  en (5), que hemos denotado  $t_{\beta_\tau}$ , puede utilizarse para contrastar la significación de  $\beta$ , es decir, para el contraste bilateral

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

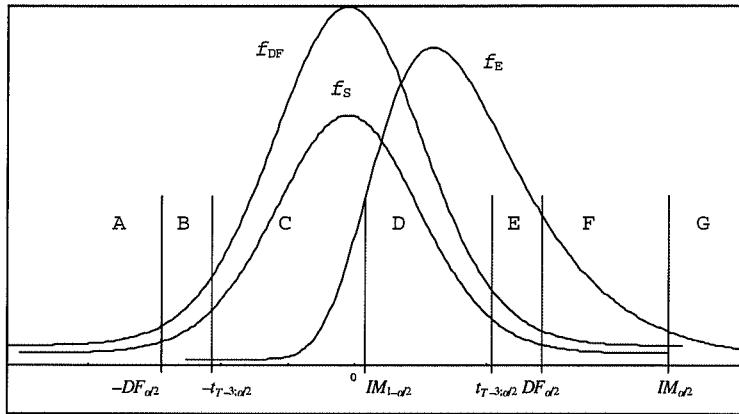
La distribución de este estadístico bajo  $H_0$ , depende del valor de  $\rho$  y de  $\mu$ . Así, cuando  $\rho = 1$  y  $\mu = 0$ , la distribución límite de  $t_{\beta_\tau}$  para  $\beta = 0$ , o lo que es igual, la distribución límite de  $\hat{\tau}_{\beta_\tau}$ , fue caracterizada por Dickey y Fuller (1981). Se trata de una distribución no conocida, y estos mismos autores proporcionan algunos valores críticos, obtenidos por simulación, para el contraste de significación de  $\beta$ , tanto para tamaños de muestra finitos como para la distribución límite (Dickey y Fuller 1981, Tabla III, p. 1062).

Hasza (1977) ha desarrollado la teoría asintótica para el caso en el que las variables exógenas en un modelo autorregresivo son polinomios de tiempo. Según esta teoría, la distribución límite de  $t_{\beta_\tau}$  para cualquier  $\beta$  (a nosotros nos interesa  $\beta = 0$ ) cuando  $|\rho| \neq 1$ , es una normal estándar, con independencia del valor que tome  $\mu$ . Por tanto, para tamaños de muestra finitos se puede emplear como aproximación la  $t$  de Student con  $T - 3$  grados de libertad.

Finalmente, si  $\beta = 0$  y  $\rho = 1$ , el teorema 3 de Nankervis y Savin (1987, p. 398) establece que cuando  $|\mu| \rightarrow \infty$ , la distribución de  $t_{\beta_\tau}$  para cada  $T$  fijo, tiende a la imagen en el espejo de la distribución de  $\hat{\tau}_\tau$  para ese valor de  $T$ . Esto significa que los valores críticos que se deben utilizar son los que Dickey (1976) tabuló para  $\hat{\tau}_\tau$ , pero cambiados de signo. Por ello, denotamos  $-\hat{\tau}_\tau$  a la imagen en el espejo de  $\hat{\tau}_\tau$ , cuya distribución empírica utilizaremos como aproximación a la distribución de  $t_{\beta_\tau}$  para  $\beta = 0$ ,  $\rho = 1$  y  $\mu \neq 0$ . En un estudio Monte Carlo realizado por Nankervis y Savin (1987, p. 504- 506), se tiene que la aproximación es buena para valores de  $\mu$  no muy grandes (por ejemplo, para  $T = 25$ , la aproximación es buena para  $\mu \geq 1$ ).

### B.2) Contraste de la hipótesis $\beta = 0$

Al igual que ocurría con el estadístico  $t$  asociado a la estimación MCO de  $\rho$ , el que la distribución de  $t_{\beta_\tau}$  para  $\beta = 0$  dependa del valor de los parámetros  $\mu$  y  $\rho$ , significa que el contraste basado en dicho estadístico es un test no similar de la hipótesis  $H_0 : \beta = 0$  en (5). Por tanto, consideraremos las tres distribuciones descritas para realizar el contraste bilateral sobre  $\beta$  planteado más arriba.



**Figura 2.** Densidad de  $t_{\beta_\tau}$  bajo  $\beta = 0$ , según los valores de  $\rho$  y  $\mu$   
**Fuente:** Elaboración propia.

La figura 2 muestra la densidad del estadístico  $t_{\beta_\tau}$  bajo la hipótesis nula ( $\beta = 0$ ) y un tamaño muestral fijo, según los valores de  $\rho$  y  $\mu$ . Así,  $f_{DF}$  representa la densidad empírica<sup>5</sup> de  $\hat{\tau}_{\beta_\tau}$  para el valor de  $T$  considerado;  $f_S$  es la densidad de la  $t_{T-3}$ ; y  $f_E$  es la densidad de la imagen en el espejo de  $\hat{\tau}_\tau$ , también para ese mismo valor de  $T$ .

Asimismo, se puede observar la simetría con respecto al origen de la distribución empírica de  $\hat{\tau}_{\beta_\tau}$ , siendo  $DF_{\alpha/2}$  el cuantil de orden  $\alpha/2$ . De esta forma, el intervalo  $(-DF_{\alpha/2}; DF_{\alpha/2})$  constituye la región de aceptación del contraste

$$\begin{aligned} H_0 : \beta &= 0 && (\text{asumiendo } \rho = 1 \text{ y } \mu = 0) \\ H_1 : \beta &\neq 0 \end{aligned}$$

En cambio, si se utilizan los valores críticos de la  $t$  de Student con  $T - 3$  grados de libertad, se obtiene la región de aceptación  $(-t_{T-3;\alpha/2}; t_{T-3;\alpha/2})$  correspondiente al contraste

$$\begin{aligned} H_0 : \beta &= 0 && (\text{asumiendo } |\rho| \neq 1) \\ H_1 : \beta &\neq 0 \end{aligned}$$

Finalmente, los valores  $IM_{1-\alpha/2}$  y  $IM_{\alpha/2}$ , representan los cuantiles de órdenes  $1 - \alpha/2$  y  $\alpha/2$ , respectivamente, de la distribución empírica de  $-\hat{\tau}_\tau$ . Así, el intervalo

<sup>5</sup> Aunque en la figura 2 no aparece reflejado, en realidad,  $f_{DF}$  es una distribución bimodal (ver Dickey y Fuller (1981), p. 1065). No obstante, este hecho no afecta al desarrollo que se realiza.

$(IM_{1-\alpha/2}; IM_{\alpha/2})$  constituye la región de aceptación del contraste

$$H_0 : \beta = 0 \quad (\text{asumiendo } \rho = 1 \text{ y } \mu \neq 0)$$

$$H_1 : \beta \neq 0$$

Puesto que lo habitual es no conocer nada acerca de  $\rho$  y  $\mu$ , la utilización simultánea de las tres distribuciones mencionadas, cuyos valores críticos determinan una serie de zonas que quedan reflejadas en la figura 2, permite sacar algunas conclusiones sobre los parámetros  $\mu$ ,  $\beta$  y  $\rho$ , dependiendo de la zona en la que caiga el valor calculado del estadístico  $t_{\beta_\tau}$  ( $\hat{t}_{\beta_\tau}$  en lo sucesivo).

Así, el intervalo  $(IM_{\alpha/2}; t_{T-3;\alpha/2})$ , **zona D** en la figura 2, es la región de aceptación de tamaño  $\alpha$  del contraste bilateral sobre  $\beta$ , ya que es la intersección de las regiones de aceptación  $(-DF_{\alpha/2}; DF_{\alpha/2})$ ,  $(-t_{T-3;\alpha/2}; t_{T-3;\alpha/2})$  e  $(IM_{1-\alpha/2}; IM_{\alpha/2})$  de los contrastes basados en  $\hat{\tau}_{\beta_\tau}$ ,  $t_{T-3}$  y  $-\hat{\tau}_\tau$ , respectivamente. Esto significa que un valor de  $\hat{t}_{\beta_\tau}$  en la zona D no permite rechazar las hipótesis nulas de los contrastes anteriores, y se concluye que  $\beta = 0$ , pudiendo  $\rho$  y  $\mu$  tomar cualquier valor.

Los intervalos  $(-\infty; -DF_{\alpha/2})$  e  $(IM_{\alpha/2}; +\infty)$ , **zonas A y G** respectivamente, constituyen la región de rechazo del contraste basado en las tres distribuciones, ya que en estas zonas el valor calculado de  $t_{\beta_\tau}$  no pertenece a las regiones de aceptación de los contrastes basados en  $\hat{\tau}_{\beta_\tau}$ ,  $t_{T-3}$  y  $-\hat{\tau}_\tau$ . En este caso la conclusión debe ser que  $\beta \neq 0$ , pudiendo  $\rho$  y  $\mu$  tomar cualquier valor.

Si el valor  $\hat{t}_{\beta_\tau}$  cae en la **zona B**, es decir, en el intervalo  $(-DF_{\alpha/2}; -t_{T-3;\alpha/2})$ , se pueden hacer las siguientes consideraciones:

- a) El valor calculado  $\hat{t}_{\beta_\tau}$  pertenece al intervalo  $(-DF_{\alpha/2}; DF_{\alpha/2})$ , y, por tanto, no se rechaza la hipótesis nula del contraste basado en  $\hat{\tau}_{\beta_\tau}$ . De esta forma, la conclusión sería que  $\beta = 0$ , asumiendo que  $\rho = 1$  y  $\mu = 0$ .
- b) En cambio, al ser  $\hat{t}_{\beta_\tau} < -t_{T-3;\alpha/2}$ , se rechazarían las hipótesis nulas de los contrastes basados en la  $t_{T-3}$  y en  $-\hat{\tau}_\tau$ . Por tanto, se concluye que  $\beta \neq 0$ , pudiendo  $\rho$  y  $\mu$  tomar cualquier valor, si bien debe excluirse el caso  $\beta \neq 0$ ,  $\rho = 1$  y  $\mu = 0$  achacable al error de tipo II del contraste basado en  $\hat{\tau}_{\beta_\tau}$ .

Por tanto, si el valor  $\hat{t}_{\beta_\tau}$  está en la zona B lo más probable es que sea porque o bien  $\beta = 0$ ,  $\rho = 1$  y  $\mu = 0$ , o bien  $\beta \neq 0$  y  $\rho$  y  $\mu$  cualesquiera (excluyendo el caso en el que de forma simultánea  $\rho = 1$  y  $\mu = 0$ ).

En el caso de que  $\hat{t}_{\beta_\tau}$  caiga en el intervalo  $(-t_{T-3;\alpha/2}; IM_{1-\alpha/2})$  o **zona C** de la figura, se plantean las siguientes situaciones:

- a) El valor calculado de  $t_{\beta_\tau}$  está dentro de las regiones de aceptación de los contrastes basados en  $\hat{\tau}_{\beta_\tau}$  y en la  $t$  de Student, lo que no nos permite rechazar la hipótesis de nula de cada uno de estos contrastes. Concretamente, para el contraste basado en  $\hat{\tau}_{\beta_\tau}$ , la conclusión sería  $\beta = 0$ , asumiendo  $\rho = 1$  y  $\mu = 0$ ; mientras que para el contraste basado en la  $t_{T-3}$ , la decisión sería  $\beta = 0$ , asumiendo que  $|\rho| \neq 1$  y pudiendo  $\mu$  tomar cualquier valor. Nótese que el único caso que queda fuera es  $\beta = 0$ ,  $\rho = 1$  y  $\mu \neq 0$ , correspondiente a la aceptación de la hipótesis nula del contraste basado en  $-\hat{\tau}_\tau$ , circunstancia que, como veremos a continuación, no se produce en esta zona.
- b) Como  $\hat{t}_{\beta_\tau} < IM_{1-\alpha/2}$ , se rechaza la hipótesis nula correspondiente al contraste basado en  $-\hat{\tau}_\tau$ , lo que significa que  $\beta \neq 0$ . Además, al estar en la zona C, es de esperar que  $\rho = 1$  y  $\mu \neq 0$ , pues, los demás casos son atribuibles al error de tipo II del contraste basado en  $\hat{\tau}_{\beta_\tau}$  o en la  $t_{T-3}$ .

Por tanto, si el valor  $\hat{t}_{\beta_\tau}$  cae en la zona C lo más probable es que sea porque es  $\beta = 0$  con  $\rho$  y  $\mu$  cualesquiera (excluyendo el caso en el que de forma simultánea  $\rho = 1$  y  $\mu \neq 0$ ) o porque  $\beta \neq 0$ ,  $\rho = 1$  y  $\mu \neq 0$ .

Por otro lado, cuando  $\hat{t}_{\beta_\tau}$  se encuentra en el intervalo  $(t_{T-3;\alpha/2}; DF_{\alpha/2})$  o **zona E**, se producen las siguientes circunstancias:

- a) Se aceptan las hipótesis nulas de los contrastes basados en  $\hat{\tau}_{\beta_\tau}$  y  $-\hat{\tau}_\tau$ , pues, el valor observado de  $t_{\beta_\tau}$  cae dentro de las respectivas regiones de aceptación. De esta forma, en relación al primer contraste, la conclusión es  $\beta = 0$ , asumiendo  $\rho = 1$  y  $\mu = 0$ ; mientras que la conclusión respecto al segundo es  $\beta = 0$ , asumiendo  $\rho = 1$  y  $\mu \neq 0$ . Esto permite concluir que  $\beta = 0$ , asumiendo  $\rho = 1$  y pudiendo  $\mu$  tomar cualquier valor.
- b) El hecho de que  $\hat{t}_{\beta_\tau} > t_{T-3;\alpha/2}$  implica el rechazo de la hipótesis nula del contraste basado en la  $t_{T-3}$ . Por tanto, la conclusión sería  $\beta \neq 0$ , y al estar en la zona E lo más probable es que  $|\rho| \neq 1$  con independencia del valor de  $\mu$ , pues, los casos en los que  $\beta \neq 0$  y  $\rho = 1$  forman parte del error de tipo II del contraste basado en  $\hat{\tau}_{\beta_\tau}$  o en  $-\hat{\tau}_\tau$ .

Así pues, cuando  $\hat{t}_{\beta_\tau}$  cae en la zona E lo más probable es que sea porque  $\beta = 0$  y  $\rho = 1$ , o porque  $\beta \neq 0$  y  $|\rho| \neq 1$ , pudiendo  $\mu$ , en ambos casos, tomar cualquier valor.

Finalmente, si el valor calculado de  $t_{\beta_\tau}$  cae en la **zona F** o intervalo  $(DF_{\alpha/2}; IM_{\alpha/2})$  podemos hacer las siguientes consideraciones:

- a) El valor  $\hat{t}_{\beta_\tau}$  cae dentro de la región de aceptación del contraste basado en  $-\hat{\tau}_\tau$ . Esto significa que se acepta que  $\beta = 0$ , asumiendo  $\rho = 1$  y  $\mu \neq 0$ .
- b) En cambio, el que  $\hat{t}_{\beta_\tau} > DF_{\alpha/2}$  indica que se rechazan las hipótesis nulas de los contrastes basados en  $\hat{\tau}_{\beta_\tau}$  y en la  $t_{T-3}$ . Esto nos lleva a concluir que  $\beta \neq 0$ , pudiendo

$\rho$  y  $\mu$  tomar cualquier valor, debiendo excluir el caso  $\beta \neq 0, \rho = 1$  y  $\mu \neq 0$  por formar parte del error de tipo II del contraste basado en  $-\hat{\tau}_\tau$ .

Por tanto, si el valor  $\hat{\tau}_{\beta_\tau}$  cae en la zona F lo más probable es que sea porque o bien  $\beta = 0, \rho = 1$  y  $\mu \neq 0$ , o bien porque  $\beta \neq 0$  con  $\rho$  y  $\mu$  cualesquiera (excluyendo el caso en el que de forma simultánea  $\rho = 1$  y  $\mu \neq 0$ ).

Como se ha podido comprobar en las **zonas B, C, E y F** se llega a conclusiones diferentes acerca del parámetro  $\beta$  según se utilicen unos valores críticos u otros. Por ello, a estas zonas las llamamos zonas de duda, la cual se resolvería si tuviésemos información acerca de  $\rho$  y  $\mu$ .

Los valores críticos de  $t_{\beta_\tau}$  que determinan las zonas de aceptación, rechazo y duda descritas para este estadístico se presentan en la tabla B del anexo para distintos tamaños muestrales. Asimismo, en la tabla 2 se recogen las consecuencias a las que se llegan según la zona en la que se encuentre el valor calculado de  $t_{\beta_\tau}$ .

**Tabla 2.** Consecuencia según la zona en la que cae el valor  $\hat{\tau}_{\beta_\tau}$

Zona	Extremos	Consecuencia
A	$(-\infty; -DF_{\alpha/2})$	$\beta \neq 0, \rho$ y $\mu$ cualesquiera
B	$(-DF_{\alpha/2}; -t_{T-3;\alpha/2})$	$\beta = 0, \rho = 1, \mu = 0$ ó $\beta \neq 0, \rho$ y $\mu$ cualesquiera*
C	$(-t_{T-3;\alpha/2}; IM_{1-\alpha/2})$	$\beta = 0, \rho$ y $\mu$ cualesquiera** ó $\beta \neq 0, \rho = 1, \mu \neq 0$
D	$(IM_{\alpha/2}; t_{T-3;\alpha/2})$	$\beta = 0, \rho$ y $\mu$ cualesquiera
E	$(t_{T-3;\alpha/2}; DF_{\alpha/2})$	$\beta = 0, \rho = 1, \mu$ cualquiera
F	$(DF_{\alpha/2}; IM_{\alpha/2})$	$\beta = 0, \rho = 1, \mu \neq 0$ ó $\beta \neq 0, \rho$ y $\mu$ cualesquiera***
G	$(IM_{\alpha/2}; +\infty)$	$\beta \neq 0,  \rho  \neq 1, \mu$ cualquiera

\* Se excluye el caso en el que de forma simultánea  $\beta \neq 0, \rho = 1, \mu = 0$ .

\*\* Se excluye el caso en el que de forma simultánea  $\beta = 0, \rho = 1, \mu \neq 0$ .

\*\*\* Se excluye el caso en el que de forma simultánea  $\beta \neq 0, \rho = 1, \mu \neq 0$ .

**Fuente:** Elaboración propia.

### C) DISTRIBUCIÓN DEL ESTADÍSTICO $t_{\mu_\tau}$ Y CONTRASTE ASOCIADO

El estadístico  $t$  de la estimación MCO de  $\mu$  en (5), que hemos denotado  $t_{\mu_\tau}$ , puede utilizarse para contrastar la significación de  $\mu$ , siendo las hipótesis del contraste

$$\begin{aligned} H_0 : \mu &= 0 \\ H_1 : \mu &\neq 0 \end{aligned}$$

El contraste basado en este estadístico es un test no similar, pues, su distribución bajo  $H_0$  depende del valor de  $\rho$  y  $\beta$ . Dickey y Fuller (1981) obtuvieron la representación de la distribución límite del estadístico  $t_{\mu_\tau}$  para  $\mu = 0$  cuando  $\rho = 1$  y  $\beta = 0$ , el cual hemos denotado  $\hat{\tau}_{\alpha_\tau}$ . Como en casos anteriores, se trata de una distribución no conocida, y estos autores proporcionan, mediante simulación, algunos valores críticos para el contraste de significación de  $\mu$ , tanto para tamaños de muestra finitos como para la distribución límite (Dickey y Fuller 1981, Tabla II, p. 1062). Se trata de una distribución bimodal y simétrica con los cuantiles para el contraste al 5% más allá del valor 2.

Según la teoría asintótica desarrollada por Hasza (1977), la distribución límite de  $t_{\mu_\tau}$  para  $\mu = 0$  cuando  $|\rho| \neq 1$  es una normal estándar.

### D) DISTRIBUCIONES DE LOS ESTADÍSTICOS $\Phi_2$ Y $\Phi_3$ Y CONTRASTES ASOCIADOS

Se analizan a continuación las distribuciones de los estadísticos  $\Phi_2$  y  $\Phi_3$  que corresponden a contrastes globales sobre el modelo (5):

$\Phi_2$ : Estadístico F para contrastar la hipótesis conjunta:  $H_0 : \rho = 1, \beta = 0, \mu = 0$ .

$\Phi_3$ : Estadístico F para contrastar la hipótesis conjunta:  $H_0 : \rho = 1, \beta = 0$ .

Dickey y Fuller (1981) caracterizaron las distribuciones límite de  $\Phi_2$  y  $\Phi_3$ , no estando la segunda afectada por el valor de  $\mu$ . Las distribuciones de estos estadísticos no son la F de Snedecor conocida y Dickey y Fuller (1981) obtuvieron empíricamente algunos valores críticos para tamaños de muestra finitos y para la distribución límite (Dickey y Fuller, 1981, Tablas V y VI, p. 1063).

Si no se rechaza la hipótesis nula del contraste basado en  $\Phi_2$ , la conclusión final es que  $\rho = 1, \beta = 0$  y  $\mu = 0$ , mientras que rechazar  $H_0$  nos lleva a no poder decidir nada, pues el rechazo puede deberse a que  $\rho \neq 1, \beta \neq 0$  y/o  $\mu \neq 0$ .

De forma análoga, si no podemos rechazar la hipótesis nula del contraste basado en  $\Phi_3$  la conclusión final es que  $\rho = 1$  y  $\beta = 0$ , con independencia del valor de  $\mu$ ; mientras

que rechazar  $H_0$  no nos permite tomar una decisión concreta, pues el rechazo puede deberse a que  $\rho \neq 1$  y/o  $\beta \neq 0$ .

## 2.2. Planteamiento y estimación del modelo solo con término constante

En caso de que se conozca que el modelo no lleva tendencia determinista lineal, podemos analizar la existencia de raíz unitaria partiendo de la estimación MCO del modelo (2), cuya representación es

$$(6) \quad Y_t = \mu + \rho Y_{t-1} + e_t \quad t = 1, 2, \dots, T$$

donde  $\mu$  y  $\rho$  son parámetros desconocidos, y para el que se asume  $Y_0 = 0$  y  $e_t \sim NI(0, 1)$ . De dicha estimación obtenemos los siguientes estimadores y estadísticos:

$\hat{\rho}_\mu$ : Estimador MCO del parámetro  $\rho$ .

$\hat{\mu}$ : Estimador MCO del parámetro  $\mu$ .

$t_{\rho_\mu}$ : Estadístico  $t$  asociado a la estimación MCO de  $\rho$ .

$t_\mu$ : Estadístico  $t$  asociado a la estimación MCO de  $\mu$ .

Cuando  $\mu = 0$  al estadístico  $t_{\rho_\mu}$  para  $\rho = 1$  se le suele dar la notación específica  $\hat{\tau}_\mu$ . Asimismo, cuando  $\rho = 1$ , al estadístico  $t_\mu$  para  $\mu = 0$  se le denota  $\hat{\tau}_{\alpha\mu}$ .

### A) CONTRASTE SOBRE $\rho$

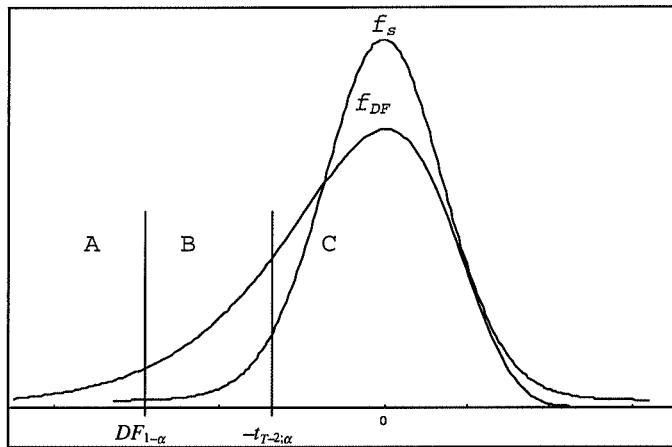
#### A.1) Distribución de $t_{\rho_\mu}$

En principio, para llevar a cabo el contraste unilateral de presencia de raíz unitaria

$$\begin{aligned} H_0 : \rho &= 1 \\ H_1 : \rho &< 1 \end{aligned}$$

se puede utilizar el estadístico  $t_{\rho_\mu}$ . En este caso la distribución de  $t_{\rho_\mu}$  bajo la hipótesis nula depende del valor del parámetro  $\mu$ , por lo que el contraste basado en este estadístico es un test no similar de la hipótesis de paseo aleatorio.

Concretamente, para  $\mu = 0$ , Dickey y Fuller (1979) caracterizaron la distribución límite



**Figura 3.** Densidad de  $t_{\rho_\mu}$  bajo  $\rho = 1$ , según el valor de  $\mu$   
**Fuente:** Elaboración propia.

de  $t_{\rho_\mu}$  para  $\rho = 1$ , es decir, la distribución límite del estadístico  $\hat{\tau}_\mu$ , bajo los supuestos<sup>6</sup>  $Y_0 = 0$  y  $e_t \sim \text{NID}(0, \sigma^2)$ . Se trata de una distribución no conocida, y Dickey (1976) obtuvo de forma empírica (por simulación) los valores críticos para el contraste de raíz unitaria, tanto para tamaños de muestra finitos como para la distribución límite (Fuller, 1996, p. 642).

Por otro lado, cuando  $\mu \neq 0$  Dickey (1976, p. 58) demuestra que, cuando  $T$  tiende a infinito, la distribución de  $t_{\rho_\mu}$  para  $\rho = 1$  es  $N(0, 1)$ . Esto hace pensar que en el caso de un tamaño de muestra finito dicha distribución se puede aproximar mediante la  $t$  de Student. En efecto, Nankervis y Savin (1985) prueban que cuando  $\mu \rightarrow \infty$ , la distribución de  $t_{\rho_\mu}$  para  $\rho = 1$  y un tamaño de muestra  $T$  fijo, tiende a una  $t$  de Student con  $T - 2$  grados de libertad, todo ello asumiendo en (6) que  $Y_0 = 0$  y que los errores son normales con media cero y varianza 1.

La distribución de  $\hat{\tau}_\mu$  que obtuvieron empíricamente Dickey y Fuller, tiene asimetría positiva. Además, para cada valor de  $T$  y un mismo nivel de significación, los valores críticos son menores que los correspondientes a una  $t$  de Student con  $T - 2$  grados de libertad. Por tanto, el resultado de Nankervis y Savin (1985) implica que, a medida que aumenta el valor de  $\mu$ , la distribución de  $\hat{\tau}_\mu$  se va desplazando a la derecha y haciéndose simétrica, acercándose a la  $t_{T-2}$ .

---

<sup>6</sup>Según Dickey y Fuller (1979, p. 430), la distribución límite de  $\hat{\tau}_\mu$  no depende de  $Y_0$  y se mantiene para  $e_t \sim \text{iid}(0, \sigma^2)$ .

Como en la práctica  $\mu$  no será infinito, la  $t_{T-2}$  es una aproximación a la distribución de  $t_{\rho_\mu}$  para  $\rho = 1$ ,  $\mu \neq 0$  y  $T$  finito que, según un estudio Monte Carlo de Nankervis y Savin (1985), es muy buena para  $\mu \geq 10$ .

### A.2) Contraste de la hipótesis $\rho = 1$

Todo lo anterior nos lleva a utilizar de forma simultánea las dos distribuciones extremas mencionadas para llevar a cabo en (6) el contraste unilateral de raíz unitaria anteriormente planteado. Así, en la figura 3,  $f_{DF}$  representa la densidad empírica de  $\hat{\tau}_\mu$  (siendo  $\rho = 1$  y  $\mu = 0$ ) para el valor de  $T$  considerado, mientras que  $f_S$  es la densidad de la  $t_{T-2}$ , también bajo  $\rho = 1$ , tomada como aproximación al caso en el que  $\mu \neq 0$ .

El valor  $DF_{1-\alpha}$  representa el cuantil de orden  $1 - \alpha$  de la distribución empírica de  $\hat{\tau}_\mu$ . De esta forma, el intervalo  $(DF_{1-\alpha}; +\infty)$  constituye la región de aceptación del contraste unilateral

$$\begin{aligned} H_0 : \rho &= 1 && \text{(asumiendo } \mu = 0) \\ H_1 : \rho &< 1 \end{aligned}$$

Por otro lado, el valor  $t_{T-2;\alpha}$  denota el cuantil de orden  $\alpha$  de la  $t$  de Student con  $T - 2$  grados de libertad. Así, el intervalo  $(-t_{T-2;\alpha}; +\infty)$  determinaría la región de aceptación del contraste unilateral

$$\begin{aligned} H_0 : \rho &= 1 && \text{(asumiendo } \mu \neq 0) \\ H_1 : \rho &< 1 \end{aligned}$$

que realizaríamos en el caso de que se conociera que  $\mu \neq 0$ .

Como lo habitual es no conocer nada sobre el valor del parámetro  $\mu$ , desarrollamos el contraste considerando las dos distribuciones simultáneamente, de forma que la región de aceptación de tamaño  $\alpha$  del contraste unilateral sería  $(-t_{T-2;\alpha}; +\infty)$ , **zona C** de la figura. Así, si el valor estimado del estadístico  $t_{\rho_\mu}$  ( $\hat{t}_{\rho_\mu}$  en lo sucesivo) cae en la zona C no se rechaza la hipótesis nula ( $\rho = 1$ ). Además, como la zona C es la intersección de las regiones de aceptación  $(DF_{1-\alpha}; +\infty)$  y  $(-t_{T-2;\alpha}; +\infty)$ , el valor  $\hat{t}_{\rho_\mu}$  llevaría a aceptar tanto la hipótesis nula del contraste basado únicamente en  $\hat{\tau}_\mu$ , como la del contraste basado solamente en la  $t_{T-2}$ . Por tanto, cuando  $\hat{t}_{\rho_\mu}$  esté en la zona C diremos que  $\rho = 1$ , pudiendo  $\mu$  tomar cualquier valor.

Los demás valores críticos determinan otras zonas que, en la figura 3, hemos llamado A y B.

La **zona A** está determinada por el intervalo  $(-\infty; DF_{1-\alpha})$ . Si el valor calculado del estadístico  $t_{\rho_\mu}$  cae en esta zona A se rechaza la hipótesis nula, y, en principio, se concluye que  $\rho < 1$ . No obstante, si tenemos en cuenta que  $DF_{1-\alpha} < -t_{T-2;\alpha}$ , entonces, si se

**Tabla 3.** Consecuencia según la zona en la que cae el valor  $\hat{t}_{\rho_\mu}$

Zona	Extremos	Consecuencia
A	$(-\infty; DF_{1-\alpha})$	$\rho < 1, \mu$ cualquiera
B	$(DF_{1-\alpha}; -t_{T-2;\alpha})$	$\rho = 1$ y $\mu = 0$ $\rho < 1$ y $\mu \neq 0$
C	$(-t_{T-2;\alpha}; +\infty)$	$\rho = 1, \mu$ cualquiera

Fuente: Elaboración propia.

ha rechazado la hipótesis nula ( $\rho = 1$ ) para la distribución empírica de  $\hat{\tau}_\mu$ , seguro que también se rechaza para la  $t_{T-2}$ . Esto nos lleva a pensar que el valor  $\hat{t}_{\rho_\mu}$  puede caer en la zona A, bien por ser  $\rho < 1$  y  $\mu = 0$ , o bien por ser  $\rho < 1$  y  $\mu \neq 0$ . Por tanto, cuando  $\hat{t}_{\rho_\mu}$  esté en la zona A la conclusión debe ser que  $\rho < 1$ , pudiendo  $\mu$  tomar cualquier valor.

Cuando el valor calculado del estadístico  $t_{\rho_\mu}$  cae en el intervalo  $(DF_{1-\alpha}; -t_{T-2;\alpha})$ , **zona B** en la figura 2, podemos encontrarnos en dos situaciones diferentes:

- a) El valor calculado de  $t_{\rho_\mu}$  está dentro del intervalo  $(DF_{1-\alpha}; +\infty)$ , por lo que no se rechaza la hipótesis nula del contraste basado en  $\hat{\tau}_\mu$ . De esta forma, la conclusión debe ser  $\rho = 1$  asumiendo que  $\mu = 0$ .
- b) Si el valor  $\hat{t}_{\rho_\mu}$  se compara con el valor crítico de la  $t_{T-2}, -t_{T-2;\alpha}$ , resulta que  $\hat{t}_{\rho_\mu} < -t_{T-2;\alpha}$ , rechazándose la hipótesis nula del contraste basado en la  $t_{T-2}$ . Por tanto, la conclusión sería  $\rho < 1$ , siendo  $\mu \neq 0$ , pero excluyendo el caso  $\rho < 1$  y  $\mu = 0$ , correspondiente al error de tipo II del contraste basado en  $\hat{\tau}_\mu$ .

Por tanto, si el valor  $\hat{t}_{\rho_\mu}$  está en la zona B lo más probable es que sea porque o bien  $\rho = 1$  y  $\mu = 0$ , o bien  $\rho < 1$  y  $\mu \neq 0$ .

Como se ha podido observar la decisión sobre  $\rho$  en la zona B es diferente dependiendo de la distribución que se utilice en el contraste. De esta forma, cuando el valor  $\hat{t}_{\rho_\mu}$  caiga en esta zona B, el no elegir la distribución correcta llevaría a tomar una decisión errónea sobre  $\rho$ . Por ello, llamamos a B, zona de duda.

Si pudiésemos determinar que  $\mu = 0$  o  $\mu \neq 0$ , podríamos decantarnos por una u otra distribución y poder así tomar una decisión sobre  $\rho$ . Esto nos lleva a analizar en la

siguiente sección la distribución del estadístico  $t$  asociado a la estimación MCO de  $\mu$  en (6).

Los valores críticos de  $t_{\rho_T}$  que determinan las zonas de aceptación, rechazo y duda descritas para este estadístico aparecen en la tabla C del anexo, para distintos tamaños muestrales.

Las consecuencias que se derivan según la zona en la que caiga el valor calculado de  $t_{\rho_T}$  aparecen resumidas en la tabla 3.

## B) CONTRASTE SOBRE $\mu$

### B.1) Distribución de $t_\mu$

El estadístico  $t$  asociado a la estimación MCO de  $\mu$  en (6), y que hemos denotado  $t_\mu$ , puede utilizarse para llevar a cabo el contraste de significación sobre  $\mu$

$$\begin{aligned} H_0 : \mu &= 0 \\ H_1 : \mu &\neq 0 \end{aligned}$$

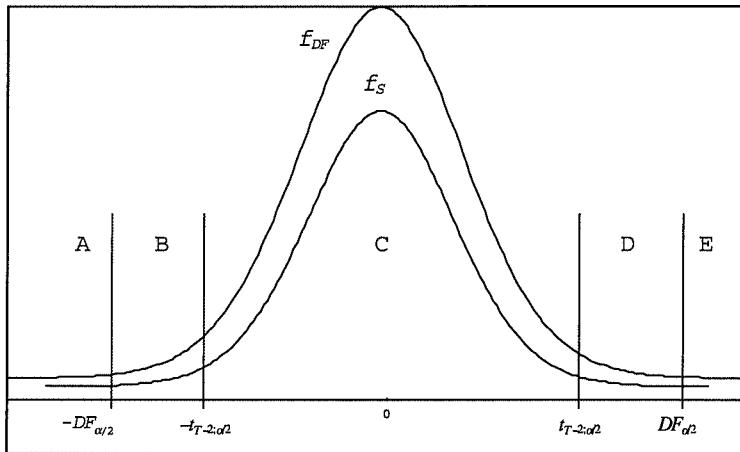
La distribución de este estadístico bajo  $H_0$ , depende del valor de  $\rho$ . Concretamente, cuando  $\rho = 1$ , el estadístico  $t_\mu$  bajo  $H_0$ , denotado por  $\hat{\tau}_{\alpha\mu}$ , tiene una distribución no conocida, cuya representación fue obtenida por Dickey y Fuller (1981). Estos autores obtuvieron por simulación algunos cuantiles, tanto para la distribución límite como para algunos tamaños de muestra finitos (Dickey y Fuller, 1981, p. 1062). Se trata de una distribución bimodal y simétrica con mayor dispersión que una  $t$  de Student.

Por otro lado, cuando  $|\rho| < 1$ , el teorema 2 de Fuller, Hasza y Goebel (1981) establece que la distribución límite del estadístico  $t_\mu$  para  $\mu = 0$ , bajo los supuestos de  $Y_0$  constante y  $e_t \sim \text{iid}(0, \sigma^2)$ , es  $N(0, 1)$ . Según el teorema 4 de estos mismos autores, esta normalidad asintótica se mantiene en el caso explosivo  $|\rho| > 1$ , si y sólo si  $e_t \sim NI(0, \sigma^2)$ .

En este último caso la  $N(0, 1)$  es la distribución asintótica, por lo que emplearemos la  $t$  de Student con  $T - 2$  grados de libertad como aproximación a la distribución de  $t_\mu$  para  $\mu = 0$ ,  $|\rho| \neq 1$  y tamaños de muestra finitos.

### B.2) Contraste de la hipótesis $\mu = 0$

Es claro que el estadístico  $t_\mu$  proporciona un test no similar de la hipótesis  $\mu = 0$ , por lo que, nuevamente, consideramos las dos distribuciones mencionadas para realizar el contraste bilateral sobre  $\mu$  planteado más arriba. En la figura 4 se representa conjuntamente



**Figura 4.** Densidad de  $t_\mu$  bajo  $\mu = 0$ , según el valor de  $\rho$   
**Fuente:** Elaboración propia.

la densidad empírica<sup>7</sup> de  $\hat{\tau}_{\alpha\mu}$  para un valor de  $T$  y la densidad de la  $t_{T-2}$ , denotadas  $f_{DF}$  y  $f_S$ , respectivamente.

En la figura 4 se puede observar la simetría de la distribución para el caso  $\rho = 1$ ,  $\mu = 0$ , aunque para un mismo nivel de significación estos valores críticos, en valor absoluto, son mayores que los de la  $t$  de Student. Concretamente,  $DF_{\alpha/2}$  representa el cuantil de orden  $\alpha/2$  de la distribución empírica de  $\hat{\tau}_{\alpha\mu}$ , de forma que el intervalo  $(-DF_{\alpha/2}; DF_{\alpha/2})$  constituye la región de aceptación del contraste bilateral

$$\begin{aligned} H_0 &: \mu = 0 \quad (\text{asumiendo } \rho = 1) \\ H_1 &: \mu \neq 0 \end{aligned}$$

que se llevaría a cabo en caso de que se conociera que  $\rho = 1$ .

En cambio, si se utilizan los valores críticos de la  $t$  de Student con  $T - 2$  grados de libertad, se obtiene la región de aceptación  $(-t_{T-2;\alpha/2}; t_{T-2;\alpha/2})$  correspondiente al contraste

$$\begin{aligned} H_0 &: \mu = 0 \quad (\text{asumiendo } |\rho| \neq 1) \\ H_1 &: \mu \neq 0 \end{aligned}$$

que realizaríamos en el caso de que se conociera que  $|\rho| \neq 1$ .

---

<sup>7</sup>Aunque no quede reflejado en la figura 4, la distribución empírica de  $\hat{\tau}_\mu$  es bimodal (ver Dickey y Fuller (1981), p. 1064).

Si no se dispone de información acerca de  $\rho$ , el contraste se lleva a cabo utilizando las dos distribuciones simultáneamente, cuyos valores críticos determinan una serie de zonas que quedan reflejadas en la figura 4. Esto permitirá sacar algunas conclusiones sobre los parámetros  $\mu$  y  $\rho$ , dependiendo de la zona en la que caiga el valor calculado del estadístico  $t_\mu$  ( $\hat{t}_\mu$  en lo sucesivo).

La **zona C**,  $(-t_{T-2;\alpha/2}; t_{T-2;\alpha/2})$ , constituye la región de aceptación del contraste sobre  $\mu$  basado en las dos distribuciones. Un valor estimado del estadístico  $t_\mu$  dentro de esta zona, pertenece al mismo tiempo a las regiones de aceptación de los contrastes basados en  $\hat{\tau}_{\alpha\mu}$  y en la  $t_{T-2}$ , lo que impide rechazar las hipótesis nulas de dichos contrastes. Por tanto, cuando el valor observado de  $t_\mu$  esté en la zona C se concluye que  $\mu = 0$ , pudiendo  $\rho$  tomar cualquier valor.

Los intervalos  $(-\infty; -DF_{\alpha/2})$  y  $(DF_{\alpha/2}; +\infty)$ , **zonas A y E**, respectivamente, constituyen la región crítica del contraste sobre  $\mu$  basado en las dos distribuciones. Cuando el valor observado del estadístico  $t_\mu$  cae en alguna de estas zonas, se tiene  $|\hat{t}_\mu| > |DF_{\alpha/2}| > |t_{T-2;\alpha/2}|$ , y se rechaza la hipótesis nula tanto del contraste basado en  $\hat{\tau}_{\alpha\mu}$  como la del contraste basado en la  $t_{T-2}$ . Por tanto, se concluye que  $\mu \neq 0$ , pudiendo  $\rho$  tomar cualquier valor.

Finalmente, si el valor calculado de  $t_\mu$  cae en la **zona B** o intervalo  $(-DF_{\alpha/2}; -t_{T-2;\alpha/2})$ , o en la **zona D** o intervalo  $(t_{T-2;\alpha/2}; DF_{\alpha/2})$ , pueden plantearse dos situaciones:

- a) Si dicho valor calculado se compara con los valores críticos de Dickey y Fuller, se llega a la conclusión de que  $\mu = 0$ , debiendo asumir que  $\rho = 1$ , pues se ha utilizado el contraste basado en  $\hat{\tau}_{\alpha\mu}$ .
- b) Si se compara con los valores críticos de la  $t_{T-2}$ , se rechaza la hipótesis nula ( $\mu = 0$ ). Además, teniendo en cuenta que el caso  $\mu \neq 0$ ,  $\rho = 1$  forma parte del error de tipo II del contraste basado en  $\hat{\tau}_{\alpha\mu}$ , se puede asumir que  $|\rho| \neq 1$ .

Por tanto, si el valor calculado del estadístico  $t_\mu$  cae en la zona B o D, lo más probable es que  $\mu = 0, \rho = 1$  o  $\mu \neq 0, |\rho| \neq 1$ .

Estamos nuevamente ante dos zonas de duda, la cual se resolvería si conociésemos si  $\rho = 1$  o  $|\rho| \neq 1$ . No obstante, si tenemos en cuenta que acudimos al contraste sobre  $\mu$  porque tenemos duda en el contraste sobre  $\rho$ , pero que a su vez puede surgir duda sobre  $\mu$  que solo se resuelve si tenemos información acerca de  $\rho$ , vemos que entramos en un círculo vicioso. Esta situación extrema la intentamos resolver con un contraste global tipo F.

Los valores críticos de  $t_\mu$  que delimitan las zonas descritas se encuentran en la tabla C del anexo. Por otro lado, la tabla 4 resume las conclusiones a las que se llega según la zona en la que cae el valor calculado de  $t_\mu$ .

**Tabla 4.** Consecuencia según la zona en la que cae el valor  $\hat{i}_\mu$

Zona	Extremos	Consecuencia
A	$(-\infty; -DF_{\alpha/2})$	$\mu \neq 0, \rho$ cualquiera
B	$(-DF_{\alpha/2}; -t_{T-2;\alpha/2})$	$\mu = 0, \rho = 1$ $\delta$ $\mu \neq 0,  \rho  \neq 1$
C	$(-t_{T-2;\alpha/2}; t_{T-2;\alpha/2})$	$\mu = 0, \rho$ cualquiera
D	$(t_{T-2;\alpha/2}; DF_{\alpha/2})$	$\mu = 0, \rho = 1$ $\delta$ $\mu \neq 0,  \rho  \neq 1$
E	$(DF_{\alpha/2}; +\infty)$	$\mu \neq 0, \rho$ cualquiera

**Fuente:** Elaboración propia.

### C) DISTRIBUCIÓN DEL ESTADÍSTICO $\Phi_1$ Y CONTRASTE ASOCIADO

A partir de la estimación MCO del modelo (6) se puede construir un estadístico tipo  $F$ , denotado  $\Phi_1$ , que permita realizar el contraste global

$$H_0 : \rho = 1, \quad \mu = 0 \\ H_1 : \text{no se cumple } H_0$$

Dickey y Fuller (1981) caracterizaron la distribución límite de  $\Phi_1$ , la cual no es una  $F$  de Snedecor, y obtuvieron empíricamente algunos valores críticos para tamaños de muestra finitos y para la distribución límite (Dickey y Fuller, 1981, Tabla IV, p. 1063).

Es evidente que si se acepta la hipótesis nula se concluye que  $\rho = 1$  y  $\mu = 0$ . Sin embargo, el rechazar  $H_0$  nos lleva a no poder decidir nada sobre  $\rho$ , pues el rechazo puede deberse a que  $\mu \neq 0$  y/o  $|\rho| \neq 1$ .

### 2.3. Planteamiento y estimación del modelo sin constante ni tendencia determinista lineal

Planteamos el modelo en su forma más simple. Por tanto, se estima por MCO el modelo

$$(7) \quad Y_t = \rho Y_{t-1} + e_t \quad t = 1, 2, \dots, T$$

donde  $\rho$  es un parámetro desconocido y para el que se asume  $Y_0 = 0$  y  $e_t \sim NI(0, 1)$ . De dicha estimación obtenemos el siguiente estimador y estadístico:

$\hat{\rho}$ : Estimador MCO del parámetro  $\rho$ .

$t_\rho$ : Estadístico  $t$  asociado a la estimación MCO de  $\rho$ .

El contraste más realizado es el de raíz unitaria ( $\rho = 1$ ), y al estadístico  $t_\rho$  para  $\rho = 1$  se le suele dar la notación específica de  $\hat{\tau}$ .

#### A) DISTRIBUCIÓN DE $\hat{\tau}$ Y CONTRASTE ASOCIADO

El estadístico  $\hat{\tau}$  permite llevar a cabo sobre (7) el contraste unilateral

$$\begin{aligned} H_0 : \rho &= 1 \\ H_1 : \rho &< 1 \end{aligned}$$

La distribución límite de este estadístico fue obtenida por Dickey y Fuller (1979) bajo los supuestos<sup>8</sup>  $Y_0 = 0$  y  $e_t \sim NID(0, \sigma^2)$ . Se trata de una distribución no conocida, para la que Dickey (1976) calculó por simulación algunos cuantiles, para tamaños de muestra finitos y para la distribución límite (Fuller, 1996, p. 642).

En este caso el estadístico proporciona un test similar ya que su distribución bajo la hipótesis nula de existencia de raíz unitaria no depende de los valores que tome ningún parámetro, puesto que no hay más parámetros. Por tanto, si el valor calculado de  $\hat{\tau}$  es menor que el valor crítico de Dickey y Fuller se rechaza la hipótesis nula y la serie  $Y_t$  no tiene raíz unitaria. En cambio, si el valor de  $\hat{\tau}$  es mayor que dicho valor crítico, concluimos que  $Y_t$  es un paseo aleatorio.

### 3. EVIDENCIA EMPÍRICA

Con el fin de ilustrar la problemática presentada, presentamos a continuación un ejemplo donde se puede apreciar el error en que se incurre a la hora de tomar una decisión sobre la existencia o no de raíz unitaria, quedando patente la incidencia de la no similaridad del contraste. Asimismo, para resolver este error, proponemos aplicar una estrategia de contraste secuencial en base a los desarrollos de la sección 2.

---

<sup>8</sup>En realidad, la distribución límite de  $\hat{\tau}$  no depende de  $Y_0$  y se mantiene para  $e_t \sim iid(0, \sigma^2)$  (Dickey y Fuller (1979, p. 430)).

Sea el siguiente proceso generador de datos:

$$Y_t = 1 + t + 0.8 Y_{t-1} + e_t \quad t = 1, 2, \dots, T$$

Como se puede apreciar, se trata de un proceso sin raíz unitaria, con constante ( $\mu = 1$ ) y con tendencia determinista lineal  $t$  de coeficiente  $\beta = 1$ . Tomando  $e_t \sim N(0, 1)$  generamos una muestra de tamaño  $T = 100$  para analizarla.

Nuestro objetivo es contrastar la existencia de raíz unitaria. Para ello partimos del supuesto de que no conocemos el proceso generador, y empezamos planteando el modelo completo:

$$(8) \quad Y_t = \mu + \beta t + \rho Y_{t-1} + e_t$$

De la estimación MCO de (8) se obtienen los valores calculados  $\hat{\rho}_\tau = -3.3588$  y  $\hat{\beta}_\tau = 3.3857$ , los cuales permiten contrastar las hipótesis nulas  $H_0 : \rho = 1$  y  $H_0 : \beta = 0$ , respectivamente.

Para llevar a cabo el contraste unilateral de raíz unitaria tal y como se suele hacer en la práctica, es decir, sin tener en cuenta la no similaridad, tomamos el valor crítico de Dickey-Fuller, que para  $T = 100$  y al 5%, es  $-3.45$  (Fuller, 1996, p. 642). En este caso,  $\hat{\tau}_\tau = \hat{\rho}_\tau = -3.3588$ , por lo que no se rechaza la existencia de raíz unitaria. Además, al utilizar el valor crítico de Dickey-Fuller, la decisión se debe tomar asumiendo que  $\beta = 0$ . Vemos, pues, que no considerar la no similaridad del contraste nos lleva a tomar una decisión errónea.

No obstante, si tenemos en cuenta la no similaridad del contraste basado en  $t_{\rho_\tau}$ , y consideramos las dos distribuciones extremas expuestas en la sección 2, se observa que el valor calculado del estadístico ( $-3.3588$ ) está comprendido entre los valores  $-3.45$  y  $-1.66$  (tabla A del anexo), es decir, está en lo que en la figura 1 hemos denominado zona B o zona de duda. Si estamos en esta zona, lo más probable es que sea porque  $\rho = 1$ ,  $\beta = 0$ , o porque  $\rho < 1$ ,  $\beta \neq 0$ , con independencia del valor de  $\mu$ , en ambos casos. Esto significa que no podemos llegar a una conclusión definitiva acerca de la existencia de raíz unitaria, ya que todo depende del valor de  $\beta$ . Por ello, planteamos el siguiente contraste:

$$\begin{aligned} H_0 &: \beta = 0 \\ H_1 &: \beta \neq 0 \end{aligned}$$

que como se vio en la sección 2, se trata de un test no similar, ya que la distribución del estadístico  $\hat{\beta}_\tau$  que nos permite llevar a cabo el contraste, depende de los valores que tomen tanto  $\rho$  como  $\mu$ . De la tabla B del anexo obtenemos los valores críticos ( $T = 100$ ) que corresponden al contraste bilateral de  $\beta$  al 5%, y como el valor calculado del estadístico  $\hat{\beta}_\tau$  es  $3.3857$ , se observa que estamos en lo que en la figura 2 hemos denominado zona F, es decir, la zona determinada por el valor crítico superior de

Dickey-Fuller (3.14) y por el valor crítico superior de la distribución  $-\hat{\tau}_T$  (3.73). Por tanto, también estamos en zona de duda en relación al contraste sobre  $\beta$ . No obstante, teniendo en cuenta lo comentado para esta zona F en la sección 2, podemos decir que lo más probable es que esta situación se deba bien a que  $\beta = 0$ ,  $\rho = 1$  y  $\mu \neq 0$ , o bien a  $\beta \neq 0$  con  $\rho$  y  $\mu$  cualesquiera. En cualquier caso, la duda persiste.

Ante esta situación recurrimos al estadístico  $\Phi_3$ , es decir, a contrastar

$$H_0 : \rho = 1, \quad \beta = 0$$

$$H_1 : \text{no se cumple } H_0$$

El valor calculado de  $\Phi_3$  resultó igual a 6.90, que comparado con el valor crítico de Dickey-Fuller al 5% para  $T = 100$  (6.49) nos lleva a rechazar la hipótesis nula, concluyendo que no se cumple al mismo tiempo que  $\rho = 1$  y  $\beta = 0$ . Esto nos permite descartar una de las dos situaciones más probables que se podían dar, y concluimos que  $\rho < 1$ ,  $\beta \neq 0$ , pudiendo  $\mu$  tomar cualquier valor.

Como se puede comprobar, al tener en cuenta la no similaridad de los contrastes de raíz unitaria, hemos llegado a tomar una decisión sobre los parámetros que coincide con el proceso generador de datos. En cambio, si no hubiésemos tenido en cuenta la no similaridad habríamos tomado una decisión errónea.

Esta estrategia de contraste utilizada en el ejemplo anterior aparece generalizada para los procesos generadores (5) y (6) en los diagramas A y B, respectivamente, del anexo.

#### 4. EXPERIMENTO MONTE CARLO

En esta sección analizamos el número de veces que los valores calculados de los estadísticos  $t$  considerados, caen en las diferentes zonas establecidas como consecuencia de tener en cuenta la no similaridad del contraste de raíz unitaria.

Para ello, en primer lugar, se diseñó un experimento Monte Carlo usando el modelo

$$(9) \quad Y_t = \mu + \rho Y_{t-1} + e_t \quad t = 1, \dots, T$$

Tomando en (9)  $Y_0 = 0$  y  $e_t \sim N(0, 1)$ , se generaron 2000 muestras de tamaño  $T = 100$  para  $\rho = 0.8, 0.9, 0.95, 0.99, 1$ , y  $\mu = 0, 0.5, 1, 5, 10$ . Todas las simulaciones se llevaron a cabo usando rutinas desarrolladas en EViews 3.0.

Con cada serie generada se estimó por MCO un modelo autorregresivo de orden 1 con constante, obteniéndose los valores calculados de los estadísticos  $t_{\rho\mu}$  y  $t_\mu$ .

Las tablas 5 y 6 recogen el número de veces en % que el valor calculado de cada estadístico cae en las diferentes zonas establecidas para cada uno de ellos en la sección

2 para un  $\alpha = 0.05$ . Por ejemplo, cuando  $\mu = 1$  y  $\rho = 0.9$ , el 74.70% de los valores calculados de  $t_{\rho\mu}$  caen en la zona A, el 25.05% en la zona B, y el 0.25% restante en la zona C.

**Tabla 5.** Número de veces (en %) que el valor calculado  $t_{\rho\mu}$  cae en cada zona

$T = 100$		ZONA		
$\alpha = 0.05$		A	B	C
		$-\infty$	$DF_{1-\alpha}$	$-t_{T-2;\alpha}$
$\mu = 0$	$\rho = 0.8$	87.10	12.90	
	$\rho = 0.9$	30.85	63.95	5.20
	$\rho = 0.95$	10.90	64.15	24.95
	$\rho = 0.99$	6.45	44.30	49.25
	$\rho = 1$	5	40.05	54.95
$\mu = 0.5$	$\rho = 0.8$	90.90	9.10	
	$\rho = 0.9$	43.20	54.70	2.10
	$\rho = 0.95$	22.15	65	12.85
	$\rho = 0.99$	5.50	31.60	62.90
	$\rho = 1$	0.85	9.20	89.95
$\mu = 1$	$\rho = 0.8$	96.60	3.40	
	$\rho = 0.9$	74.70	25.05	0.25
	$\rho = 0.95$	66.10	32.25	1.65
	$\rho = 0.99$	19.75	45.60	34.65
	$\rho = 1$	0.30	6.20	93.50
$\mu = 5$	$\rho = 0.8$	100		
	$\rho = 0.9$	100		
	$\rho = 0.95$	100		
	$\rho = 0.99$	100		
	$\rho = 1$	0.25	5.25	94.50
$\mu = 10$	$\rho = 0.8$	100		
	$\rho = 0.9$	100		
	$\rho = 0.95$	100		
	$\rho = 0.99$	100		
	$\rho = 1$	0.40	4.35	95.25

**Fuente:** Elaboración propia.

**Tabla 6.** Número de veces (en %) que el valor calculado  $t_\mu$  cae en cada zona

		ZONA					
		A	B	C	D	E	
		$-\infty$	$-DF_{\alpha/2}$	$-t_{T-2;\alpha/2}$	$t_{T-2;\alpha/2}$	$DF_{\alpha/2}$	$+\infty$
$\mu = 0$	$\rho = 0.8$	0.45	2.85	92.65	3.65	0.40	
	$\rho = 0.9$	0.55	3.60	91.35	4.25	0.25	
	$\rho = 0.95$	0.90	4.70	88.35	5.45	0.60	
	$\rho = 0.99$	2.25	9.05	77.60	9.15	1.95	
	$\rho = 1$	2.05	10.90	72.95	11.50	2.60	
$\mu = 0.5$	$\rho = 0.8$			2.25	30.90	66.85	
	$\rho = 0.9$			9.25	50.40	40.35	
	$\rho = 0.95$			14.15	54.50	31.35	
	$\rho = 0.99$			14.70	43.25	42.05	
	$\rho = 1$			12.10	38.10	49.80	
$\mu = 1$	$\rho = 0.8$				3.95	96.05	
	$\rho = 0.9$			0.20	13.70	86.10	
	$\rho = 0.95$			0.10	8.40	91.50	
	$\rho = 0.99$				1.10	98.90	
	$\rho = 1$				0.20	99.80	
$\mu = 5$	$\rho = 0.8$					100	
	$\rho = 0.9$					100	
	$\rho = 0.95$					100	
	$\rho = 0.99$					100	
	$\rho = 1$					100	
$\mu = 10$	$\rho = 0.8$					100	
	$\rho = 0.9$					100	
	$\rho = 0.95$					100	
	$\rho = 0.99$					100	
	$\rho = 1$					100	

**Fuente:** Elaboración propia.

La tabla 5 muestra que la distribución empírica de  $t_{\rho_\mu}$  bajo  $\rho = 1$  (hipótesis nula) está fuertemente influenciada por el valor de  $\mu$ . Así, cuando  $\mu = 0$  y  $\rho = 1$ , el 5% de los

valores calculados caen en la zona A, es decir, a la izquierda del valor crítico de Dickey-Fuller, confirmándose la distribución empírica que para el estadístico  $\hat{\tau}_\mu$  obtuvieron estos autores. No obstante, cuando  $\mu$  aumenta ( $\mu = 1, 5, 10$ ), aproximadamente el 95% de los valores calculados caen en la zona C, es decir, a la derecha del valor crítico de la  $t$  de Student, y el resto en la zona B. Este hecho confirma el resultado de Nankervis y Savin (1985) relativo a que cuando  $\mu$  tiende a infinito la distribución de  $t_{\rho_\mu}$  bajo  $\rho = 1$  tiende a una  $t$  de Student con  $T - 2$  grados de libertad.

Por otro lado, se observa que cuando  $\rho$  está próximo a la unidad (0.9, 0.95, 0.99) un alto porcentaje de valores calculados de  $t_{\rho_\mu}$  caen en la zona de duda B. Por ejemplo, cuando  $\mu = 0.5$  y  $\rho = 0.95$ , el 65% de estos valores caen en la citada zona B. De esta forma, si en este caso se usan los valores críticos de Dickey-Fuller para llevar a cabo el contraste de raíz unitaria, aceptaríamos erróneamente la hipótesis nula en el 65% de los casos. En cambio, si se utilizan los valores críticos de la  $t$  de Student se tomaría la decisión correcta. Vemos que la baja potencia que caracteriza al contraste de Dickey-Fuller para alternativas próximas a la unidad es debida a la no similaridad de dicho contraste. De esta forma, si se tiene en cuenta la no similaridad llevando a cabo el contraste en la forma expuesta en el ejemplo de la sección 3, podemos mejorar, en parte, la potencia.

Por otra parte, en la tabla 6 se puede observar la influencia del parámetro  $\rho$  en la distribución empírica del estadístico  $t_\mu$  bajo  $\mu = 0$ . Concretamente, cuando  $\rho = 1$ , aproximadamente el 95% de los valores calculados de  $t_\mu$  caen entre los valores críticos de Dickey-Fuller (zonas B, C y D), lo que confirma los resultados que sobre la distribución de  $\hat{\tau}_{\alpha\mu}$  obtuvieron estos autores. En cambio, a medida que disminuye  $\rho$ , ese 95% se va concentrando en la zona C, aproximándose al comportamiento de una  $t$  de Student.

En las alternativas, es decir, en los casos en que  $\mu \neq 0$ , el valor del parámetro  $\rho$  apenas influye en la distribución del porcentaje de valores calculados en las distintas zonas. Así, cuando  $\mu = 5$  y 10, la totalidad de estos valores caen en la zona E o zona de rechazo con ambas distribuciones. El problema surge cuando  $\mu = 0.5$  y 1, pues, en estos casos un alto porcentaje de valores calculados caen en la zona de duda D. Este porcentaje (con un valor máximo de 54.50 en  $\mu = 0.5$  y  $\rho = 0.95$ ) corresponde a aceptaciones erróneas de la hipótesis nula ( $\mu = 0$ ), motivadas por el uso indebido, para el contraste, de la distribución de Dickey-Fuller, en lugar de la  $t$  de Student.

De forma análoga, se diseñó un experimento Monte Carlo usando el modelo

$$(10) \quad Y_t = \mu + \beta t + \rho Y_{t-1} + e_t \quad t = 1, 2, \dots$$

Tomando en (10)  $Y_0 = 0$  y  $e_t \sim N(0, 1)$ , se generaron 2000 muestras de tamaño  $T = 100$  para  $\rho = 0.8, 0.9, 0.95, 0.99, 1$ ,  $\beta = 0, 0.1, 1$ , y  $0, 1, 10$ .

Con cada serie generada se estimó por MCO un modelo autorregresivo de orden 1 con constante y tendencia determinista lineal, obteniéndose los valores calculados de los estadísticos  $t_{\rho_\tau}$  y  $t_{\beta_\tau}$ .

**Tabla 7.** Número de veces (en %) que el valor calculado  $t_{\rho_T}$  cae en cada zona

		$T = 100$		ZONA		
		$\alpha = 0.05$		A	B	C
		$-\infty$		$DF_{1-\alpha}$		$-t_{T-3;\alpha}$
$\rho = 0.8$	$\mu = 0$	$\beta = 0$	64.35	35.65		
		$\beta = 0.1$	66.20	33.75	0.05	
		$\beta = 1$	100			
	$\mu = 1$	$\beta = 0$	78.25	21.25		
		$\beta = 0.1$	71.70	28.30		
		$\beta = 1$	96.75	3.25		
	$\mu = 10$	$\beta = 0$	100			
		$\beta = 0.1$	100			
		$\beta = 1$	100			
$\rho = 0.9$	$\mu = 0$	$\beta = 0$	19.45	78.80	1.75	
		$\beta = 0.1$	28.90	70.15	0.95	
		$\beta = 1$	100			
	$\mu = 1$	$\beta = 0$	34.60	64.45	0.95	
		$\beta = 0.1$	18.95	79.50	1.55	
		$\beta = 1$	100			
	$\mu = 10$	$\beta = 0$	100			
		$\beta = 0.1$	100			
		$\beta = 1$	82.45	17.55		
$\rho = 0.95$	$\mu = 0$	$\beta = 0$	8.70	80.95	10.35	
		$\beta = 0.1$	37.60	58.90	3.50	
		$\beta = 1$	100			
	$\mu = 1$	$\beta = 0$	14.25	76.10	9.65	
		$\beta = 0.1$	12.55	74.60	12.85	
		$\beta = 1$	100			
	$\mu = 10$	$\beta = 0$	100			
		$\beta = 0.1$	100			
		$\beta = 1$	100			
$\rho = 0.99$	$\mu = 0$	$\beta = 0$	4.75	72.60	22.65	
		$\beta = 0.1$	17.65	62.50	19.85	
		$\beta = 1$	100			
	$\mu = 1$	$\beta = 0$	3.50	57.10	39.40	
		$\beta = 0.1$	11.75	61.50	26.75	
		$\beta = 1$	100			
	$\mu = 10$	$\beta = 0$	19.30	62.35	18.35	
		$\beta = 0.1$	5.50	70.80	23.70	
		$\beta = 1$	100			
$\rho = 1$	$\mu = 0$	$\beta = 0$	5.05	71.60	23.35	
		$\beta = 0.1$	0.10	7.30	92.60	
		$\beta = 1$	0.10	4.50	95.40	
	$\mu = 1$	$\beta = 0$	5.40	69.05	25.55	
		$\beta = 0.1$	0.10	6.45	93.45	
		$\beta = 1$	0.10	5.65	94.25	
	$\mu = 10$	$\beta = 0$	5.65	69.95	24.40	
		$\beta = 0.1$	0.10	6.30	93.60	
		$\beta = 1$		5.85	94.15	

**Fuente:** Elaboración propia.

Las tablas 7 y 8 recogen el número de veces en % que el valor calculado de cada estadístico cae en las diferentes zonas establecidas para cada uno de ellos en la sección 2 para un  $\alpha = 0.05$ .

En la tabla 7 se puede observar la influencia del valor del parámetro  $\beta$  en la distribución empírica de  $t_{\rho_T}$  bajo  $\rho = 1$ . Cuando  $\beta = 0$ , se cumple lo establecido por Dickey y Fuller, es decir, aproximadamente el 95% de los valores calculados de  $t_{\rho_T}$  caen a la derecha del valor crítico de Dickey-Fuller (zonas B y C). A medida que aumenta el valor de  $\beta$ , ese 95% se concentra en la zona C, esto es, a la derecha del valor crítico de la  $t$  de Student. Asimismo, se observa que este comportamiento no se ve afectado por el valor  $\mu$ , confirmándose lo comentado en la sección 2 sobre la independencia de este parámetro en la distribución de  $t_{\rho_T}$  bajo  $\rho = 1$ .

Para los demás valores de  $\rho$ , se observa una influencia muy fuerte de la tendencia determinista lineal, pues, para  $\beta = 1$  la totalidad de valores calculados del estadístico  $t_{\rho_T}$  caen en la zona A, es decir, en una zona donde se rechaza tanto con la distribución de Dickey-Fuller como con la  $t$  de Student. En cambio, cuando  $\beta = 0.1$ , esta influencia es menor, produciéndose un alto porcentaje de dudas. Por ejemplo, cuando  $\rho = 0.9$ ,  $\mu = 1$  y  $\beta = 0.1$ , el 79.5% caen en zona de duda (zona B). Esto se traduciría en una decisión errónea, si para el contraste de raíz unitaria se utilizan los valores críticos de Dickey-Fuller. De nuevo mejoraría la potencia del contraste si tenemos en cuenta de forma simultánea las dos distribuciones implicadas.

Finalmente, cuando  $\rho < 1$  y  $\beta = 0$ , el porcentaje de valores calculados en las zonas B y C, pone de manifiesto la baja potencia del test de Dickey-Fuller para alternativas próximas a la unidad.

Por último, en la tabla 8 se observa la influencia de los parámetros  $\mu$  y  $\rho$  en la distribución empírica de  $t_{\beta_T}$  bajo  $\beta = 0$ . Por un lado, cuando  $\rho = 1$  y  $\mu = 0$ , el 95% de los valores calculados del estadístico  $t_{\beta_T}$  se distribuyen de una forma simétrica en las zonas B, C, D y E, tal y como establecen Dickey y Fuller. En cambio, cuando  $\mu$  aumenta, este 95% se concentra en las zonas D, E y F, presentando un mayor porcentaje las dos primeras (36.75% y 49.60%, respectivamente, cuando  $\mu = 1$ ). Esto confirma lo comentado en la sección 2 en relación a que la distribución de  $t_{\beta_T}$  para  $\beta = 0$  y  $\rho = 1$ , tiende a la imagen en el espejo de la distribución de  $\hat{\tau}_T$ . Por otro lado, cuando  $\rho \neq 1$  se observa una mayor concentración de los valores calculados de  $t_{\beta_T}$  en las zonas delimitadas por los valores críticos de la  $t_{T-3}$ , no afectando el valor del parámetro  $\mu$ .

En los casos en que  $\beta \neq 0$  y  $\rho < 1$ , el porcentaje de valores calculados que caen en la zona E corresponde a aceptaciones indebidas de la hipótesis nula provocadas por el uso indebido de los valores críticos de Dickey-Fuller en lugar de los de la  $t$  de Student. Encontramos un valor máximo de 59.75% para  $\beta = 0.1$ ,  $\mu = 1$  y  $\rho = 0.9$ .

**Tabla 8.** Número de veces (en %) que el valor calculado  $t_{\beta_T}$  cae en cada zona

		T = 100 $\alpha = 0.05$						
		ZONA						
		A	B	C	D	E	F	G
		$-\infty$	$-DF_{\alpha/2}$	$-t_{T-3}\alpha/2$	$IM_{1-\alpha/2}$	$t_{T-3;\alpha/2}$	$DF_{\alpha/2}$	$IM_{\alpha/2}$
$\rho = 0.8$	$\mu = 0$	$\beta = 0$	0.30	4.40	65	25.70	4.25	0.20
		$\beta = 0.1$				0.10	15.75	35.85
		$\beta = 1$					0.10	99.90
	$\mu = 1$	$\beta = 0$	0.10	3.70	62.75	27.85	5	0.40
		$\beta = 0.1$				0.05	16.85	32.50
		$\beta = 1$					0.55	7.70
$\rho = 0.9$	$\mu = 10$	$\beta = 0$	0.10	2.65	69.10	26.05	2	0.10
		$\beta = 0.1$						100
		$\beta = 1$						100
	$\mu = 0$	$\beta = 0$	0.75	6.15	61.35	26.10	4.90	0.50
		$\beta = 0.1$				2.20	46.20	31.65
		$\beta = 1$						19.95
$\rho = 0.95$	$\mu = 1$	$\beta = 0$	0.10	2.70	58	30.60	8.05	0.40
		$\beta = 0.1$				7.60	59.75	22.35
		$\beta = 1$						100
	$\mu = 10$	$\beta = 0$	0.10	1.75	70.40	25.10	2.50	0.15
		$\beta = 0.1$						100
		$\beta = 1$				0.20	8.25	22.90
$\rho = 0.99$	$\mu = 0$	$\beta = 0$	0.85	9.10	55.60	24.80	8.85	0.60
		$\beta = 0.1$			0.05	2.70	26.20	28.95
		$\beta = 1$						42.10
	$\mu = 1$	$\beta = 0$	0.05	1.75	46.60	39.90	10.45	1.10
		$\beta = 0.1$			0.45	18.90	55.60	17.35
		$\beta = 1$						7.70
$\rho = 1$	$\mu = 10$	$\beta = 0$	0.10	2.05	68.20	25.90	3.70	0.05
		$\beta = 0.1$					0.10	99.90
		$\beta = 1$						100
	$\mu = 0$	$\beta = 0$	2.10	11.35	48.85	25	10.65	1.50
		$\beta = 0.1$					0.40	1.25
		$\beta = 1$						98.35
$\rho = 1$	$\mu = 1$	$\beta = 0$		0.05	20.70	46.20	28.80	2.950
		$\beta = 0.1$				0.10	5.20	9.45
		$\beta = 1$						85.25
	$\mu = 10$	$\beta = 0$	0.05	1.10	64.35	30.40	40	0.05
		$\beta = 0.1$			2.45	36.30	50.35	8.40
		$\beta = 1$						2.50
	$\mu = 0$	$\beta = 0$						100
		$\beta = 0.1$						
		$\beta = 1$						
$\rho = 1$	$\mu = 1$	$\beta = 0$			2.95	36.75	49.60	8
		$\beta = 0.1$					0.10	0.95
		$\beta = 1$						98.95
	$\mu = 10$	$\beta = 0$			2.30	38.60	48.35	7.90
		$\beta = 0.1$			2.40	25.75	41.85	16.45
		$\beta = 1$						13.55
	$\mu = 0$	$\beta = 0$						100
		$\beta = 0.1$						
		$\beta = 1$						

**Fuente:** Elaboración propia.

## 5. CONCLUSIONES

Con todo el desarrollo hasta aquí expuesto, del análisis de las distribuciones de los estadísticos que se emplean para los contrastes de raíces unitarias, podemos concluir lo siguiente:

- Queda patente la gran incidencia que tiene el carácter no similar de dichos contrastes.
- Si no se tiene en cuenta este problema, se puede incurrir claramente en decisiones erróneas cuando se realizan aplicaciones econométricas utilizando sólo los valores críticos de Dickey y Fuller para rechazar o no que  $\rho = 1$ .
- Como consecuencia del carácter no similar de los contrastes, se originan zonas de duda en el espacio paramétrico de los estadísticos, que crean incertidumbre a la hora de decidir sobre el rechazo de la hipótesis nula.
- Sin embargo, del estudio realizado se deduce que en caso de que exista incertidumbre en el contraste de  $\rho$  y también en el de  $\beta$ , la aplicación de los contrastes  $\Phi_1$  o  $\Phi_3$  puede llevarnos a la toma de decisión adecuada.
- De los resultados del experimento Monte Carlo se deduce que el porcentaje de casos que darían lugar a la aceptación indebida de la hipótesis nula, puede ser importante, habiendo resultado un valor superior al 79% en el modelo autorregresivo de orden 1 con constante y tendencia determinista lineal, y un 65% en el modelo solo con constante.

Actualmente, se está desarrollando una estrategia de contraste desde el punto de vista de la no similaridad, es decir, teniendo en cuenta las diferentes distribuciones correspondientes a los estadísticos implicados. En este sentido, se ha llevado a cabo un estudio Monte Carlo para analizar el funcionamiento de la misma en el modelo autorregresivo de orden 1 con constante (Díos y Roldán, 1998).

## 6. REFERENCIAS BIBLIOGRÁFICAS

- Ahn, S.K. (1993). «Some test for unit roots in autorregressive-integrated-moving average models with deterministic trends», *Biometrika*, 80, 855-868.
- Bhargava, A. (1986). «On the theory of testing for unit roots in observed time series», *Review of Economic Studies*, 53, 369-384.
- Bierens, H.J. and Guo, S. (1993). «Testing stationary and trend stationary against the unit root hypothesis», *Econometrics Reviews*, 12, 1-32.
- Choi, I. (1990). «Most U.S. Economic Time Series Do Not Have Unit Roots: Nelson and Plosser. Results Reconsidered», manuscrito, Universidad Estatal de Ohio.

- Dickey, D.A. (1976). «Estimation and Hypothesis Testing in Nonstationary Time Series». Ph.D. dissertation, Iowa State University.
- (1984). «Powers of unit root tests», *Proceedings of The American Statistical Association*, Business and Economic Statistics Section, 489-493.
- Dickey, D.A. and Fuller, W.A. (1979). «Distribution of the Estimators for Autoregressive Time Series With a Unit Root», *Journal of the American Statistical Association*, 74, 427-431.
- (1981). «Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root», *Econometrica*, 49, 1057-1072.
- Dios, R. and Roldán, J.A. (1998). «Analysis of a non-similar unit root test: a Monte Carlo Investigation», *First International Conference on Applied Sciences and the Environment (ASE 98)*, 5-7 October 1998, Cadiz, Spain.
- Evans, G.B.A. and Savin, N.E. (1984). «Testing for unit roots: 2», *Econometrica*, 52, 5, 1241-1269.
- Ferretti, N. and Romo, J. (1996). «Unit root bootstrap tests for AR(1) models», *Biometrika*, 83, 4, 849-860.
- Fukushige, M.; Hatanaka, M. and Koto, Y. (1990). «A test on the attainability of equilibrium and its applications to international capital markets», Invited paper at the 6th World Congress of Econometric Society, August 1990.
- Fuller, Wayne A. (1976). *Introduction to Statistical Times Series*, New York: John Wiley & Sons.
- (1996). *Introduction to Statistical Times Series*, Second Edition, New York: John Wiley & Sons.
- Fuller, W.A., Hasza, P.D. and Goebel, J.J. (1981). «Estimation of the parameters of stochastic difference equations», *The Annals of Statistics*, 9, 3, 531-543.
- Guilkey, D.K. and Schmidt, P. (1989). «Extended tabulations for Dickey-Fuller tests», *Economics Letters*, 31, 355-357.
- Hall, A. (1989). «Testing for a unit root in the presence of moving average errors», *Biometrika*, 76, 49-56.
- (1992). «Joint hypothesis tests for a random walk based on instrumental variables estimators», *Journal of Time Series Analysis*, 12, 29-45.
- Hasan, M.N. and Koenker, R.W. (1997). «Robust rank tests of the unit root hypothesis», *Econometrica*, 65, 1, 133-161.
- Hasza, D.P. (1977). «Estimation in Nonstationary Time Series». Ph.D. dissertation, Iowa State University.
- Hwang, J. and Schmidt, P. (1996). «Alternative methods of detrending and the power of unit root tests», *Journal of Econometrics*, 71, 227-248.
- Kahn, J.A. and Ogaki, M. (1992). «A consistent test for the null of stationarity against the alternative of a unit root», *Economics Letters*, 39, 7-11.
- Kiviet, J.F. and Phillips, G.D.A. (1992). «An exact similar t-type test for unit roots», *University of Amsterdam*, mimeographed.

- (1992). «Exact similar tests for unit roots and cointegration», *Oxford Bulletin of Economics and Statistics*, 54, 3, 349-367.
- Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P. and Shin, Y. (1992). «Testing the null hypothesis of stationary against the alternative of a unit root. How sure are we that economic time series have a unit root?», *Journal of Econometrics*, 54, 159-178.
- Leybourne, S.J. and McCabe, B.M.P. (1994). «A consistent test for a unit root», *Journal of Business and Economic Statistics*, 2, 157-166.
- Nankervis, J.C. and Savin, N.E. (1985). «Testing the autoregressive parameter with the  $t$  statistic», *Journal of Econometrics*, 27, 143-161.
- (1987). «Finite sample distributions of  $t$  and  $F$  statistics in an AR(1) model with an exogenous variable», *Econometric Theory*, 3, 387-408.
- Nelson, C.R. and Plosser, C.I. (1982). «Trends and random walks in macroeconomic time series: Some evidence and implications», *Journal of Monetary Economics*, 10, 139-162.
- Ogaki, M. and Park, J.Y. (1990). «A cointegration approach to estimating preference parameters», Paper presented at the Conference on Empirical Applications of Structural Models organized by Econometrica and Social Systems Research Institute (University of Wisconsin, Madison, WI).
- Pantula, S.G. and Hall, A. (1991). «Testing for unit roots in autoregressive moving average models», *Journal of Econometrics*, 48, 325-353.
- Phillips, P.C.B. (1987). «Time series regression with a unit root», *Econometrica*, 55, 2, 277-301.
- Phillips, P.C.B. and Perron, P. (1988). «Testing for a unit root in time series regression», *Biometrika*, 75, 335-346.
- Said, S.E. and Dickey, D.A. (1984). «Testing for unit roots in autoregressive-moving average models of unknown order», *Biometrika*, 71, 599-607.
- Sargan, J.D. and Bhargava, A. (1983). «Testing residuals from least squares regression for being generated by the gaussian random walk», *Econometrica*, 51, 153-174.
- Schmidt, P. and Phillips, P.C.B. (1992). «LM test for a unit root in the presence of deterministic trends», *Oxford Bulletin of Economics and Statistics*, 54, 257-287.
- Schwert, G.W. (1989). «Tests for unit roots: A Monte Carlo investigation», *Journal of Business and Economic Statistics*, 7, 147-159.
- Shin, D.W. and So, B. (1997). «Semiparametric unit root tests based on symmetric estimators», *Statistical Probability Letters*, 33, 2, 177-184.
- Solo, V. (1984). «The order of differencing in ARIMA models», *Journal of the American Statistical Association*, 79, 388, 916-921.

## ANEXO

**Tabla A.** Valores críticos del estadístico  $t_{\rho_T}$  para  $\rho = 1$ , según los valores de  $\beta$   
(modelo  $Y_t = \mu + \beta t + \rho Y_{t-1} + e_t$ )

$T$	$\alpha$	$DF_{1-\alpha}$	$-t_{T-3;\alpha}$
25	0.01	-4.38	-2.51
	0.025	-3.95	-2.07
	0.05	-3.60	-1.72
	0.10	-3.24	-1.32
50	0.01	-4.15	-2.41
	0.025	-3.80	-2.01
	0.05	-3.50	-1.68
	0.10	-3.18	-1.30
100	0.01	-4.04	-2.37
	0.025	-3.73	-1.98
	0.05	-3.45	-1.66
	0.10	-3.15	-1.29
250	0.01	-3.99	-2.33
	0.025	-3.69	-1.96
	0.05	-3.43	-1.65
	0.10	-3.13	-1.28
$\infty$	0.01	-3.96	-2.34
	0.025	-3.66	-1.97
	0.05	-3.41	-1.65
	0.10	-3.12	-1.28

Fuente: Elaboración propia

$T$ : tamaño muestral.

$\alpha$ : nivel de significación

$DF$ : valores críticos de Dickey-Fuller.

$t_{T-3}$  : valores críticos de la  $t$  de Student

**Tabla B.** Valores críticos del estadístico  $t_{\beta_T}$  para  $\beta = 0$  según los valores de  $\rho$  y  $\mu$   
(modelo  $Y_t = \mu + \beta t + \rho Y_{t-1} + e_t$ )

$T$	$\alpha$	Zona B   Zona C		Zona E   Zona F		
		$-DF_{\alpha/2}$	$-t_{T-3;\alpha/2}$	$IM_{1-\alpha/2}$	$t_{T-3;\alpha/2}$	$DF_{\alpha/2}$
25	$\alpha/2 = 0.025$	-3.25	-2.07	0.50	2.07	3.25
	$\alpha/2 = 0.05$	-2.85	-1.72	0.80	1.72	2.85
50	$\alpha/2 = 0.025$	-3.18	-2.01	0.58	2.01	3.18
	$\alpha/2 = 0.05$	-2.81	-1.68	0.87	1.68	2.81
100	$\alpha/2 = 0.025$	-3.14	-1.98	0.62	1.98	3.14
	$\alpha/2 = 0.05$	-2.79	-1.66	0.90	1.66	2.79
250	$\alpha/2 = 0.025$	-3.12	-1.96	0.64	1.96	3.12
	$\alpha/2 = 0.05$	-2.79	-1.65	0.92	1.65	2.79
$\infty$	$\alpha/2 = 0.025$	-3.11	-1.96	0.66	1.96	3.11
	$\alpha/2 = 0.05$	-2.78	-1.65	0.94	1.65	2.78

Fuente: Elaboración propia

$T$ : tamaño muestral.

$\alpha$ : nivel de significación

$DF$ : valores críticos de Dickey-Fuller.

$t_{T-3}$  : valores críticos de la  $t$  de Student

$IM$ : valores de la imagen en el espejo de los valores críticos de Dickey-Fuller para  $\hat{\tau}_T$

**Tabla C. Valores críticos del estadístico  $t_{\rho\mu}$  para  $\rho = 1$ , según los valores de  $\mu$  (modelo  $Y_t = \mu + \rho Y_{t-1} + e_t$ )**

$T$	$\alpha$	$DF_{1-\alpha}$	$-t_{T-2;\alpha}$
25	0.01	-3.75	-2.50
	0.025	-3.33	-2.07
	0.05	-3.00	-1.71
	0.10	-2.63	-1.32
50	0.01	-3.58	-2.41
	0.025	-3.22	-2.01
	0.05	-2.93	-1.68
	0.10	-2.60	-1.30
100	0.01	-3.51	-2.37
	0.025	-3.17	-1.98
	0.05	-2.89	-1.66
	0.10	-2.58	-1.29
250	0.01	-3.46	-2.33
	0.025	-3.14	-1.96
	0.05	-2.88	-1.65
	0.10	-2.57	-1.28
$\infty$	0.01	-3.43	-2.34
	0.025	-3.12	-1.97
	0.05	-2.86	-1.65
	0.10	-2.57	-1.28

Fuente: Elaboración propia

$T$ : tamaño muestral.

$\alpha$ : nivel de significación

$DF$ : valores críticos de Dickey-Fuller.

$t_{T-2}$ : valores críticos de la  $t$  de Student

**Tabla D. Valores críticos del estadístico  $t_\mu$  para  $\mu = 0$  según los valores de  $\rho$  (modelo  $Y_t = \mu + \rho Y_{t-1} + e_t$ )**

$T$	$\alpha$	Zona B		
		$-DF_{\alpha/2}$	$-t_{T-2;\alpha/2}$	$t_{T-2;\alpha/2}$
25	$\alpha/2 = 0.025$	-2.97	-2.07	2.07
	$\alpha/2 = 0.05$	-2.61	-1.71	1.71
50	$\alpha/2 = 0.025$	-2.89	-2.01	2.01
	$\alpha/2 = 0.05$	-2.56	-1.68	1.68
100	$\alpha/2 = 0.025$	-2.86	-1.98	1.98
	$\alpha/2 = 0.05$	-2.54	-1.66	1.66
250	$\alpha/2 = 0.025$	-2.54	-1.66	1.66
	$\alpha/2 = 0.05$	-2.53	-1.65	1.65
$\infty$	$\alpha/2 = 0.025$	-2.83	-1.97	1.97
	$\alpha/2 = 0.05$	-2.52	-1.65	1.65

$T$	$\alpha$	Zona C		
		$-DF_{\alpha/2}$	$-t_{T-2;\alpha/2}$	$t_{T-2;\alpha/2}$
25	$\alpha/2 = 0.025$	-2.97	-2.07	2.07
	$\alpha/2 = 0.05$	-2.61	-1.71	1.71
50	$\alpha/2 = 0.025$	-2.89	-2.01	2.01
	$\alpha/2 = 0.05$	-2.56	-1.68	1.68
100	$\alpha/2 = 0.025$	-2.86	-1.98	1.98
	$\alpha/2 = 0.05$	-2.54	-1.66	1.66
250	$\alpha/2 = 0.025$	-2.54	-1.66	1.66
	$\alpha/2 = 0.05$	-2.53	-1.65	1.65
$\infty$	$\alpha/2 = 0.025$	-2.83	-1.97	1.97
	$\alpha/2 = 0.05$	-2.52	-1.65	1.65

$T$	$\alpha$	Zona D		
		$DF_{\alpha/2}$	$-t_{T-2;\alpha/2}$	$t_{T-2;\alpha/2}$
25	$\alpha/2 = 0.025$	2.97	-2.07	2.07
	$\alpha/2 = 0.05$	2.61	-1.71	1.71
50	$\alpha/2 = 0.025$	2.89	-2.01	2.01
	$\alpha/2 = 0.05$	2.56	-1.68	1.68
100	$\alpha/2 = 0.025$	2.86	-1.98	1.98
	$\alpha/2 = 0.05$	2.54	-1.66	1.66
250	$\alpha/2 = 0.025$	2.54	-1.66	1.66
	$\alpha/2 = 0.05$	2.53	-1.65	1.65
$\infty$	$\alpha/2 = 0.025$	2.83	-1.97	1.97
	$\alpha/2 = 0.05$	2.52	-1.65	1.65

Fuente: Elaboración propia

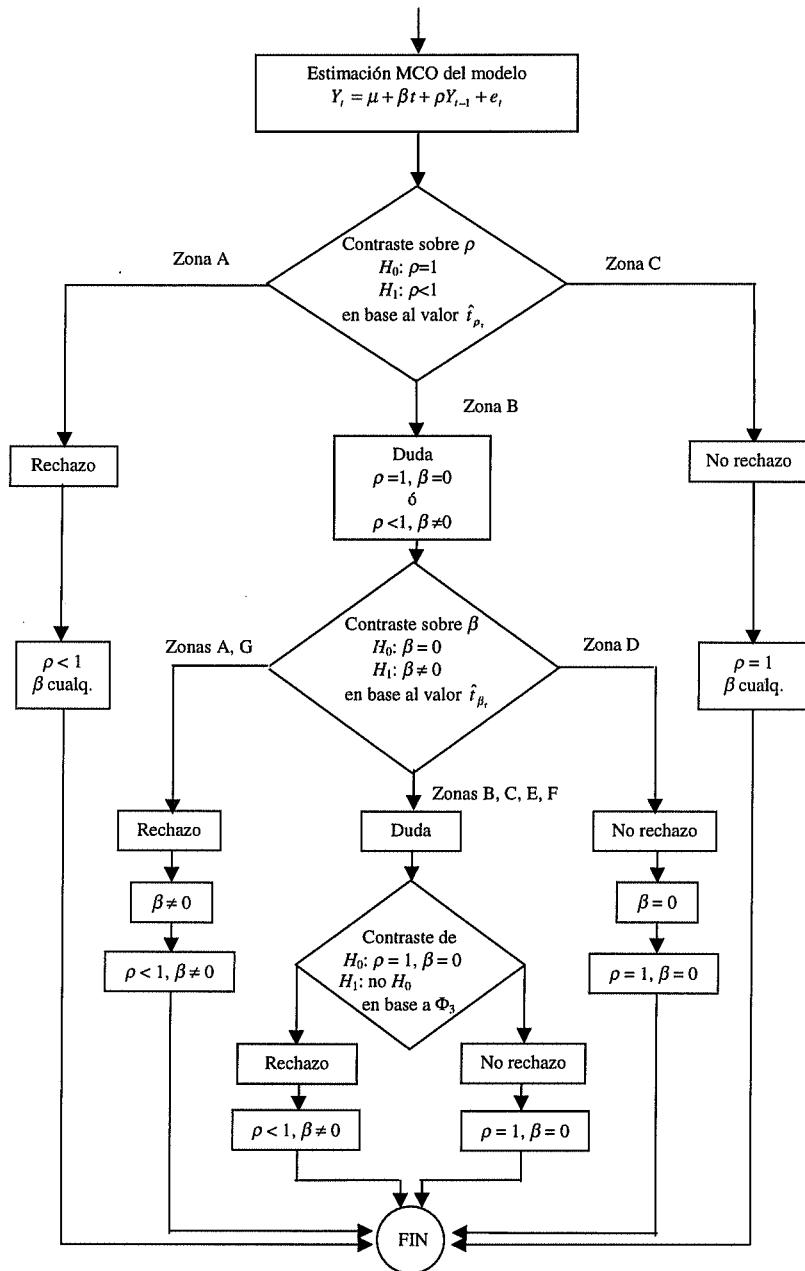
$T$ : tamaño muestral.

$\alpha$ : nivel de significación

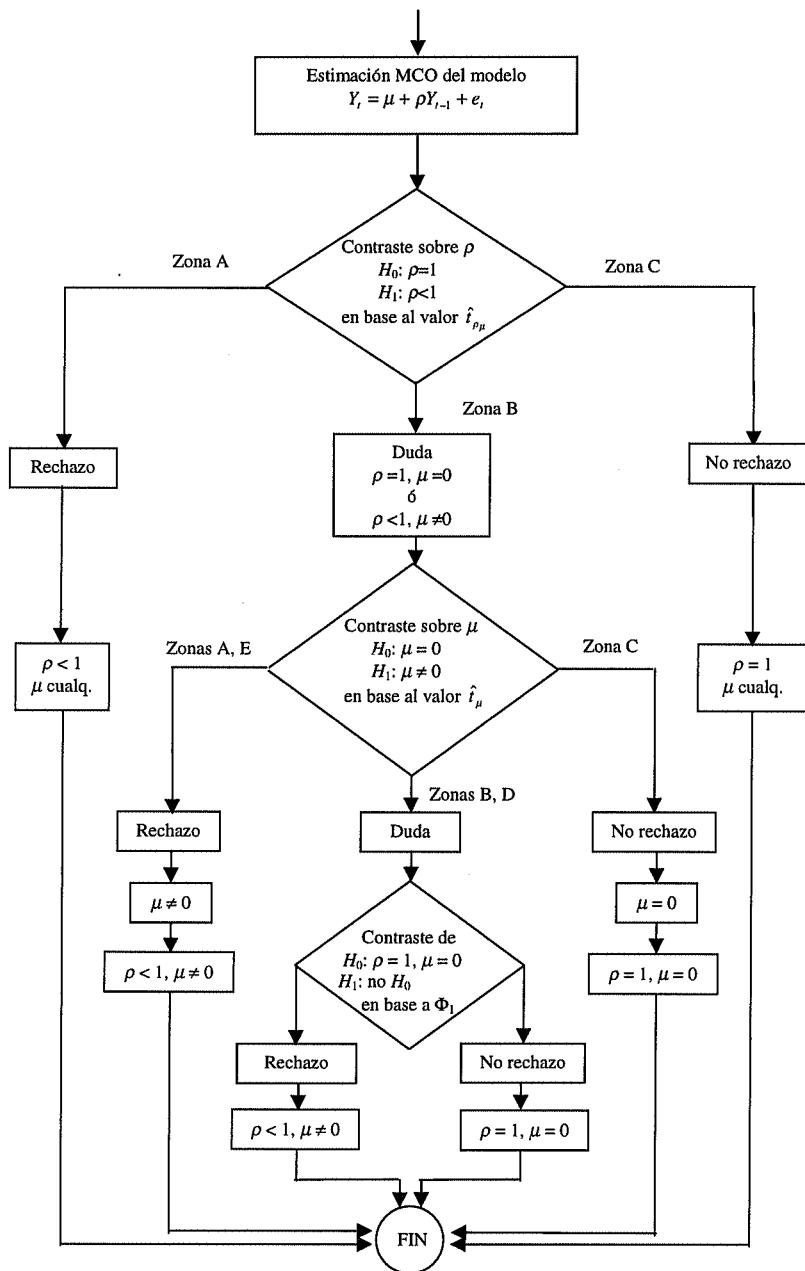
$DF$ : valores críticos de Dickey-Fuller.

$t_{T-2}$ : valores críticos de la  $t$  de Student

**Diagrama A: ESTRATEGIA DE CONTRASTE**



**Diagrama B: ESTRATEGIA DE CONTRASTE**



## ENGLISH SUMMARY

### TESTING FOR UNIT ROOTS IN TIME SERIES. AN ANALYSIS WITH NON-SIMILAR TESTS

JOSÉ ÁNGEL ROLDÁN CASAS

RAFAELA DIOS PALOMARES

Universidad de Córdoba\*

*The finite and limiting sample distributions of the t and F statistics are analyzed in the context of a first-order autoregressive model with an unknown intercept and/or a linear trend, where the initial value of the variable is a known constant. The sensitivity of the distributions to the values of the nuisance parameters of the models considered is examined. We propose a strategy for testing the unit root hypothesis in these models based on the no-similarity. It is an unusual one-sided test since it involves two distributions: the Student's t distribution and the distribution tabulated by Dickey and Fuller under the null hypothesis of unit root. In some cases, the critical values tabulated by Dickey and Fuller do not allow us to reject the hypothesis of a unit root, while we can reject it if we use Student's t. To solve these situations we test for the significance of the intercept or the linear trend and if the doubt continues we use F tests proposed by Dickey and Fuller (1981). Finally, in order to demonstrate the relevance of no-similarity Monte Carlo simulations are used to show that the testing strategy is more powerful than the one-sided test considered by Dickey and Fuller at stable alternatives.*

**Keywords:** unit root, stochastic trend, non-similar test, Dickey-Fuller test, Monte Carlo experiment

**AMS Classification (MSC 2000):** 62M10

---

\* Universidad de Córdoba. Dpto. de Estadística e Investigación Operativa. E.S.I.A y de Montes.  
Apartado 3048. 14080 Córdoba.

-Received March 1998.  
-Accepted March 2000.

A problem arising in many time-series applications is the question of whether a series should be differenced. This is equivalent to asking if the time series has a unit root.

Currently, practitioners may decide to difference a time series on the basis of a visual inspection of the autocorrelation function, or a statistic test.

Dickey and Fuller (1979, 1981) have proposed some test statistics for the unit root hypothesis for an observed time series. They derive the finite and limiting distributions of test statistics for a unit root when the estimated model is a random walk, a random walk with shift in mean, and a random walk with shift in mean and a linear time trend. The distribution of Dickey and Fuller tests relied on the innovation process  $e_t$  being white noise. These authors have restricted their work to autoregressive (AR) series. More precisely, in 1979 they presented a class of test statistics, known as Dickey-Fuller (DF) statistics, assuming that DGP is a pure AR(1) process. In 1981 they extended the DF test to an AR( $p$ ) process which is called ‘augmented’ Dickey-Fuller (ADF) test.

On the other hand, Nankervis and Savin (1985, 1987) showed that the statistics proposed by Dickey and Fuller yield non-similar tests of the unit root hypothesis. No similarity implies that the distribution of a test statistic is affected by the value, under the null, of a nuisance parameter. If the nuisance parameter is unknown it is possible for a mistake to be in either the rejection or non-rejection. Nankervis and Savin’s non-similar tests of the random walk hypothesis involve two distributions and are substantially more powerful at most alternatives of interest than the similar tests considered by Dickey and Fuller (1981).

In this paper we assume that the data is generated by the first order autoregressive processes

$$\begin{aligned} (1) \quad Y_t &= \rho Y_{t-1} + e_t & t = 1, 2, \dots \\ (2) \quad Y_t &= \mu + \rho Y_{t-1} + e_t & t = 1, 2, \dots \\ (3) \quad Y_t &= \mu + \beta t + \rho Y_{t-1} + e_t & t = 1, 2, \dots \end{aligned}$$

where  $Y_0 = 0$ ,  $\mu$ ,  $\beta$  y  $\rho$  are unknown real numbers and  $\{e_t\}$  is a sequence of independent normal random variables with mean zero and variance  $\sigma^2$ . We study the distributions of the  $t$  and  $F$  statistics for models considered, taking into account the values of the nuisance parameters. In the context of models (2) and (3), Dickey and Fuller (1979) obtained the limiting distribution of the  $t$  statistics of  $\rho$  for  $\rho = 1$ , under the assumptions that  $\mu = 0$  and  $\beta = 0$ , respectively. These distributions are non-standard and Dickey and Fuller obtained the empirical quantiles of the limiting and finite sample distribution by Monte Carlo methods (Fuller 1996, p. 6).

On the other hand, for a fixed sample size, the distributions of the  $t$  statistics for both the exogenous and lagged dependent variables converge to Student's  $t$  as the drift coefficient and/or the coefficient of the exogenous variable increases in absolute value (Nankervis y Savim, 1985, 1987).

Therefore, the  $t$  statistics mentioned above yield non-similar tests of the unit root hypothesis since their distributions are influenced by the values of the nuisance parameters  $\mu$  and  $\beta$ .

From the previous discussion we propose a testing strategy to test the unit root hypothesis in the context of models (2) and (3). This testing strategy takes into account the no-similarity and is based on the two distributions test introduced by Nankervis and Savin (1985), since it involves the two distributions of the  $t$  statistics of  $\rho$  under the null: the Student's  $t$  distribution and the distribution tabulated by Dickey and Fuller. In some cases, the critical values tabulated by Dickey and Fuller do not allow us to reject the hypothesis of a unit root, while we may reject it if we use Student's  $t$ . To solve these situations we test for the significance of the intercept or the linear trend, and if the doubt continues we use  $F$  tests proposed by Dickey and Fuller (1981).

Finally, in order to demonstrate the relevance of no-similarity, Monte Carlo simulations are used to show that the testing strategy proposed is more powerful than the one-sided test considered by Dickey and Fuller at stable alternatives.



# **Investigació Operativa**



## DEL POLIEDRO DEL AGENTE VIAJERO GRÁFICO AL DE RUTAS DE VEHÍCULOS CON DEMANDA COMPARTIDA

CARMEN MARTÍNEZ

ENRIQUE MOTA

Universitat de València\*

*En este trabajo abordamos el estudio del poliedro asociado al Problema de Rutas de Vehículos con Demanda Compartida, problema de distribución que surge cuando hay que repartir mercancías a un conjunto de clientes utilizando una flota fija de vehículos de capacidad limitada. El objetivo es diseñar las rutas de forma que se minimice la distancia total recorrida. Se diferencia de otros problemas más conocidos de rutas con capacidades en que se permite abastecer la demanda de cada cliente utilizando más de un vehículo. Presentamos resultados poliédricos que se obtienen como extensión de otros previamente conocidos para el Problema del Agente Viajero Gráfico.*

**From the graphical traveling salesman polyhedron to the one associated with the split delivery vehicle routing problem**

**Palabras clave:** Rutas de vehículos, demanda compartida, poliedro, facetas

**Clasificación AMS (MSC 2000):** 90C10, 90C27, 90C35

---

\* Departament d'Estadística i Investigació Operativa. Universitat de València. Dr. Moliner, 50. Burjassot, 46100. España. E-mail: M.Carmen.Martinez@uv.es

-Recibido en octubre de 1999.

-Aceptado en octubre de 2000.

## 1. INTRODUCCIÓN

El *Problema de Rutas de Vehículos con Demanda Compartida*, en adelante *SDVRP*, nos permite modelizar aquellos problemas de distribución en los que una empresa debe satisfacer, de la forma más económica posible, la demanda de una cartera de clientes, utilizando un número fijo de vehículos de capacidad limitada. El objetivo es diseñar una ruta para cada vehículo de manera que se abastezcan todas las demandas, ningún vehículo sirva más de lo que indica su capacidad y se minimice la distancia total recorrida por el conjunto de rutas. Así planteado, el problema responde a la versión clásica del *Problema de Rutas de Vehículos con Capacidades, CVRP* (ver Bodin *et al.*, 1983 y Christofides, 1985), sin embargo, en el *CVRP* no se contempla la posibilidad de servir la demanda de un cliente utilizando varios vehículos, algo admisible en el *SDVRP*, de ahí la calificación «*demandas compartidas*».

El *SDVRP* fue estudiado por primera vez en Dror y Trudeau (1989), descubriendose entonces su potencial para aportar soluciones más económicas para los problemas de distribución, pues, no únicamente reduce el número de vehículos necesario para abastecer la demanda total, sino que además, en ocasiones esto va acompañado de importantes ahorros en la distancia total recorrida. Otro problema resuelto al trabajar con el *SDVRP* es el de que siempre se conoce el número mínimo de vehículos necesario para abastecer toda la demanda, pues basta con que calculemos el cociente entre la demanda total y la capacidad del vehículo. En el caso del *CVRP* determinar el mínimo número de vehículos necesario implica resolver un problema *NP – duro*, el problema de *Bin-Packing* asociado a las demandas de los clientes y a cajas de capacidad igual a la de los vehículos, Garey y Johnson (1979). Sin embargo, aunque el *SDVRP* es una relajación del *CVRP*, y por tanto puede parecer más sencillo de resolver, es también un problema *NP – duro* (ver Dror y Trudeau, 1990), de ahí el interés que despierta el estudio del poliedro de soluciones asociado. Cabe esperar, como ocurre con otros problemas de rutas, como el *Problema del Agente Viajero* (Padberg y Rinaldi, 1991) o el mismo *CVRP* (Cornuéjols y Harche, 1993, y Augerat *et al.*, 1995), que el conocimiento de la estructura poliédrica del conjunto de soluciones asociado ayudará a resolver de forma exacta instancias del problema cada vez más grandes. Ampliar el estudio poliédrico presentado en Belenguer *et al.* (2000) y Martínez *et al.* (2000) es el objetivo principal que se pretende conseguir con el trabajo que aquí presentamos. Los dos trabajos anteriores junto con la revisión bibliográfica llevada a cabo en Belenguer *et al.* (2000), y el algoritmo tabú diseñado por Belenguer y Martínez (2000) constituyen un buen resumen del estado del arte en el problema.

Hemos organizado el artículo de la siguiente forma: en la Sección 2 introducimos la notación que vamos a utilizar y definimos formalmente el problema. En la Sección 3 revisamos las propiedades que deben cumplir las desigualdades válidas para el poliedro asociado al *Problema de Rutas de Vehículos con Demanda Compartida*. Estudiamos las desigualdades *path*, *wheelbarrow* y *bicycle*, introducidas en Cornuéjols *et al.* (1985)

para el poliedro del *Agente Viajero Gráfico* (*GTSP*, versión del *TSP* en la que se permite visitar las ciudades más de una vez), y damos condiciones suficientes bajo las cuales dichas desigualdades inducen facetas del poliedro cuyo estudio nos ocupa. Terminamos el artículo con una sección dedicada a consideraciones finales.

## 2. NOTACIÓN Y DEFINICIÓN DEL PROBLEMA

Sea  $G = (V, E)$  un grafo completo y no dirigido, en donde el conjunto de vértices,  $V$ , representa un conjunto de clientes,  $\{1, \dots, n\}$ , con demandas enteras conocidas,  $d_i > 0$ , y un vértice distinguido,  $\{0\}$ , al que denominaremos depósito, en el que hay localizada una flota de  $K$  vehículos idénticos de capacidad  $Q \geq \max\{d_i, i \in V \setminus \{0\}\}$ . Cada arista  $e = (i, j)$  del conjunto de aristas,  $E$ , representa el camino entre sus extremos.

Dado  $S \subseteq V$ ,  $\bar{S}$  denota el conjunto de vértices en  $V \setminus S$  y  $d(S)$  la suma de las demandas de los clientes de  $S$ .  $E(S)$  representa el conjunto de aristas con los dos extremos en  $S$  y  $\delta(S)$  el conjunto de aristas con un único extremo en  $S$ . Dados  $S$  y  $S' \subseteq V$ ,  $(S, S')$  denota el conjunto de aristas con un extremo en  $S$  y el otro en  $S'$ .

Dado un subconjunto  $T$  de aristas, que puede contener en general más de una copia de una misma arista,  $V(T)$  representa el conjunto de vértices que son extremos de aristas de  $T$ , y  $G(T) = (V(T), T)$ , el subgrafo de  $G$  inducido por las aristas de  $T$ . Definimos el vector de incidencia,  $x^T$ , asociado a  $T$  como un vector real indexado en  $E$  cuyas componentes,  $x_e$  (o  $x_{ij}$ ), indican el número de veces que la arista  $e$  aparece en  $T$ .  $x(T)$  indicará  $\sum_{e \in T} x_e$ .

Un *tour* es un camino cerrado que puede utilizar tanto aristas como vértices repetidos. Dado un subconjunto de vértices, un *ciclo Hamiltoniano* es un tour que visita exactamente una vez cada vértice del subconjunto. Salvo que se indique lo contrario supondremos que todos los vectores,  $x$ , son vectores columna.  $x'$ , denotará el vector traspuesto. El cardinal de un conjunto  $W$  lo denotamos como  $|W|$ .

Dado  $P$  un poliedro en  $\mathbb{R}^{|E|}$ , una desigualdad  $a^t x \geq a_0$  definida en  $\mathbb{R}^{|E|}$  se dice que es válida para  $P$  si la cumple cualquier punto de  $P$ ; se dice que es soporte si  $\{x \in \mathbb{R}^{|E|} \mid a^t x = a_0, x \in P\}$  es no vacío. Decimos que la desigualdad induce una faceta de  $P$  si es soporte y  $\{x \in \mathbb{R}^{|E|} \mid a^t x = a_0, x \in P\}$  es una faceta. Tanto los resultados de teoría poliédrica como las técnicas de demostración que vamos a utilizar pueden consultarse en cualquiera de los textos clásicos, Nemhauser y Wolsey (1988) o Pulleyblank (1983).

De ahora en adelante supondremos que el número de vehículos disponible,  $K$ , es suficiente como para abastecer la demanda de todos los clientes. En el contexto de demanda compartida basta con garantizar que  $K \geq \lceil \frac{d(V)}{Q} \rceil$ , en donde  $\lceil \frac{d(V)}{Q} \rceil$  representa el menor entero mayor que  $\frac{d(V)}{Q}$ . La notación que acabamos de establecer nos permite dar la siguiente definición:

**Definición 1.** Un conjunto de aristas,  $T$ , no necesariamente distintas, se dice que es una solución del SDVRP si es posible establecer una partición de  $T$  en  $K$  subconjuntos,  $T_h$ , de manera que, el grafo inducido por cada  $T_h$ ,  $G(T_h) = (V(T_h), T_h)$ , es un tour que contiene al depósito y además es posible asociar con cada vértice,  $i$ , y con cada ruta,  $h$ , un valor,  $d_i^h$ , que indica la parte de la demanda de  $i$  servida por el vehículo  $h$ , de forma que:

- cada vértice en  $V \setminus \{0\}$  está asignado a al menos una ruta,
- $V = \bigcup_{h=1}^K V(T_h)$ ,
- que el cliente  $i$  esté asignado a la ruta  $h$  no significa que  $d_i^h > 0$  pero, si  $i \notin V(T_h)$ , entonces  $d_i^h = 0$ ,
- $\sum_{h=1}^K d_i^h = d_i \quad \forall i \in \{1, 2, \dots, n\}$  y,
- $\sum_{i \in V(T_h)} d_i^h \leq Q \quad \forall h \in \{1, 2, \dots, K\}$ .

En Belenguer *et al.* (2000) se demuestra que la envoltura convexa de los vectores de incidencia asociados a los subconjuntos de aristas,  $T$ , que cumplen la definición anterior, es un poliedro no acotado de dimensión completa. En adelante nos referiremos a él como  $P_{SDVRP}$ . En el mismo trabajo se prueba que los vectores de incidencia,  $x$ , de las soluciones del SDVRP cumplen:

- (1)  $x(\delta(0)) \geq 2K \quad$  y par
- (2)  $x(\delta(i)) \geq 2 \quad$  y par  $\quad \forall i \in V \setminus \{0\}$
- (3)  $x(\delta(S)) \geq 2 \lceil \frac{d(S)}{Q} \rceil \quad \forall S \subseteq V \setminus \{0\}, \quad 2 \leq |S| \leq n - 1$
- (4)  $x_{ij} \geq 0, \quad x_{ij} \quad$  entero  $\quad \forall (i, j) \in E$

en donde  $x_{ij}$  representa el número de veces que se utiliza la arista  $(i, j)$  en la solución. Por último, demuestran que las restricciones de grado, (1) y (2), las de capacidad, (3), y las de no negatividad, (4), inducen facetas de  $P_{SDVRP}$ .

Modificando ligeramente el valor de los términos de la derecha de las desigualdades anteriores se consiguen formulaciones de problemas de rutas más conocidos, como el *GTSP*, que hemos mencionado antes, o la versión gráfica del *CVRP*, conocida como *GVRP*, en donde sigue sin permitirse servir la demanda de un cliente utilizando más de un vehículo, pero sí se admite la posibilidad de visitarlo más de una vez. Véanse respectivamente los trabajos de Cornuéjols *et al.* (1985), y Cornuéjols y Harche (1993), en donde se presentan interesantes resultados, tanto teóricos como prácticos, que se han obtenido considerando como base para formular los problemas las desigualdades anteriores.

Es sencillo comprobar que las restricciones (1) – (4) proporcionan una relajación del SDVRP, ya que hay vectores que satisfacen todas las desigualdades y no corresponden al vector de incidencia de ninguna solución posible del problema, véase Belenguer *et al.*

(2000). Sin embargo, esta formulación ha proporcionado también buenos resultados en el campo del *SDVRP*, tanto desde el punto de vista teórico como práctico. Belenguer *et al.* (2000) resuelven óptimamente instancias de hasta 50 clientes, y Martínez *et al.* (2000) consiguen encontrar la descripción lineal completa del poliedro asociado a una instancia de 5 clientes y 2 vehículos, la única que se conoce por el momento para un problema de estas características. Creemos que estos son motivos suficientes para continuar la investigación en esta línea.

A continuación describimos brevemente la técnica que utilizaremos para demostrar algunos de los resultados que vamos a presentar, es conocida como método indirecto. Como el  $P_{SDVRP}$  es un poliedro no acotado de dimensión completa, sabemos que no hay ninguna ecuación implícita en su descripción lineal, esto nos permite afirmar (ver Nemhauser y Wolsey, 1988) lo siguiente: *Para cada faceta del  $P_{SDVRP}$  hay una única (salvo producto por un escalar no negativo) desigualdad válida que la induce. Además, para demostrar que una desigualdad  $a^t x \geq a_0$  define una faceta basta con demostrar que si  $f^t x \geq f_0$  es una faceta del  $P_{SDVRP}$  que cumple:*

$$\{x \in P_{SDVRP} \mid a^t x = a_0\} \subseteq \{x \in P_{SDVRP} \mid f^t x = f_0\},$$

entonces, existe un escalar  $\lambda$  no negativo, que cumple  $f_e = \lambda a_e$ ,  $\forall e \in E$  y,  $f_0 = \lambda a_0$ .

### 3. DEL $P_{GTSP}$ AL $P_{SDVRP}$

En esta sección extendemos al poliedro del *Problema de Rutas de Vehículos con Demanda Compartida* algunas desigualdades que se sabe son facetas del  $P_{GTSP}$ . En las demostraciones de los resultados que vamos a presentar se utilizan, de forma implícita, las propiedades de las desigualdades válidas para el  $P_{SDVRP}$  que resumimos a continuación. Tales propiedades pueden encontrarse, junto con sus respectivas demostraciones, en Martínez (1995).

- Las desigualdades válidas para el  $P_{SDVRP}$  son de la forma:

$$\sum_{e \in E} a_e x_e \geq a_0, \text{ con } a_e \geq 0, \quad \forall e \in E.$$

- Las únicas restricciones,  $a^t x \geq a_0$ , distintas de  $x_e \geq 0$ ,  $e \in E$ , que inducen facetas del  $P_{SDVRP}$ , cumplen  $a_0 > 0$ .
- Si  $a^t x \geq a_0$  induce una faceta de  $P_{SDVRP}$ , entonces, o es una desigualdad trivial, i.e.  $x_e \geq 0$ ,  $\forall e \in E$ , o sus coeficientes satisfacen la desigualdad triangular.
- Si  $a^t x \geq a_0$  es una desigualdad válida para el  $P_{SDVRP}$  y sus coeficientes cumplen la desigualdad triangular entonces, para cualquier arista  $e \in E$  existe una solución  $T$  del *SDVRP*, que utiliza la arista  $e$  y cuyo vector de incidencia cumple  $a^t x = a_0$ .

De las definiciones del *GTSP*, Cornuéjols *et al.* (1985), el *GVRP*, Cornuéjols y Harche (1993) y la del *SDVRP*, se deduce fácilmente que la relación existente entre sus poliedros es  $P_{GVRP} \subseteq P_{SDVRP} \subseteq P_{GTSP}$ , por lo tanto, cualquier desigualdad válida para el *GTSP* también lo será para el *SDVRP*. Es evidente que esto no implica que se herede también la propiedad de ser facetas, de hecho, hasta ahora no se conoce ningún resultado general que permita establecer una relación entre las facetas de dichos poliedros. Solamente se conocen resultados parciales, como por ejemplo, que bajo distintas condiciones, las desigualdades (1)-(4), adaptadas convenientemente a cada problema, inducen facetas en los tres casos.

Las desigualdades que estudiamos aquí están asociadas a estructuras más complejas. Se trata de las definidas a partir de las *configuraciones path*, *wheelbarrow* y *bicycle* introducidas en Cornuéjols *et al.* (1985) para el poliedro de soluciones del *GTSP*. Más tarde, Cornuéjols y Harche (1993) demuestran que bajo determinadas condiciones las restricciones asociadas también inducen facetas del poliedro del *GVRP*. La relación existente entre estos poliedros y el  $P_{SDVRP}$  invita a pensar que estas desigualdades también inducirán, como veremos, facetas del mismo. Tanto la notación como las definiciones que damos a continuación han sido adoptadas del trabajo de Cornuéjols *et al.* (1985). Notar, no obstante, que nosotros trabajamos sobre un grafo completo.

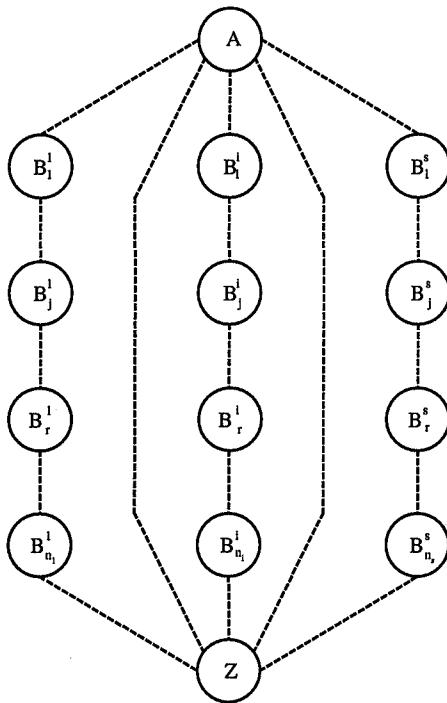
Dado  $G = (V, E)$  un grafo completo no dirigido, a la estructura de la figura 1 se la denomina *configuración path*, en donde,  $s$  es un entero,  $s \geq 3$  e impar y  $\{A, Z, B_j^i, i = 1, \dots, s, j = 1, \dots, n_i\}$ , con  $n_i \in \mathbb{N}$ ,  $n_i \geq 2$ ,  $\forall i = 1, 2, \dots, s$ , una partición del conjunto de vértices,  $V$ , en subconjuntos no vacíos.

Se define la *desigualdad*,  $a^t x \geq a_0$ , asociada a una *configuración path* como aquélla cuyos coeficientes son:

$$a_e = \begin{cases} 1 & \text{para } e \in (A, Z) \\ \frac{|j-p|}{n_i-1} & \text{para } e \in (B_j^i, B_p^i), j \neq p \\ \max \left\{ \frac{p}{n_r-1} - \frac{j-2}{n_i-1}, \frac{j}{n_i-1} - \frac{p-2}{n_r-1} \right\} & \text{para } e \in (B_j^i, B_p^r), i \neq r \\ 0 & \text{en otro caso} \end{cases}$$

y el término independiente es:

$$a_0 = 1 + \sum_{i=1}^s \frac{n_i+1}{n_i-1}.$$



**Figura 1.** Configuración *path*.

Si  $Z = \emptyset$ , la configuración resultante se denomina *configuración wheelbarrow*. Si tanto  $A$  como  $Z$  son vacíos la configuración se denomina *configuración bicycle*. Análogamente se definen las desigualdades asociadas.

Por convenio,  $A = B_0^i$ ,  $Z = B_{n_i+1}^i$ ,  $i = 1, \dots, s$  y  $B_{n_{s+1}}^{s+1} = B_{n_1}^1$ . Esperamos que esta notación aporte claridad a las demostraciones. Por el mismo motivo,  $P_i = \bigcup_{j=1}^{n_i} B_j^i$ ,  $i = 1, \dots, s$ , representará el conjunto de vértices de cada uno de los  $s$  caminos que hay en la configuración uniendo  $A$  con  $Z$ .

Evidentemente, las desigualdades *path* son válidas para el  $PSDVRP$  ya que lo son para el  $PGTSP$ , Cornuéjols *et al.* (1985). La cuestión que nos planteamos aquí es averiguar si también definen facetas. En ese sentido, es evidente que bajo las condiciones establecidas por Cornuéjols y Harche (1993) para garantizar que dichas restricciones inducen facetas del  $PGVRP$ , podemos afirmar que también lo son del  $PSDVRP$ , ya que cualquier solución posible del  $GVRP$  lo es del  $SDVRP$ .

Sin embargo, tales condiciones son difíciles de comprobar, pues implican averiguar el número de vehículos necesario para garantizar la existencia de ciertas particiones

de los clientes, particiones que a su vez deben cumplir una serie de condiciones muy concretas dirigidas a garantizar la existencia de ciertos tipos de soluciones. De hecho, para averiguar dicho número de vehículos es necesario resolver previamente distintos problemas de *Bin-Packing*.

La idea en este trabajo es explotar la ventaja que supone compartir la demanda para extender al  $P_{SDVRP}$  los resultados propuestos en Cornuéjols *et al.* (1985) para el  $P_{GTSP}$ . Se trata de derivar condiciones, sencillas de comprobar, que sigan garantizando que las desigualdades *path* inducen facetas del  $P_{SDVRP}$ , sin necesidad de recurrir a la resolución de problemas tipo *Bin-Packing* para poder verificarlas.

**Teorema 1.** *Dados K vehículos de capacidad Q, si 0 ∈ A (o si 0 ∈ Z), y se cumplen las condiciones siguientes:*

$$i.- \quad K \geq \frac{s-1}{2},$$

$$ii.- \quad Q \geq d(P_s) + d(P_{s-1}) + d(P_{s-2}), \text{ siendo } d(P_1) \leq d(P_2) \leq \dots \leq d(P_s), \text{ y}$$

$$iii.- \quad Q \geq 2 \left\lceil \frac{d(Z) + \sum_{i=1}^s d(P_i)}{s-1} \right\rceil$$

entonces, las desigualdades «*path*» inducen facetas de  $P_{SDVRP}$ .

#### Demostración

Podemos suponer, sin pérdida de generalidad, que el depósito está en A. Observar que como A y Z juegan papeles equivalentes en las configuraciones *path*, cualquier resultado obtenido para uno de ellos es válido para el otro.

Ya sabemos que cualquier desigualdad,  $a'x \geq a_0$ , asociada a una configuración *path* es válida para el  $P_{SDVRP}$ . Demostraremos que es soporte comprobando que las soluciones que vamos a construir pertenecen a la cara inducida por la desigualdad. Obsérvese también que añadiendo dos copias extras de cualquiera de las aristas con coeficiente estrictamente positivo en  $a$ , obtenemos una solución que cumple la desigualdad con mayor estricto, lo cuál demuestra que  $a'x \geq a_0$  es una desigualdad propia.

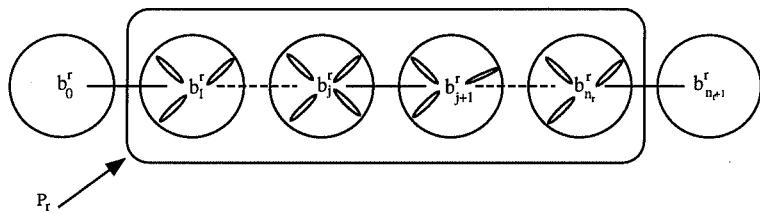
En primer lugar asociaremos a cada uno de los subconjuntos de la partición que define la configuración *path* un vértice que utilizaremos como su *representante*. El representante de A será el depósito, al que nos referiremos como  $b_0^i = 0$ ,  $i = 1, \dots, s$ . Para cada  $B_j^i$ , con  $i = 1, \dots, s$ , y  $j = 1, \dots, n_i$ , elegiremos como representante un  $b_j^i \in B_j^i$ , y finalmente, tomaremos  $z = b_{n_i+1}^i$ ,  $i = 1, \dots, s$  como representante de Z.

A continuación definiremos las soluciones que nos van a permitir demostrar el resultado utilizando la técnica conocida como método indirecto. Para cada  $i = 1, \dots, s$  construi-

mos la solución  $T^i$ , que diremos asociada al camino  $P_i$ , tomando:

- $2K$  copias de cada arista  $(0, v)$ , con  $v \in A$  i.e.,  $x_{0v} = 2K$ . Esto garantiza que todos los vehículos visitan a todos los vértices de  $A$ , ya que es posible asignar a cada uno de ellos dos de las aristas incidentes con  $v$ .
- Añadir  $\frac{s-1}{2}$  copias de cada una de las aristas,  $e$ , de un ciclo Hamiltoniano,  $C$  en  $Z$ . Garantizamos así que cada vértice de  $Z$  es visitado por  $\frac{s-1}{2}$  vehículos distintos, basta con asignar una arista a cada uno de ellos.
- Completamos la solución tomando  $x_e = 1$ ,  $\forall e = (b_j^r, b_{j+1}^r)$ ,  $j = 0, \dots, n_r$  y  $r = 1, \dots, s$ ,  $r \neq i$ . De este modo tenemos la seguridad de que para cada  $P_r$  ( $r \neq i$ ) hay un camino desde el depósito  $b_0^r$  hasta  $z$  que solamente utiliza aristas consecutivas de  $E(P_r)$ .
- Añadiendo  $x_{b_j^r v} = 2$ ,  $\forall v \in B_j^r$ ,  $j = 1, \dots, n_r$ ,  $r = 1, \dots, s$ , estamos seguros de que visitamos todos los vértices de  $B_j^r$  utilizando el único vehículo que entrará en el subconjunto.

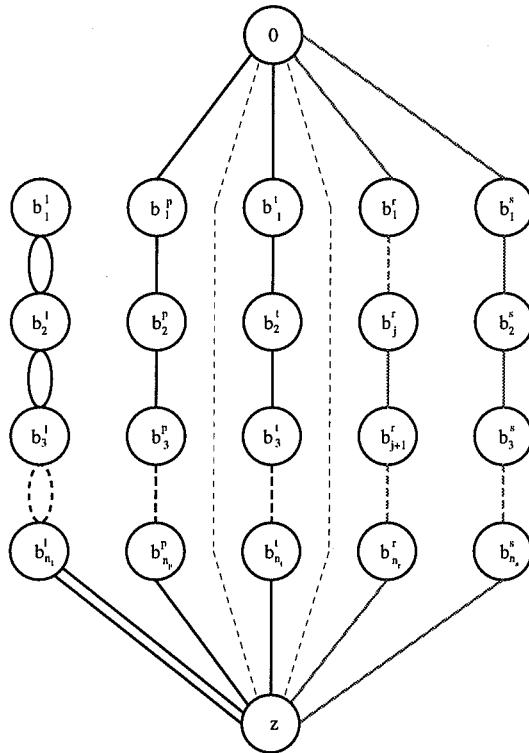
En la figura 2 se ilustra parcialmente la estructura de la solución que estamos construyendo.



**Figura 2.** Vista parcial de  $T^i$  incluyendo el camino  $P_r$ .

- Notar que todos los caminos tienen como vértices extremos a  $0$  y a  $z$ , esto nos permite definir  $\frac{s-1}{2} - 1$  tours distintos emparejando los caminos  $P_r$ ,  $r = 1, \dots, s$ ,  $r \neq i$ , de dos en dos. La hipótesis  $K \geq \frac{s-1}{2}$  garantiza que es posible asignar un vehículo distinto a cada uno de estos tours. Además, cada uno de estos vehículos puede servir completamente la demanda de los clientes que visita a lo largo de los dos caminos que lo forman sin exceder su capacidad; recordar que por hipótesis,  $Q \geq d(P_s) + d(P_{s-1}) + d(P_{s-2})$ , lo cual significa que la demanda de dos, e incluso tres,  $P_r$  cualesquiera puede abastecerse completamente utilizando un único vehículo.
- Para terminar, tomamos 2 copias de cada una de las aristas  $(b_j^i, b_{j+1}^i)$ ,  $j = 1, \dots, n_i$ , que junto con el tour que nos quedaba por asignar en el apartado anterior proporciona una nueva ruta que asignamos al vehículo que seguro todavía nos queda, recordar que por hipótesis  $K \geq \frac{s-1}{2}$ . Este vehículo abastecerá completamente la demanda de los tres caminos que lo forman más parte de la demanda de  $A$  y de la de  $Z$  si fuese necesario. Notar que la arista  $(b_0^i, b_1^i)$  no se utiliza.

En total hemos construido  $\frac{s-1}{2}$  rutas en  $V \setminus A$ . Obsérvese que entre todos los vehículos asignados a estas rutas pueden servirse  $Q \frac{s-1}{2}$  unidades de demanda, lo que garantiza, gracias a la hipótesis,  $Q \geq 2 \lceil \frac{d(Z) + \sum_{i=1}^s d(P_i)}{s-1} \rceil$ , que entre todos ellos es posible abastecer toda la demanda de  $Z$ , pues todos estos vehículos entran en  $Z$  y visitan a todos sus clientes. Además, el hecho de que los  $K$  vehículos disponibles visiten a todos los clientes de  $A$  garantiza la existencia de una asignación adecuada de las demandas a los vehículos de forma que se cumplan todas las condiciones establecidas en la definición 1. Por lo tanto, la solución que acabamos de construir es una solución posible del *SDVRP*, y como el procedimiento constructivo que hemos utilizado es independiente del  $P_i$  elegido, podemos obtener hasta  $s$  soluciones distintas. La figura 3 corresponde a  $T^1$ .



**Figura 3.** Solución asociada a  $P_1$ .

Nótese que, precisamente, el hecho de permitir atender la demanda de un cliente utilizando varios vehículos, es lo que garantiza que las soluciones construidas a lo largo de todo el artículo sean soluciones posibles para el *SDVRP*, pero no necesariamente del

*GVRP.* Pensemos por ejemplo en el caso en que  $Z$  es un único cliente. Las soluciones consideradas solamente serán soluciones posibles del *GVRP* cuando la capacidad residual de, al menos uno de los vehículos que visitan a  $Z$ , es suficiente para servirlo completamente. Además, hemos de tener en cuenta que dicho vehículo debe servir también la demanda de los caminos  $P_i$  asignados a su ruta; sin duda alguna ésta es una condición más fuerte que la que nosotros hemos establecido.

Veamos a continuación que las soluciones que hemos construido pertenecen a la cara inducida por la desigualdad *path*. Obsérvese que en cada  $P_r \neq P_i$ , utilizamos  $n_r + 1$  aristas de coeficiente  $\frac{1}{n_r - 1}$  y en cada  $P_i$ ,  $2n_i$  aristas de coeficiente  $\frac{1}{n_i - 1}$ , luego:

$$\begin{aligned} a^t x^{T^i} &= \sum_{r=1, r \neq i}^s \frac{n_r + 1}{n_r - 1} + \frac{2n_i}{n_i - 1} = \sum_{r=1, r \neq i}^s \frac{n_r + 1}{n_r - 1} + \frac{n_i + 1 + n_i - 1}{n_i - 1} = \\ &1 + \sum_{r=1}^s \frac{n_r + 1}{n_r - 1} = a_0. \end{aligned}$$

Supongamos que existe una desigualdad  $f^t x \geq a_0$  que induce una faceta de  $P_{SDVRP}$  cumpliendo,  $\{x \in P_{SDVRP} \mid a^t x = a_0\} \subseteq \{x \in P_{SDVRP} \mid f^t x = a_0\}$ , entonces, utilizando las soluciones que acabamos de construir, comprobaremos que  $f_e = a_e, \forall e \in E$ , con lo que daremos por demostrado el resultado.

1.  $f_e = a_e = 0, \forall e \in E(B_j^i), i = 1, \dots, s, j = 0, \dots, n_i + 1$ . Dada una solución  $T^i$  de las construidas, añadiendo 2 copias de la arista  $e$  elegida se obtiene una nueva solución  $T'$ , que cumple  $a^t x^{T'} = a_0$ , lo cual implica que  $f^t x^{T'} = a_0$ , y como  $f^t x^{T'} = f^t x^{T^i} + 2f_e$ , con  $f^t x^{T^i} = a_0$ ,  $f_e$  ha de ser necesariamente igual a 0.
2.  $f_e = f_{b_j^r b_{j+1}^r}, \forall e \in (B_j^r, B_{j+1}^r), r = 1, \dots, s, j = 0, \dots, n_r$ . Dada una arista  $e = (u, v)$  cualquiera de la cortadura  $(B_j^r, B_{j+1}^r)$  es fácil construir a partir de  $T^i$ , con  $i \neq r$ , una nueva solución  $T'$  que utilice la arista  $e$  y que cumpla  $f^t x^{T'} = a_0$ . Basta con sustituir las aristas  $(u, b_j^r)$ ,  $(b_j^r, b_{j+1}^r)$  y  $(b_{j+1}^r, v)$ , que como sabemos forman parte de  $T^i$ , por la nueva arista  $(u, v)$ , es decir,  $T' = T^i - (u, b_j^r) - (b_j^r, b_{j+1}^r) - (b_{j+1}^r, v) + (u, v)$ . Evidentemente,  $a^t x^{T'} = a_0$  ya que los coeficientes en  $a$  de las aristas  $(u, b_j^r)$  y  $(b_{j+1}^r, v)$  son cero y  $a_{uv} = \frac{1}{n_r - 1} = a_{b_j^r b_{j+1}^r}$ . Consecuentemente,  $f^t x^{T'} = a_0$ , y por tanto:

$$\begin{aligned} f^t x^{T'} &= f^t x^{T^i} - f_{ub_j^r} - f_{b_j^r b_{j+1}^r} - f_{b_{j+1}^r v} + f_{uv} = a_0, \\ f^t x^{T^i} &= a_0. \end{aligned}$$

Restando respectivamente los términos de la derecha y de la izquierda de las dos ecuaciones anteriores, y teniendo en cuenta que por lo visto en el punto (1),  $f_{ub_j^r} = f_{b_{j+1}^r v} = 0$ , tenemos,  $-f_{b_j^r b_{j+1}^r} + f_{uv} = 0$ , lo que demuestra el resultado i.e.,

$$f_{uv} = f_{b_j^r b_{j+1}^r}, \forall (u, v) \in (B_j^r, B_{j+1}^r).$$

La demostración es independiente de la cortadura elegida, lo único importante es elegir adecuadamente la solución  $T^i$ , por lo tanto será cierto para cualquiera de ellas.

3.  $f_e = a_e, \forall e \in (B_j^i, B_{j+1}^i), i = 1, \dots, s, j = 0, \dots, n_i$ . En primer lugar demostraremos que todas las aristas de cualquiera de las cortaduras definidas por dos subconjuntos consecutivos en un mismo  $P_i$  tienen el mismo coeficiente, al que denominaremos  $\pi_i$ . Es decir, dado un  $i = 1, \dots, s$ , queremos ver que  $f_e = f_{e'} \forall e \in (B_j^i, B_{j+1}^i)$  y  $\forall e' \in (B_k^i, B_{k+1}^i)$  siendo  $j$  y  $k$  dos índices distintos de  $\{0, \dots, n_i\}$ . Podemos suponer sin pérdida de generalidad que  $j = 0$  y que  $k = 1$  i.e., que  $e \in (B_0^i, B_1^i)$  y  $e' \in (B_1^i, B_2^i)$ . Si no fuese así modificaríamos adecuadamente la solución  $T^i$  para que no contuviese ninguna copia de la arista  $e$ . A partir de  $T^i$  construimos  $T'$  eliminando las dos copias que hay de la arista  $e'$  y añadiendo dos copias de  $e$ , de esta forma,  $f^t x^{T'} = a_0$ , y por tanto,  $f^t x^{T'} = a_0$  (ver figura 4).

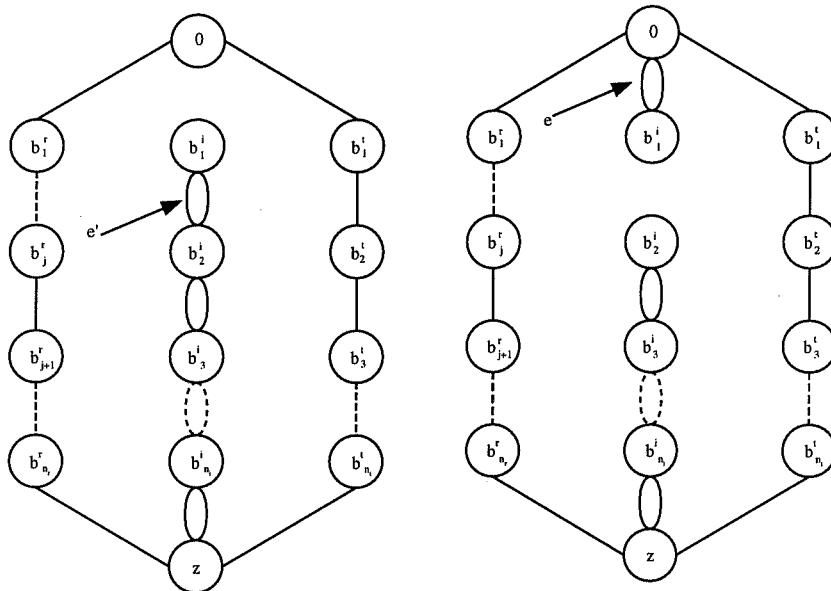


Figura 4. Rutas de  $T^i$  y  $T'$  que contienen a  $e'$  y a  $e$ , respectivamente.

Luego,

$$\left. \begin{array}{l} f^t x^{T^i} = a_0 \\ f^t x^{T'} = a_0 \end{array} \right\} \rightarrow f^t x^{T^i} - f^t x^{T'} = f^t x^{T^i} - f^t x^{T'} + 2f_{e'} - 2f_e = 0 \rightarrow f_{e'} = f_e$$

Como  $f^t x^{T^i} = a_0$  para cualquier  $i$ , sustituyendo todas las soluciones en la ecuación obtenemos el siguiente sistema de  $s$  ecuaciones con  $s$  incógnitas  $\pi_1, \dots, \pi_s$ :

$$\begin{aligned}
& 2n_1\pi_1 + (n_2 + 1)\pi_2 + \dots + (n_{s-1} + 1)\pi_{s-1} + (n_s + 1)\pi_s = a_0 \\
& (n_1 + 1)\pi_1 + 2n_2\pi_2 + \dots + (n_{s-1} + 1)\pi_{s-1} + (n_s + 1)\pi_s = a_0 \\
& \quad \vdots \\
& (n_1 + 1)\pi_1 + \dots + (n_{i-1} + 1)\pi_{n_{i-1}} + 2n_i\pi_i + \\
& \quad (n_{i+1} + 1)\pi_{n_{i+1}} + \dots + (n_s + 1)\pi_s = a_0 \\
& \quad \vdots \\
& (n_1 + 1)\pi_1 + \dots + (n_i + 1)\pi_{n_i} + \dots + 2n_s\pi_s = a_0
\end{aligned}$$

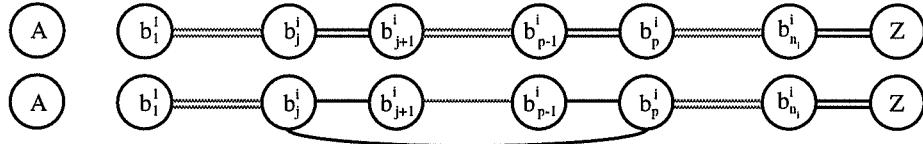
cuya única solución es:

$$\pi_i = \frac{1}{n_i - 1}, \quad i = 1, \dots, s,$$

lo que finalmente demuestra que  $f_e = a_e$ ,  $\forall e \in (B_j^i, B_{j+1}^i)$ ,  $i = 1, \dots, s$  y  $j = 0, \dots, n_i$ .

4.  $f_e = a_e$ ,  $\forall e \in (B_j^i, B_p^i)$  con  $p \geq j+2$  e  $i = 1, \dots, s$ . Observar que no hay pérdida de generalidad en suponer que  $j \geq 1$ . Si no lo fuese modificaríamos adecuadamente la solución  $T^i$  para que la siguiente operación tuviese sentido (véase la figura 5),

$$T' = T^i - \sum_{r=0}^{p-j-1} (b_{j+r}^i, b_{j+r+1}^i) + (b_j^i, b_p^i).$$



**Figura 5.** Construcción de  $T'$ , para probar que  $f_e = a_e$ ,  $\forall e \in (B_j^i, B_p^i)$  con  $p \geq j+2$ .

Es fácil comprobar que  $a^t x^{T'} = a_0$ , y por tanto  $f^t x^{T'} = a_0$ , es decir:

$$f^t x^{T'} = f^t x^{T^i} - \sum_{r=0}^{p-j-1} \frac{1}{n_i - 1} + f_{b_j^i b_p^i} = a_0.$$

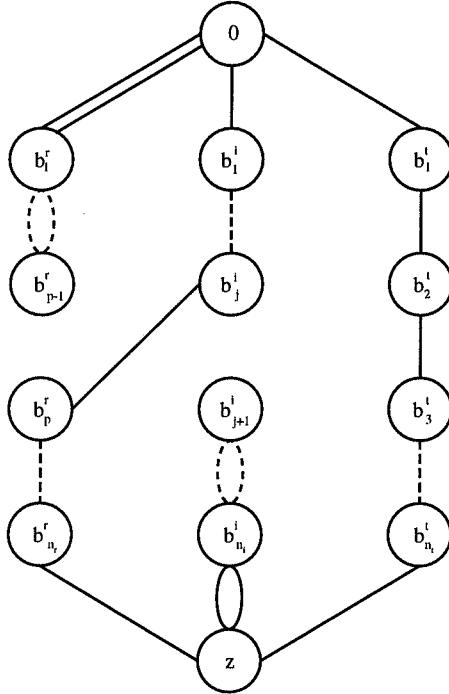
Luego,  $-\frac{p-j}{n_i - 1} + f_{b_j^i b_p^i} = 0$  y entonces,  $f_{b_j^i b_p^i} = \frac{p-j}{n_i - 1}$ , lo cual demuestra el resultado enunciado.

5.  $f_e = a_e$ ,  $\forall e \in (B_j^i, B_p^r)$ ,  $i \neq r$ ,  $1 \leq j \leq n_i$  y  $1 \leq p \leq n_r$ . Para demostrarlo construiremos una solución  $T'$  que cumplirá la restricción  $a^t x \geq a_0$  con igualdad y cuyas aristas

coinciden con las que se utilizan en  $T^i$  excepto las aristas de los caminos  $P_i$  y  $P_r$ . En su lugar tomaremos,

$$\begin{aligned}x_{b_j^i b_p^r} &= 1, \\x_{b_d^l b_{d+1}^i} &= 1, \quad d = 0, \dots, j-1, \\x_{b_d^r b_{d+1}^i} &= 2, \quad d = 0, \dots, p-2, \\x_{b_{j+d}^i b_{j+d+1}^r} &= 2, \quad d = 1, \dots, n_i - j, \\x_{b_{p+d}^r b_{p+d+1}^i} &= 1, \quad d = 0, \dots, n_r - p.\end{aligned}$$

En la figura 6 puede verse la parte de  $T'$  que difiere de  $T^i$ .



**Figura 6.** Vista parcial de la solución  $T'$ .

Podemos suponer sin pérdida de generalidad que:

$$a_e = \max \left\{ \frac{p}{n_r - 1} - \frac{j-2}{n_i - 1}, \frac{j}{n_i - 1} - \frac{p-2}{n_r - 1} \right\} = \frac{j}{n_i - 1} - \frac{p-2}{n_r - 1},$$

lo cual nos permite comprobar que efectivamente  $T'$  satisface,

$$d^t x^{T'} =$$

$$\sum_{q \neq i, q \neq r} \frac{n_q+1}{n_q-1} + \frac{j}{n_i-1} + \frac{j}{n_r-1} - \frac{p-2}{n_r-1} + 2 \frac{p-1}{n_r-1} + \frac{n_r+1-p}{n_r-1} + 2 \frac{n_i-j}{n_i-1} =$$

$$\sum_{q \neq i, q \neq r} \frac{n_q+1}{n_q-1} + \frac{j+j+2n_i-2j}{n_i-1} + \frac{-p+2+2p-2+n_r+1-p}{n_r-1} =$$

$$\sum_{q \neq i, q \neq r} \frac{n_q+1}{n_q-1} + \frac{n_i+1+n_i-1}{n_i-1} + \frac{n_r+1}{n_r-1} = \sum_{q=1}^s \frac{n_q+1}{n_q-1} + \frac{n_i-1}{n_i-1} =$$

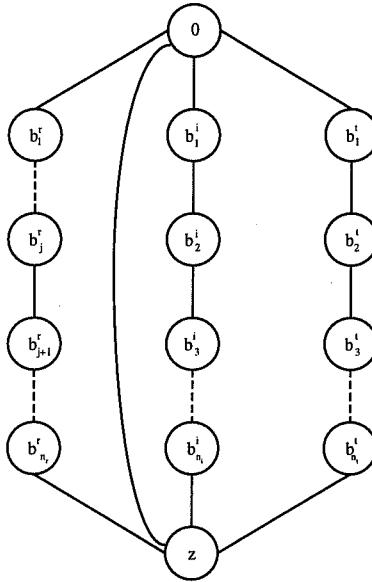
$$1 + \sum_{q=1}^s \frac{n_q+1}{n_q-1} = a_0,$$

y por tanto,  $f^t x^{T'} = a_0$ . De donde,  $f^t x^{T^i} - f^t x^{T'} = 0$  y, consecuentemente,

$$f^t x^{T^i} - f^t x^{T'} = \frac{2n_i}{n_i-1} + \frac{n_r+1}{n_r-1} - [\frac{j}{n_i-1} + 2 \frac{n_i-j}{n_i-1} + f_e + 2 \frac{p-1}{n_r-1} + \frac{n_r+1-p}{n_r-1}]$$

$\downarrow$

$$f_e = \frac{2n_i-j-2n_i+2j}{n_i-1} + \frac{n_r+1-2p+2-n_r-1+p}{n_r-1} = \frac{j}{n_i-1} - \frac{p-2}{n_r-1} = a_e.$$



**Figura 7.** Vista parcial de la solución  $T'$  utilizada para calcular  $f_e$ ,  $\forall e \in (A, Z)$ .

6.  $f_e = 1, \forall e \in (A, Z)$ . El resultado se obtiene directamente comparando una cualquiera de las soluciones  $T^i$  con la solución  $T'$  dada en la figura 7, y que se obtiene a partir de ella de la siguiente forma:

$$T' = T^i - \sum_{r=1}^{n_i} (b_r^i, b_{r+1}^i) + (b_0^i, b_1^i) + (0, z).$$

Evidentemente,  $a^t x^{T'} = \frac{n_i+1}{n_i-1} + 1 + \sum_{r \neq i} \frac{n_r+1}{n_r-1} = a_0$ , lo que implica,

$$f^t x^{T^i} - f^t x^{T'} = 0 \longrightarrow \frac{2n_i}{n_i-1} - \frac{n_i+1}{n_i-1} - f_e = 0,$$

y por tanto:

$$f_e = \frac{2n_i - n_i - 1}{n_i - 1} = \frac{n_i - 1}{n_i - 1} = 1 = a_e.$$

Hemos visto que  $a_e = f_e$  para cualquier arista  $e \in E$ , lo que demuestra que las desigualdades *path* inducen facetas del  $P_{SDVRP}$ .

□

En el apéndice adjunto demostramos que las condiciones (i) y (ii) son suficientes para garantizar que las desigualdades *wheelbarrow* inducen facetas del  $P_{SDVRP}$ . La técnica de demostración es la misma que la utilizada para probar el resultado anterior.

Como ya hemos señalado, las condiciones establecidas en el teorema 1 son sencillas de comprobar, no requieren la resolución de problemas tipo *Bin-Packing* y, en ciertos casos, son menos fuertes que las exigidas en Cornuéjols y Harche (1993). Evidentemente, si en una instancia concreta, en la que se ha fijado el número de vehículos y la capacidad de los mismos, se cumplen las condiciones de Cornuéjols y Harche (1993), también se cumplirán las establecidas en el teorema anterior. El siguiente ejemplo, con el que concluimos esta cuestión, muestra que el recíproco no es cierto:

**Ejemplo:** Dada la instancia del *SDVRP* en la que hay dos vehículos de demanda igual a 16, y 11 clientes cuyas demandas son,  $d_1 = d_2 = 5$ ,  $d_i = 1$ ,  $i = 3, \dots, 10$ , y  $d_{11} = 14$ , si definimos:

$$\begin{aligned} A &= \{0\}; & Z &= \{11\}; & P_1 &= \{1, 2\}, & (B_{11} = 1, B_{12} = 2); \\ P_2 &= \{3, 4\}, & (B_{21} = 3, B_{22} = 4); & P_3 &= \{5, 6\}, & (B_{31} = 5, B_{32} = 6); \\ P_4 &= \{7, 8\}, & (B_{41} = 7, B_{42} = 8); & P_5 &= \{9, 10\}, & (B_{51} = 9, B_{52} = 10), \end{aligned}$$

es evidente que la desigualdad asociada a esta configuración *path* induce una faceta del  $P_{SDVRP}$ , ya que se cumplen las condiciones establecidas en el enunciado del teorema. Sin embargo, no podemos decir lo mismo cuando nos referimos al  $P_{GVRP}$  asociado. En

este caso es sencillo comprobar que no existe ninguna partición del conjunto de clientes en 2 subconjuntos tales que se cumplan las condiciones establecidas en Cornuéjols y Harche (1993).

Llegado este punto queda por analizar qué pasa cuando el depósito pertenece a cualquier otro subconjunto de la partición, i.e., cuando  $0 \in B_j^i$  con  $i = 1, \dots, s$  y  $j = 1, \dots, n_i$ . Cornuéjols y Harche (1993) distinguen claramente dos situaciones, dependiendo de si el número de vehículos que se necesita para servir la demanda de  $V \setminus B_j^i$  es o no superior a uno. Nosotros distinguiremos también la misma casuística.

**Teorema 2.** *Si  $0 \in B_j^i$ , con,  $i = 1, \dots, s$ ,  $j = 1, \dots, n_i$ , y  $d(V \setminus B_j^i) \leq Q$ , entonces, las desigualdades «path», «wheelbarrow» y «bicycle» inducen facetas del PSDVRP.*

#### Demostración

Este resultado se deduce directamente de los teoremas 3.2 y 3.4 de Cornuéjols *et al.* (1985), y del teorema 2.6 de Naddef y Rinaldi (1991), que establecen, respectivamente, que dichas desigualdades inducen facetas del  $P_{GTSP}$ . Basta con tener en cuenta que bajo la hipótesis establecida cualquier solución del  $GTSP$  puede extenderse a una solución del  $SDVRP$  añadiendo, únicamente, aristas con coeficiente nulo en la desigualdad.

Veámoslo, sea  $C_{B_j^i}$  un ciclo Hamiltoniano en  $B_j^i$  y  $K$  el número de vehículos disponible para abastecer la demanda de  $V \setminus \{0\}$ . Evidentemente, cualquier tour asociado a una solución del  $GTSP$  recorrerá todos los vértices del grafo, si añadimos a este tour  $K - 1$  copias de cada una de las aristas de  $C_{B_j^i}$  obtendremos un tour en el que  $K$  vehículos visitan todos los vértices de  $B_j^i$ , y solamente uno de ellos visitará a los vértices del complementario. Por hipótesis, sabemos que se necesita un único vehículo para abastecer la demanda de  $V \setminus B_j^i$ , luego seguro que hay una asignación posible de demandas de clientes a vehículos de manera que el conjunto de aristas que acabamos de dar define una solución del  $SDVRP$ . Obsérvese que las únicas aristas añadidas son aristas de  $E(B_j^i)$ , como el coeficiente de estas aristas es igual a cero, si la solución del  $GTSP$  de la que partíamos pertenecía a la cara del  $P_{GTSP}$  inducida por  $a^T x \geq a_0$ , la solución del  $SDVRP$  que hemos construido pertenecerá también a la cara del  $P_{SDVRP}$  inducida por la misma desigualdad.

□

Hasta aquí los resultados que nos han permitido extender facetas del  $P_{GTSP}$  a facetas del  $P_{SDVRP}$ . Como demostramos a continuación, cuando el número de vehículos necesario para servir la demanda de  $V \setminus B_1^i$  es superior a uno el tipo de conclusiones a las que se llega no es tan satisfactorio. En el caso de necesitar al menos tres, concluimos que las desigualdades *path* no son ni tan siquiera soporte. Lo que ocurre en este caso en relación con las desigualdades *wheelbarrow* o *bicycle* es, por el momento, un tema abierto.

**Proposición 1.** Si el depósito pertenece a  $B_1^i$ , para algún  $i = 1, \dots, s$ , y el número de vehículos necesario para servir la demanda de los vértices de  $V \setminus B_1^i$  es al menos tres, las desigualdades «path» no son soporte del PSDVRP.

#### Demostración

Sea  $a^t x \geq a_0$  una desigualdad asociada a una configuración *path*. Supongamos que  $T$  es una solución del *SDVRP* cuyo vector de incidencia minimiza el término de la izquierda de la restricción y veamos que con las hipótesis establecidas esto implica,  $a^t x^T > a_0$ .

Podemos suponer sin pérdida de generalidad que las únicas aristas de  $T$  incidentes con  $B_1^i$  son aristas de  $(A, B_1^i) \cup (B_1^i, B_2^i)$ . Si no fuera así, aplicando un número finito de veces el procedimiento que describimos a continuación obtendríamos una solución  $T'$  que cumpliría la condición y seguiría minimizando  $a^t x$ .

Si en  $T$  hay una arista  $e \in (B_1^i, B_j^i)$ , con  $3 \leq j \leq n_i + 1$ , sustituyendo cada una de las copias de  $e = (u, v)$  por el camino  $u - b_2^i - \dots - b_r^i - b_{r+1}^i - \dots - b_{j-1}^i - v$ , obtenemos una nueva solución  $T'$ , que cumple  $a^t x^T = a^t x^{T'}$ , y no contiene a  $e$ .

Análogamente, si  $e \in (B_1^i, B_p^r)$ ,  $r \neq i$ ,  $p \geq 1$ , sustituyéndola por la arista  $(u, b_0^i)$  y el camino  $b_0^i = b_0^r - b_1^r - \dots - b_j^r - \dots - b_{p-1}^r - v$ , obtenemos una solución  $T'$ , que cumple  $a^t x^T \geq a^t x^{T'}$ . Basta con observar que la contribución al término izquierdo de la desigualdad de la nueva arista y del camino es  $\frac{1}{n_i-1} + \frac{p}{n_r-1}$ , y que,

$$a_e \geq \frac{p}{n_r-1} - \frac{1-2}{n_i-1} = \frac{1}{n_i-1} + \frac{p}{n_r-1}.$$

Sea  $\bar{k} = \lceil \frac{d(V \setminus B_1^i)}{Q} \rceil$  el número de vehículos que necesitamos para servir la demanda de  $V \setminus B_1^i$ . Como el depósito pertenece a  $B_1^i$ , la cortadura  $(B_1^i, \bar{B}_1^i)$  se atravesará al menos  $2\bar{k}$  veces utilizando, como hemos visto, solamente aristas de  $(A, B_1^i) \cup (B_1^i, B_2^i)$ . Estas aristas se distribuirán, si nos olvidamos de cuál es la cortadura concreta, siguiendo uno de los patrones siguientes:

Cortadura 1	Cortadura 2	Eliminando	Quedan
$2\bar{k}$	0	$(2\bar{k}-2, 0)$	$(2, 0)$
$2\bar{k}-1$	1	$(2\bar{k}-2, 0)$	$(1, 1)$
$2\bar{k}-2$	2	$(2\bar{k}-4, 0)$	$(2, 2)$
$2\bar{k}-3$	3	$(2\bar{k}-4, 2)$	$(1, 1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\bar{k}$	$\bar{k}$	Si $\bar{k}$ es par $(\bar{k}-2, \bar{k}-2)$ Si $\bar{k}$ es impar $(\bar{k}-1, \bar{k}-1)$	$(2, 2)$ $(1, 1)$

Observar que a partir de una solución cualquiera del *SDVRP* en la que una arista  $e$  aparece más de dos veces, siempre podemos conseguir una solución del *GTSP* en la que dicha arista se utiliza, a lo sumo, dos veces. Basta con eliminar un número adecuado de copias de la misma, a saber,  $2p$  copias, si hubiese  $2p+1$ , con  $p \geq 1$ , o  $2(p-1)$ , si hubiese  $2p$ , con  $p \geq 2$ . Por lo tanto, para un  $\bar{k}$  fijo, siempre es posible eliminar al menos  $2(\bar{k}-2)$  aristas de la cortadura  $(B_1^i, \bar{B}_1^i)$  de manera que se sigue teniendo un tour,  $T'$ , que recorre todos y cada uno de los vértices de  $V$ . Este tour es una solución del *GTSP* y por tanto,

$$a^t x^{T'} = a^t x^T - \frac{2(\bar{k}-2)}{n_i-1} \geq a_0,$$

por ser las desigualdades asociadas a las configuraciones *path* válidas para el *GTSP*, Cornuéjols *et al.* (1985). Despejando obtenemos el resultado que buscábamos,

$$a^t x^T \geq a_0 + \frac{2(\bar{k}-2)}{n_i-1} > a_0.$$

Obsérvese que la hipótesis  $\bar{k} \geq 3$  garantiza que  $2(\bar{k}-2) \geq 2$ , y consecuentemente también,  $\frac{2(\bar{k}-2)}{n_i-1} \geq \frac{2}{n_i-1} > 0$ .

□

El único resultado conocido por el momento para el caso en que el número de vehículos que se necesita para servir  $d(V \setminus B_1^i)$  es dos, es el siguiente:

**Proposición 2.** *Si el depósito pertenece a  $B_1^i$  para algún  $i = 1, \dots, s$  y el número de vehículos necesario para abastecer la demanda de  $V \setminus B_1^i$  es exactamente 2, entonces, en cualquier solución del *SDVRP* cuyo vector de incidencia pertenezca a la cara inducida por una desigualdad «path», la demanda de  $V \setminus P_i$  debe servirse utilizando un único vehículo, i.e., la cortadura  $(P_i, \bar{P}_i)$  solamente se atraviesa dos veces.*

La demostración puede consultarse con todo detalle en Martínez (1995).

Con la proposición que presentamos a continuación damos por terminado el estudio de las relaciones poliédricas entre el *GTSP* y el *SDVRP*.

**Proposición 3.** *Las desigualdades asociadas a configuraciones «bicycle» con  $s = 3$  y  $n_i = 2$ ,  $\forall i = 1, \dots, s$  (conocidas como «bicycle 2-regulares») están dominadas por  $a^t x \geq a_0 + 2$ , si  $\lceil \frac{d(V \setminus P_i)}{Q} \rceil = 2$ , siendo  $P_i$  el camino que contiene el subconjunto  $B_j^i$  al que pertenece el depósito.*

### *Demostración*

Supongamos que el depósito pertenece a  $B_j^i$ , como para servir la demanda de  $V \setminus P_i$  necesitamos dos vehículos, la cortadura  $(P_i, \bar{P}_i)$  será atravesada al menos dos veces por dos vehículos distintos. Por lo tanto, en cualquier solución  $T$  del *SDVRP* debe haber al menos cuatro aristas que pertenezcan a dicha cortadura.

Consideremos el grafo inducido por las aristas de  $T$ , teniendo en cuenta el resultado básico de teoría de grafos que establece,

«el doble del número de aristas de un grafo es igual a la suma de los grados de sus vértices»

deducimos que en  $T$  habrá al menos 7 aristas. Observar que el grado del subconjunto que contiene al depósito es al menos cuatro y el de los otros cinco subconjuntos que completan la configuración es al menos dos.

Teniendo en cuenta que cuatro de estas aristas deben pertenecer a la cortadura  $(P_i, \bar{P}_i)$ , y que el coeficiente en  $a$  de las mismas es al menos dos, tenemos,

$$\sum_{e \in T \cap \delta(P_i)} a_e x_e^T \geq 8.$$

Notar que si la desigualdad fuese soporte debería cumplirse lo siguiente:

$$a^T x^T = \sum_{e \in E(P_i)} a_e x_e^T + \sum_{e \in T \cap \delta(P_i)} a_e x_e^T + \sum_{e \in E(\bar{P}_i)} a_e x_e^T = a_0 = 10,$$

y sin embargo, como necesitamos al menos tres aristas más para poder completar una solución del *SDVRP*,

$$a^T x^T \geq 8 + 3 = 11 > a_0$$

pues los coeficientes de estas aristas son mayores o iguales que 1.  $\square$

Es sencillo comprobar que esta demostración no puede generalizarse para  $s \geq 5$ , pues,

$$\text{número aristas} = \frac{\Sigma \text{grados vértices}}{2} \geq \frac{4+2(2s-1)}{2} = 2 + (2s - 1) = 2s + 1,$$

$$a^T x^T \geq 8 + ((2s + 1) - 4) = 2s + 5,$$

y como  $a_0 = 3s + 1$  tenemos que:

$$2s + 5 > 3s + 1 \Leftrightarrow 4 > s.$$

Por último, comentar que en Martínez (1995) se demuestra que las desigualdades *path* y las *wheelbarrow* son precisamente las restricciones que surgen cuando se intentan generalizar al contexto del *SDVRP* algunas de las desigualdades *2-matching generalizadas* introducidas por Araque (1989) para el *CVRP* con demandas unitarias. Concretamente, las que surgen cuando la capacidad del vehículo se establece en un valor tal, que la demanda de 6 clientes cualesquiera puede servirse completamente utilizando un único vehículo.

#### 4. CONSIDERACIONES FINALES

El estudio poliédrico que hemos realizado nos ha permitido extender resultados conocidos para el poliedro del *Agente Viajero Gráfico* al poliedro del *Problema de Rutas de Vehículos con Demanda Compartida*. Hemos visto cómo las desigualdades asociadas a las configuraciones *path*, *wheelbarrow* y *bicycle* son también válidas para el  $P_{SDVRP}$  y bajo ciertas condiciones incluso inducen facetas.

Las desigualdades que hemos considerado en este trabajo no son las únicas familias de desigualdades válidas conocidas para el *SDVRP*. Como hemos dicho en la introducción, en Martínez *et al.* (2000) y Belenguer *et al.* (2000) se presentan nuevas restricciones, que bajo condiciones adecuadas, también inducen facetas del  $P_{SDVRP}$ . Además, por lo que nosotros sabemos, estos trabajos son, hoy por hoy, los únicos en los que se estudia la estructura facial del  $P_{SDVRP}$ . No obstante, cabe la posibilidad de continuar la investigación en esta línea tratando de deducir condiciones bajo las cuáles, desigualdades como, las *path*, *wheelbarrow* y *bicycle* con *capacidades*, introducidas en Cornuéjols y Harche (1993) para el *GVRP*, las clásicas restricciones *comb* y *clique tree* del *TSP* (Grötschel y Padberg, 1985), o las de *capacidad generalizada* del *CVRP* (Harche y Rinaldi, 1991), induzcan facetas del  $P_{SDVRP}$ . La adaptación de estas restricciones al *SDVRP* y la demostración de su validez para el correspondiente poliedro pueden consultarse en Martínez (1995).

Por último, nos gustaría destacar que, aparte del interés estrictamente teórico que pueden tener los resultados que hemos presentado en este trabajo, la sencillez de las condiciones establecidas los hacen también atractivos desde el punto de vista práctico. Sin duda, será interesante investigar el desarrollo de algoritmos que permitan identificar restricciones de este tipo cuando sean violadas por soluciones fraccionarias, lo cuál ayudaría a resolver instancias del problema de tamaño superior a las resueltas hasta el momento. Desde este punto de vista, el trabajo que hemos presentado abre la puerta a futuras investigaciones en relación con el tema, a la vez que amplia el conocimiento de la estructura facial del  $P_{SDVRP}$ .

## AGRADECIMIENTOS

Nos gustaría expresar nuestro agradecimiento a los dos evaluadores anónimos por sus comentarios sobre el trabajo, los cuales nos han permitido mejorar el enfoque y la calidad del mismo.

## REFERENCIAS

- Araque, J.R. (1989). *A Polyhedral Approach to Vehicle Routing*. Comunicación presentada en ORSA/TIMS Meeting. New York.
- Augerat, P., Belenguer, J.M., Benavent, E., Corberán, A., Naddef, D. y Rinaldi, G. (1995). «Computational Results with a Branch and Cut Code for the Vehicle Routing Problem». *Technical Report RR949-M*. Institut IMAG. Grenoble.
- Belenguer, J.M. y Martínez, M.C. (2000). «Búsqueda Tabú en Problemas de Rutas de Vehículos con Demanda Compartida». *Technical Report 1-00*. Universitat de València.
- Belenguer, J.M., Martínez, M.C. y Mota, E. (2000). «A Lower Bound for the Split Delivery Vehicle Routing Problem». *Operations Research*, 48, 801-810.
- Bodin, L.D., Golden, B.L., Assad, A.A. y Ball, M.O. (1983). «Routing and Scheduling of Vehicles and Crews: The State of the Art». *Computers and Operations Research*, 10, 63-211.
- Christofides, N. (1985). «Vehicle Routing», in: *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan y D.B. Shmoys eds. Wiley, Chichester, 431-448.
- Cornuéjols, G., Fonlupt, J. y Naddef, D. (1985). «The Traveling Salesman Problem on a Graph and some Related Integer Polyhedra». *Mathematical Programming*, 33, 1-27.
- Cornuéjols, G. y Harche, F. (1993). «Polyhedral Study of the Capacitated Vehicle Routing Problem». *Mathematical Programming*, 60, 21-52.
- Dror, M. y Trudeau, P. (1989). «Savings by Split Delivery Routing». *Transportation Science*, 23, 141-145.
- Dror, M. y Trudeau, P. (1990). «Split Delivery Routing». *Naval Research Logistics*, 37, 383-402.
- Garey, M.R. y Johnson, D.S. (1979). *Computers and Intractability: A Guide to the Theory of NP-completeness*. Freeman, San Francisco.
- Grötschel, M. y Padberg, M.W. (1985). «Polyhedral Theory», in: *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan y D.B. Shmoys eds. Wiley, Chichester, 251-305.
- Martínez, M.C. (1995). *Rutas de Vehículos con Demanda Compartida*. Tesis Doctoral. Universitat de València.

- Martínez, M.C., Mota, E. y Rinaldi, G. (2000). «The Split Delivery Vehicle Routing Polyhedron: New Families of Facet Defining Inequalities». En proceso de revisión.
- Naddef, D. y Rinaldi, G. (1991). «The Symmetric Traveling Salesman Polytope: New Facets from the Graphical Relaxation». *Technical Report*, 248. Istituto di Analisi dei Sistemi ed Informatica del CNR. Roma.
- Nemhauser, G.L. y Wolsey, L.A. (1988). *Integer and Combinatorial Optimization*. Wiley, New York.
- Padberg, M.W. y Rinaldi, G. (1991). «A Branch and Cut Algorithm for the Resolution of Large-Scale Symmetric Traveling Salesman Problems». *SIAM Review*, 33 (1), 60-100.
- Pulleyblank, W.R. (1983). «Polyhedral Combinatorics», in: *Mathematical Programming: The State of the Art*, A. Bachem, M. Grötschel and B. Korte eds. Springer, Berlin.

## APÉNDICE

*Dados  $K$  vehículos de capacidad  $Q$ , si  $0 \in A$  y se cumplen las dos condiciones:*

i.-  $K \geq \frac{s-1}{2}$ , y

ii.-  $Q \geq d(P_s) + d(P_{s-1}) + d(P_{s-2})$ , siendo  $d(P_1) \leq d(P_2) \leq \dots \leq d(P_s)$ .

*entonces, las desigualdades «wheelbarrow» inducen facetas de  $PSDVRP$ .*

### Demostración

Análogamente a como hemos hecho con las desigualdades *path*, vamos a construir  $s$  soluciones distintas, una para cada camino  $P_i$ , que nos permitirán demostrar el resultado vía el método indirecto. Las soluciones se construyen siguiendo el mismo procedimiento que con las *path* excepto en la forma de unir las parejas de caminos. Se unirán utilizando la arista que une a los representantes de sus subconjuntos extremos,  $B'_{n_r}$ . En la figura 8 presentamos la solución asociada al camino  $P_i$ , la llamaremos  $T^i$ . No es difícil comprobar que las hipótesis establecidas garantizan la existencia de una asignación de demandas a vehículos de forma que con  $\frac{s-1}{2}$  vehículos se sirven los clientes de  $V \setminus A$ .

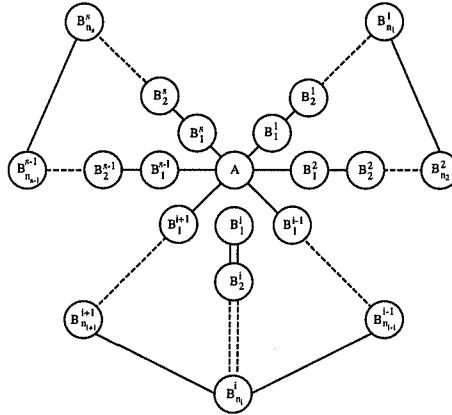


Figura 8. Solución  $T^i$  para una configuración *wheelbarrow*.

Demostraremos a continuación que las soluciones así construidas cumplen con igualdad la desigualdad *wheelbarrow*  $a^T x \geq 1 + \sum_{r=1}^s \frac{n_r+1}{n_r-1}$ . Probado esto es evidente que se

trata de una desigualdad propia, ya que basta con duplicar cualquiera de las aristas utilizadas con coeficiente estrictamente positivo, para encontrar una solución que cumple la desigualdad con mayor estricto.

A fin de simplificar las cuestiones de notación comprobaremos el resultado para la solución  $T^{s-1}$ , la demostración para el resto de soluciones,  $T^i$ , sería análoga.

Observar que las únicas aristas en la solución con coeficiente no nulo son, las que unen a dos subconjuntos  $B_p^r$  consecutivos en cada camino  $P_r$ , con  $r = 1, \dots, s$ ,  $r \neq i$ , y las que utilizamos para unir parejas de caminos. Supongamos que los caminos se han unido de dos en dos empezando por el primero y en orden ascendente, es decir, la demanda de los clientes en  $P_{2r-1}$  y  $P_{2r}$  se sirve utilizando el mismo vehículo, para cada  $r = 1, \dots, \frac{s-3}{2}$ . Evaluaremos  $a^T x^{T^{s-1}}$  contando las aristas que se utilizan en la solución y cuánto valen sus coeficientes. Para cada uno de estos caminos se utilizan  $n_r$  aristas de peso  $\frac{1}{n_r-1}$  y una arista de peso  $\frac{n_{2r}}{n_{2r}-1} - \frac{n_{2r-1}-2}{n_{2r-1}-1}$  por cada par de caminos unidos. Además, hay  $2(n_{s-1} - 1)$  aristas de peso  $\frac{1}{n_{s-1}-1}$  que recorren los subconjuntos de  $P_{s-1}$ ,  $n_{s-2}$  aristas de peso  $\frac{1}{n_{s-2}-1}$  para  $P_{s-2}$  y  $n_s$  de peso  $\frac{1}{n_s-1}$  para  $P_s$ , mas las dos aristas que unen a  $P_{s-2}$  con  $P_{s-1}$  y a éste con  $P_s$ .

Observar que el coeficiente de las aristas que unen los subconjuntos finales de dos caminos cualesquiera,  $P_{2r-1}$  y  $P_{2r}$  es:

$$\max \left\{ \frac{n_{2r}}{n_{2r}-1} - \frac{n_{2r-1}-2}{n_{2r-1}-1}, \frac{n_{2r-1}}{n_{2r-1}-1} - \frac{n_{2r}-2}{n_{2r}-1} \right\},$$

puede comprobarse fácilmente que en este caso el máximo es único e igual a:

$$\frac{1}{n_{2r}-1} + \frac{1}{n_{2r-1}-1}.$$

Luego,

$$a^T x^{T^{s-1}} = \sum_{r=1, r \neq s-1}^s \frac{n_r}{n_r-1} + \sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}}{n_{2r}-1} - \frac{n_{2r-1}-2}{n_{2r-1}-1} \right) + \\ \frac{n_s}{n_s-1} - \frac{n_{s-1}-2}{n_{s-1}-1} + \frac{2n_{s-1}-2}{n_{s-1}-1},$$

es decir,

$$a^T x^{T^{s-1}} = \sum_{r=1}^s \frac{n_r}{n_r-1} + \sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}}{n_{2r}-1} - \frac{n_{2r-1}-2}{n_{2r-1}-1} \right) + \frac{n_s}{n_s-1} = \\ \sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}}{n_{2r}-1} + \frac{n_{2r-1}}{n_{2r-1}-1} + \frac{n_{2r}}{n_{2r}-1} - \frac{n_{2r-1}-2}{n_{2r-1}-1} \right) + \frac{2n_s}{n_s-1} = \\ \sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}}{n_{2r}-1} + \frac{n_{2r-1}}{n_{2r-1}-1} + \frac{n_{2r}(n_{2r-1}-1)-(n_{2r}-1)(n_{2r-1}-2)}{(n_{2r}-1)(n_{2r-1}-1)} \right) + \frac{2n_s}{n_s-1},$$

que por unicidad del máximo es igual a,

$$\sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}}{n_{2r-1}} + \frac{n_{2r-1}}{n_{2r-1}-1} + \frac{1}{n_{2r-1}} + \frac{1}{n_{2r-1}-1} \right) + \frac{2n_s}{n_s-1} =$$

$$\sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}+1}{n_{2r-1}} + \frac{n_{2r-1}+1}{n_{2r-1}-1} \right) + \frac{2n_s}{n_s-1} =$$

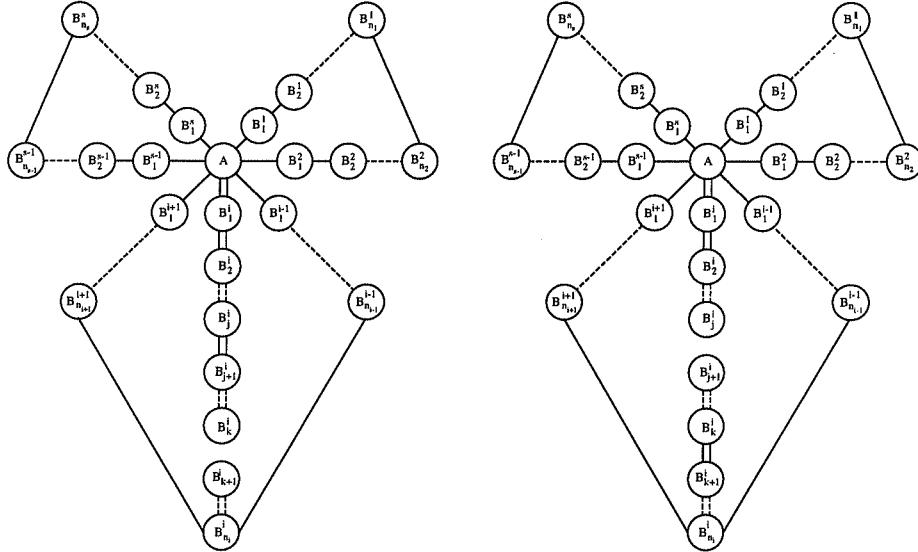
$$1 + \sum_{r=1}^s \frac{n_r+1}{n_r-1}.$$

Probado que la desigualdad *wheelbarrow*,  $a^t x \geq a_0$ , es soporte y no trivial, supongamos ahora que existe una desigualdad  $f^t x \geq a_0$  que define una faceta del poliedro y cumple:

$$\{x \in P_{SDVRP} \mid a^t x = a_0\} \subseteq \{x \in P_{SDVRP} \mid f^t x = a_0\}.$$

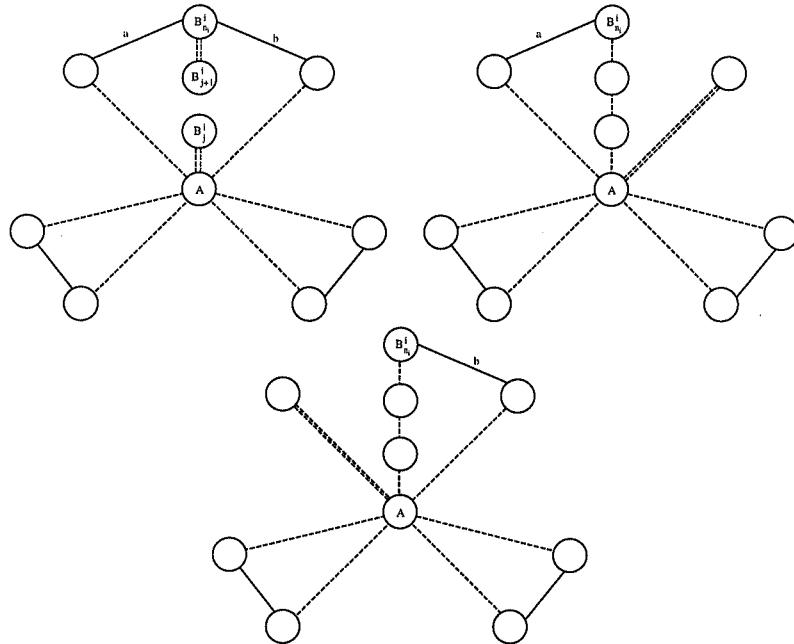
Demostraremos que  $f_e = a_e$ ,  $\forall e \in E$  y que por tanto  $a^t x \geq a_0$  induce una faceta del  $P_{SDVRP}$ .

La demostración de los dos primeros casos es análoga a la realizada para las desigualdades *path* pero teniendo en cuenta la solución  $T^i$  que hemos descrito para las configuraciones *wheelbarrow*.



**Figura 9.** Soluciones demostrando que  $f_e = f_{e'}, \forall e \in (B_j^i, B_{j+1}^i), \forall e' \in (B_k^i, B_{k+1}^i)$ .

1.  $f_e = a_e = 0, \forall e \in E(B_j^i), i = 1, \dots, s, j = 0, \dots, n_i$ .
2.  $f_e = f_{b_j^r b_{j+1}^r}, \forall e \in (B_j^r, B_{j+1}^r), r = 1, \dots, s, j = 0, \dots, n_r - 1$ .
3.  $f_e = a_e, \forall e \in (B_j^i, B_{j+1}^i), i = 1, \dots, s, j = 0, \dots, n_i - 1$ . En primer lugar demostraremos que todas las aristas de cualquiera de las cortaduras definidas por dos subconjuntos consecutivos en un mismo  $P_i$  tienen el mismo coeficiente, que denominaremos  $\pi_i$ . Es decir, dado un  $i = 1, \dots, s$ , queremos ver que  $f_e = f_{e'} \forall e \in (B_j^i, B_{j+1}^i)$  y  $\forall e' \in (B_k^i, B_{k+1}^i)$  siendo  $j$  y  $k$  dos índices distintos de  $\{0, \dots, n_i - 1\}$ . Para demostrarlo basta con comparar las soluciones de la figura 9, observar que solamente se diferencian en aristas que pertenecen a estas cortaduras, además ambas pertenecen a la cara inducida por la desigualdad *wheelbarrow*, y por lo tanto también a la inducida por  $f^t x \geq a_0$ .



**Figura 10.**  $T^i, T^{i-1}, T^{i+1}$  soluciones del SDVRP pertenecientes a la cara inducida por la desigualdad *wheelbarrow*.

Una vez visto que  $f_e = \pi_i, \forall e \in (B_j^i, B_{j+1}^i)$ , para cada  $i = 1, \dots, s$ , completaremos la demostración resolviendo el sistema que se obtiene al sustituir en  $f^t x$  los vectores de incidencia asociados a las soluciones  $T^{i-1}, T^i$  y  $T^{i+1}$ ,  $i = 2, \dots, s - 1$ , figura 10. Estas soluciones corresponden a los tours utilizados en Cornuéjols *et al.* (1985) para demostrar el teorema 3.5. Gracias a las hipótesis establecidas sobre el número de

vehículos disponibles y su capacidad no es difícil comprobar que estos tours definen soluciones posibles del *SDVRP*. Como los tres vectores de incidencia cumplen  $a^t x = a_0$ , también cumplirán  $f^t x = a_0$ , de donde obtenemos,

$$\begin{aligned} f^t x^{T^i} &= f^t x^{T^{i-1}} \\ f^t x^{T^i} &= f^t x^{T^{i+1}} \end{aligned}$$

es decir,

$$\begin{aligned} n_{i-1}\pi_{i-1} + 2(n_i - 1)\pi_i + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} + n_{i+1}\pi_{i+1} &= \\ 2n_{i-1}\pi_{i-1} + n_i\pi_i + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} + n_{i+1}\pi_{i+1} & \\ n_{i-1}\pi_{i-1} + 2(n_i - 1)\pi_i + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} + n_{i+1}\pi_{i+1} &= \\ n_{i-1}\pi_{i-1} + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + n_i\pi_i + 2n_{i+1}\pi_{i+1} & \\ \downarrow & \\ (n_i - 2)\pi_i + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} &= \\ n_{i-1}\pi_{i-1} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} & \\ (n_i - 2)\pi_i + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} &= \\ n_{i+1}\pi_{i+1} + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} & \end{aligned}$$

que, considerándolo para  $i = 2, \dots, s-1$ , proporciona el sistema:

$$\begin{aligned} (n_2 - 2)\pi_2 + f_{b_{n_1}^1 b_{n_2}^2} + f_{b_{n_2}^2 b_{n_3}^3} &= n_1\pi_1 + f_{b_{n_2}^2 b_{n_3}^3} \\ (n_2 - 2)\pi_2 + f_{b_{n_1}^1 b_{n_2}^2} + f_{b_{n_2}^2 b_{n_3}^3} &= n_3\pi_3 + f_{b_{n_1}^1 b_{n_2}^2} \\ \vdots & \\ (n_i - 2)\pi_i + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} &= n_{i-1}\pi_{i-1} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} \\ (n_i - 2)\pi_i + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} + f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} &= n_{i+1}\pi_{i+1} + f_{b_{n_{i-1}}^{i-1} b_{n_i}^i} \\ \vdots & \\ (n_{s-1} - 2)\pi_{s-1} + f_{b_{n_{s-2}}^{s-2} b_{n_{s-1}}^{s-1}} + f_{b_{n_{s-1}}^{s-1} b_{n_s}^s} &= n_{s-2}\pi_{s-2} + f_{b_{n_{s-1}}^{s-1} b_{n_s}^s} \\ (n_{s-1} - 2)\pi_{s-1} + f_{b_{n_{s-2}}^{s-2} b_{n_{s-1}}^{s-1}} + f_{b_{n_{s-1}}^{s-1} b_{n_s}^s} &= n_s\pi_s + f_{b_{n_{s-2}}^{s-2} b_{n_{s-1}}^{s-1}} \end{aligned}$$

cuya solución es:

$$\begin{aligned} \pi_i &= \frac{h}{n_i - 1} \\ f_{b_{n_i}^i b_{n_{i+1}}^{i+1}} &= \frac{h}{n_i - 1} + \frac{h}{n_{i+1} - 1} \quad i = 1, \dots, s \end{aligned}$$

Sustituyendo en  $f^t x$  el vector de incidencia asociado a  $T^{s-1}$  y teniendo en cuenta que ha de ser igual a  $a_0$ , tenemos:

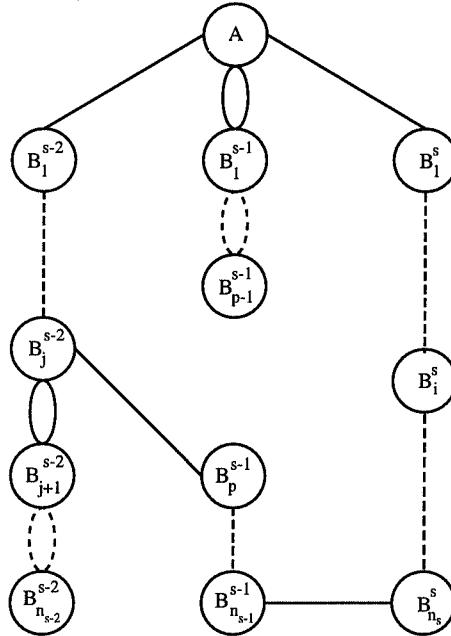
$$f^t x^{T^{s-1}} = \sum_{r=1}^{\frac{s-1}{2}} \left( \frac{n_{2r}h}{n_{2r}-1} + \frac{n_{2r-1}h}{n_{2r-1}-1} + \frac{h}{n_{2r}-1} + \frac{h}{n_{2r-1}-1} \right) + \frac{2n_sh}{n_s-1}$$

de donde,

$$f^t x^{T^{s-1}} = h \left( 1 + \sum_{r=1}^{\frac{s}{2}} \frac{n_r+1}{n_r-1} \right) = ha_0 = a_0$$

luego,  $h = 1$ , lo que demuestra que los coeficientes de las aristas que estamos considerando coinciden en  $a$  y en  $f$ .

4.  $f_e = a_e, \forall e \in (B_j^i, B_p^r), i \neq r, 1 \leq j \leq n_i \text{ y } 1 \leq p \leq n_r$ . No hay pérdida de generalidad en hacer la demostración para una arista de la cortadura  $e \in (B_j^{s-2}, B_p^{s-1})$ . Ya hemos demostrado que los coeficientes en  $a$  y  $f$  coinciden para el resto de las aristas, lo que nos permitirá demostrar el resultado basándonos simplemente en el hecho de que la solución de la figura 11,  $T'$ , es una solución del SDVRP que cumple  $a^t x^{T'} = a_0$  con igualdad, por lo que también cumplirá  $f^t x^{T'} = a_0$ .



**Figura 11.** Solución  $T'$ .

En efecto:

$$\begin{aligned}
a^T x^{T'} &= \sum_{r=1}^{s-3} \frac{n_r+1}{n_r-1} + \frac{n_{s-2}-j}{n_{s-2}-1} + \frac{n_{s-2}}{n_{s-2}-1} + \\
&\quad \frac{j}{n_{s-2}-1} - \frac{p-2}{n_{s-1}-1} + \frac{2(p-1)}{n_{s-1}-1} + \frac{n_{s-1}-p}{n_{s-1}-1} + \\
&\quad \frac{1}{n_{s-1}-1} + \frac{1}{n_s-1} + \frac{n_s}{n_s-1} = \\
&\sum_{r=1}^{s-3} \frac{n_r+1}{n_r-1} + \frac{n_{s-2}-j+n_{s-2}+j}{n_{s-2}-1} + \frac{-p+2+2p-2+n_{s-1}-p+1}{n_{s-1}-1} + \frac{1+n_s}{n_s-1} = \\
&\sum_{r=1}^{s-3} \frac{n_r+1}{n_r-1} + \frac{n_{s-2}+1+n_{s-2}-1}{n_{s-2}-1} + \frac{n_{s-1}+1}{n_{s-1}-1} + \frac{n_s+1}{n_s-1} = \\
&1 + \sum_{r=1}^s \frac{n_r+1}{n_r-1} = a_0
\end{aligned}$$

como consecuencia,

$$f^T x^{T'} - a^T x^{T'} = f_e - a_e = 0 \longrightarrow f_e = a_e$$

Concluimos así que las desigualdades asociadas a las configuraciones *wheelbarrow* inducen facetas del  $P_{SDVRP}$ .

## **ENGLISH SUMMARY**

# **FROM THE GRAPHICAL TRAVELING SALESMAN POLYHEDRON TO THE ONE ASSOCIATED WITH THE SPLIT DELIVERY VEHICLE ROUTING PROBLEM**

CARMEN MARTÍNEZ

ENRIQUE MOTA

Universitat de València\*

*In this paper we study the polyhedron associated with the Split Delivery Vehicle Routing Problem, a distribution problem arising in the context of servicing goods to clients using a fixed fleet of identical vehicles with limited capacity. The objective is to design a set of routes in such a way that the total distance travelled is minimized. The main difference with other known routing problems is that the demand of a given client can be serviced among several vehicles. We present some polyhedral results that are obtained as extensions of previously known results for the Graphical Traveling Salesman Problem.*

**Keywords:** Vehicle routing, split delivery, polyhedron, facets

**AMS Classification (MSC 2000):** 90C10, 90C27, 90C35

---

\* Departament d'Estadística i Investigació Operativa. Universitat de València. Dr. Moliner, 50. Burjassot, 46100. España. E-mail: M.Carmen.Martinez@uv.es

– Received October 1999.

– Accepted October 2000.

## 1. INTRODUCTION

The *Split Delivery Vehicle Routing Problem (SDVRP)* is a relaxation of the classical *Capacitated Vehicle Routing Problem (CVRP)*, in which the demand of any client can be serviced by more than one vehicle. This relaxation may result in cheaper solutions, both in terms of the total distance travelled, as well as in the number of vehicles to be used. Moreover, the minimum number of vehicles needed to service the total demand is always known in advance. However, the *SDVRP* remains an *NP-hard* problem (Dror and Trudeau, 1990). Hopefully, the knowledge of the structure of the polyhedron associated with the *SDVRP* feasible solutions will allow the optimal resolution of bigger and bigger instances. The purpose of this paper is to continue the polyhedral study presented in Belenguer *et al.* (2000) and in Martínez *et al.* (2000).

## 2. NOTATION AND PROBLEM DEFINITION

Let  $G = (V, E)$  be a complete and undirected graph, where the set of vertices,  $V$ , represents  $n$  clients with known integer demands,  $d_i$ , and a depot, vertex 0, where a fleet of  $K$  identical vehicles of capacity  $Q$  is located. Each edge  $e = (i, j)$  in  $E$  represents the (shortest) path between  $i$  and  $j$ .

In Belenguer *et al.* (2000), the feasible solutions of the *SDVRP* are defined on  $G$ , in terms of a multiset of edges subject to problem constraints. It is also proved that the convex hull of the incidence vectors associated with the *SDVRP* feasible solutions is an unbounded and full dimensional polyhedron, denoted in what follows  $P_{SDVRP}$ . Moreover, such vectors satisfy a simple set of constraints, (1) – (4), which in fact defines a relaxation of the *SDVRP*. These constraints, subject to small changes, have also been considered in the formulation of the *Graphical TSP (GTSP)* and in the graphical relaxation of the *CVRP*, the *GVRP*. Using the above relaxation, Belenguer *et al.* (2000) solved to optimality *SDVRP* instances up to 50 clients, and Martínez *et al.* (2000) obtained the complete linear description of the polyhedron associated to a *SDVRP* instance with 5 clients and 2 vehicles.

## 3. FROM $P_{GTSP}$ TO $P_{SDVRP}$

From the *GVRP*, *SDVRP*, and *GTSP* problem definitions it is easy to see that  $P_{GVRP} \subseteq P_{SDVRP} \subseteq P_{GTSP}$ , and therefore, every *GTSP* valid inequality will also be a valid inequality for the *SDVRP*. However, the facial structures can not be so easily related.

In this section we extend to the  $P_{SDVRP}$  some inequalities that are facet defining inequalities for the  $P_{GTSP}$ . These inequalities are associated with some complex structures,

known as *path*, *wheelbarrow* and *bicycle configurations*, that were first introduced in Cornuéjols *et al.* (1985) for the  $P_{GTSP}$ . Obviously, *path*, *wheelbarrow* and *bicycle inequalities* are valid inequalities for the  $P_{SDVRP}$ , since they are valid for the  $P_{GTSP}$ .

Moreover, if the conditions given in Cornuéjols and Harche (1993) hold, the *path inequalities* not only induce facets of  $P_{GVRP}$ , but also of  $P_{SDVRP}$ . Next theorem presents another set of conditions, easier to check and, at least in some cases, weaker than the ones given in Cornuéjols and Harche (1993).

**Theorem 1.** *Given  $K$  vehicles of capacity  $Q$ , if  $0 \in A$  (or if  $0 \in Z$ ), and the following conditions hold:*

$$i.- \quad K \geq \frac{s-1}{2},$$

$$ii.- \quad Q \geq d(P_s) + d(P_{s-1}) + d(P_{s-2}), \text{ such that } d(P_1) \leq d(P_2) \leq \dots \leq d(P_s), \text{ and}$$

$$iii.- \quad Q \geq 2 \left\lceil \frac{d(Z) + \sum_{i=1}^s d(P_i)}{s-1} \right\rceil$$

then, *path inequalities induce facets of  $P_{SDVRP}$* .

In the appendix it is shown that conditions i) and ii) are sufficient to guarantee that *wheelbarrow inequalities* induce facets of  $P_{SDVRP}$ .

Following the same lines as in Cornuéjols and Harche (1993), we have proved the following result:

**Theorem 2.** *If  $0 \in B_j^i$ , with,  $i = 1, \dots, s$ ,  $j = 1, \dots, n_i$ , and  $d(V \setminus B_j^i) \leq Q$ , then, *path*, *wheelbarrow* and *bicycle inequalities induce facets of  $P_{SDVRP}$* .*

Finally, this section includes some results, showing situations in which *path inequalities* are not even support inequalities (**Proposition 1**), exhibiting a special kind of a *path configuration* for which *path inequalities* can not define facets (**Proposition 2**), or conditions under which some *bicycle inequalities* are dominated (**Proposition 3**).

In Martínez (1995) it is shown that *path* and *wheelbarrow* inequalities are in fact the *2-matching generalized inequalities*, first introduced in Araque (1989) for the *CVRP* with unit demands.

#### 4. FINAL COMMENTS

New families of valid inequalities, that under given conditions are also facet-defining inequalities, are presented in Martínez *et al.* (2000) and Belenguer *et al.* (2000). In

Martínez (1995), valid inequalities for the  $P_{SDVRP}$  are obtained from other inequalities defined for the  $P_{GVRP}$ , for the  $P_{TSP}$  and for the  $P_{CVRP}$ . Further research remains to be done in order to show if there are conditions under which they induce facets of  $P_{SDVRP}$ .

The results presented in this paper have more than just a theoretical interest, since the simple conditions imposed in order to guarantee the facial character of *path inequalities* could be incorporated, via new identification routines, into a general procedure that would allow the resolution of bigger *SDVRP* instances.

# Estadística Oficial



## AVALUACIÓ DELS MERCATS DELS SERVEIS D'INFORMACIÓ ELECTRÒNICA

TOMÀS BAIGET

Institut d'Estadística de Catalunya\*

*Es presenten les característiques i els principals resultats del «MSStudy II», un projecte per avaluar la situació dels mercats dels serveis d'informació electrònica a l'àrea econòmica europea (AEE), [Study for assessing the situation of the markets for electronic information services in the European economic area (EEA)] que, impulsat per la Direcció General XIII de la Comissió Europea, ha dut a terme a Espanya l'Institut d'Estadística de Catalunya (Idescat), amb la col-laboració de l'Asociación Multisectorial de la Información (Asedie). El MSStudy II, que durant 1998 i 1999 s'ha realitzat simultàniament en 16 països europeus, és una activitat estadística experimental –recollida, però, en el Pla Estadístic de Catalunya 1997-2000– i ha tractat d'obtenir dades desglossades per tipus d'informació, tipus de suport o mitjà pel qual es distribueix la informació, tipus d'usuari, etc. Els resultats de cada país han estat integrats l'estiu del 2000 per un equip de professionals alemanys –els mateixos que varen planificar i dirigir aquest exercici per encàrrec de la CE.*

**Electronic information services markets assessment**

**Paraules clau:** Mercat, informació electrònica, programes europeus, societat de la informació, tecnologies de la informació, continguts informatius

**Classificació AMS (MSC 2000):** 62P20, 91B44

\* Institut d'Estadística de Catalunya, Assistència Tècnica Estadística. Via Laietana, 58. 08003 Barcelona. Spain. E-mail: baiget@idescat.es

– Rebut el desembre de 1999.

– Acceptat el novembre de 2000.

## **1. ANTECEDENTS**

Un dels objectius del programa *Info2000* de la CE (que en el seu període de vigència 1996-1999 va tractar d'estimular la producció de continguts informatius i multimèdia per a empreses, administracions públiques i ciutadans en l' emergent Societat de la Informació) va ser avaluar l'estat de desenvolupament dels mercats de serveis d'informació electrònica –i també alguns altres serveis electrònics colaterals, com ara de comunicació, de publicitat, o transaccionals– en els països de l'AEE (la UE més Noruega i Islàndia), i comparar els resultats amb els d'altres parts del món. A partir d'aquest exercici la CE podria emprendre les accions necessàries per tal d'impulsar serveis més competitius i de més qualitat.

L'enfocament d'aquest estudi es basa en un anterior *MSStudy I* [1] que es dugué a terme els anys 1996/97 sobre dades de 1994 en el marc del programa *Impact II* (l'antecessor d'*Info2000*) de la CE amb objectius similars. La principal diferència ha estat en que ara es va abastar una temàtica una mica més àmplia, cobrint no solament els serveis d'informació estrictament professionals com llavors, sino també les novetats i canvis produïts durant els dos anys anteriors:

- productes multimèdia, tant online com en cd-rom;
- explosió del fenòmen Internet;
- serveis online que el productor d'informació dirigeix directament al consumidor (sense passar per intermediaris, siguin aquests tant empreses distribuïdores com centres de documentació);
- serveis electrònics en general (no necessàriament d'informació) com ara correu-e, comerç-e, hostatjament de webs comercials, etc., que tenen una gran influència en el desenvolupament del sector de la informació.

Per tant, calia considerar molts altres agents dins d'aquesta indústria de la informació, sense oblidar també les llars privades com un grup d'usuaris –consumidors, com hem dit– que cada vegada va adquirint més importància.

De la mateixa manera que el seu predecessor, el *MSStudy II* ha hagut de ser coordinat de la manera més rigorosa possible per tal que els resultats dels diferents països fossin comparables i agregables a nivell europeu. Igualment, s'ha cuidat també la comparabilitat amb el *MSStudy I* per comprovar els canvis produïts en aquest interval de temps [2].

## **2. PROPÒSITS DEL MSSTUDY II**

- Contribuir a un coneixement millor de l'estat de desenvolupament de la societat de la informació en els països de l'AEE.

- Proveir als països participants amb dades comparatives de la seva posició nacional en relació als altres.
- Tenir una base per prendre decisions polítiques.
- Poder oferir als agents dels mercats de serveis electrònics una informació estratègica el més acurada possible.

Aquests propòsits generals s'han assolit mitjançant l'anàlisi detallada dels diversos aspectes que conformen el sector. Així, el *MSStudy*:

- ha donat una visió panoràmica del volum, estructura i dinàmica d'aquests mercats d'informació electrònica a cada país;
- ha donat pautes per gestionar els aspectes estratègics, així com avaluar els punts forts i febles;
- ha identificat els dèficits i desequilibris en l'oferta i la demanda;
- ha analitzat les barreres econòmiques, psicològiques, tècniques i normatives en els diversos segments de mercat, així com eventuals situacions monopolístiques i de dependència.

Els resultats de les enquestes nacionals han estat agregats a nivell europeu per tal de:

- poder assessorar la *CE* sobre els punts forts i febles del mercat europeu, comparant-los amb els d'altres parts del món (sobre tot, Nord-Amèrica);
- ajudar els agents d'aquests mercats a extreure conclusions estratègiques a partir de les dades empíriques.

### **3. OBJECTIUS DEL MSSTUDY II**

Conèixer millor:

- proveïdors de productes i serveis d'informació electrònica
- usuaris d'aquests proveïdors

Alguns proveïdors ofereixen també altres serveis electrònics, especialment en les àrees de publicitat (aquesta és decisiva per a l'existència de moltes publicacions) i transaccions (compres i vendes tant de publicacions com d'altres bens, reserves de viatges i hotels, etc.), que, encara que marginalment, també interessen pel *MSStudy II*.

### **4. DESCRIPCIÓ GENERAL DE LES TASQUES I INSTRUMENTS DE RECERCA MÉS ADEQUATS**

#### **4.1. Sumari de les tasques de recerca**

El període de referència ha estat l'any 1997, encara que es varen recollir algunes estimacions del 1998 en els casos en què això va ser possible.

La taula següent dóna una visió de conjunt de les feines realitzades:

Àrees de recerca dels mercats de la informació	Metodologia emprada
<b>I. Recollida de dades: any 1997</b> (i pels instituts nacionals que varen voler fer-ho voluntàriament: resultats esperats pel 1998)	
1. Oferta	– Enquesta a tots els subministradors d'informació electrònica
2. Demanda	– Enquesta a una mostra representativa de tota la població d'usuaris a les llars privades
3. Empreses	– Estudi de cas d'un segment concret que sigui d'especial interès per al país.
<b>II. Recollida d'informació: perspectives</b> 4. Mercats totals: desenvolupaments futurs	– Entrevistes a experts nacionals sobre tendències, impediments i oportunitats
<b>III. Descripció, anàlisi i recomanacions</b> 5. Mercats totals: passat, present i futur	– Paradigma de característiques en el context de <i>benchmarkings</i> intra-europeus

#### 4.2. Primer nivell de recerca: enquesta sobre l'oferta

S'enfocà a la recollida de dades sobre els volums de vendes i els beneficis dels proveïdors.

La següent matriu dóna una visió de conjunt dels tipus de proveïdors enuestats a tot Europa, les funcions que realitzen i la metodologia proposada per l'equip de coordinació. El mètode que després es va aplicar en cada país va dependre de les característiques locals. Així, per exemple, a Espanya sempre s'ha pogut fer un cens de tots els tipus de proveïdors, encara que d'alguns d'ells a l'època considerada no n'hi havia (audiotex, proveïdors de pagament dirigits a consumidors o usuaris privats, etc.).

Tipus de proveïdor	Funció	Metodologia
Proveïdor d'informació electrònica principalment per a ús de les empreses	Proveir informació-e a nivell nacional i/o internacional	Cens
	Proveir accés a/per Internet, nacional i/o internacional	Mostra representativa
	Proveir en àrees regionals i/o locals (municipals)	Treball de gabinet (ús de fonts diverses ja existents)

Tipus de proveïdor	Funció	Metodologia
Proveïdor d'informació electrònica principalment per a ús privat	Proveir informació-e <i>online</i> a nivell nacional i/o internacional	Cens
	Proveir informació-e <i>offline</i> a nivell nacional i/o internacional	Mostra representativa
	Proveir accés a/per Internet, nacional i/o internacional	Mostra representativa
	Proveir en àrees regionals i/o locals (municipals)	Treball de gabinet
	Grans serveis adreçats als consumidors	Cens
Altres proveïdors de serveis electrònics, la majoria dels quals no són d'informació	Altres serveis als consumidors	Mostra representativa o treball de gabinet
	Comunicació Oci Transaccions Educació Qualsevol altre servei fet electrònicament	Treball de gabinet i entrevistes a experts

#### 4.3. Segon nivell de recerca: enquesta sobre la demanda

Es va fer un estudi sobre els usuaris a les llars.

Tipus d'usuari	Funció o servei	Metodologia
Privats i d'empreses (però sempre obtenint les respostes des de les llars)	Ús i intensitat d'ús dels serveis electrònics tant en llars com en empreses	Mostra aleatòria representativa de la població total (llars): 1.600 entrevistes a cada país*. Qüestionari/enquesta telefònica ajudada per ordinador

\* En el cas d'Espanya es van fer 1.600 entrevistes addicionals sols a Catalunya, és a dir, en total 3.200.

Aquesta ha estat una evaluació que ha permès establir comparacions immediates i fer projeccions a tota Europa.

#### 4.4. Tercer nivell de recerca: medi empresarial

Amb les restriccions de temps i de pressupost amb les quals s'ha treballat no va ser possible descriure i analitzar tots els factors que influeixen en els mercats nacionals. La feina –realitzada bàsicament per l'equip coordinador europeu– s'ha concentrat, doncs, en dues tasques:

- descripció àmplia dels mercats d’informació-e en el medi empresarial;
- èmfasi especial en els punts forts i febles relatius en el medi empresarial i en el context d’un *benchmarking* europeu.

MSStudy II ha pogut aprofitar els resultats del preexistent projecte *Esis (European survey of information society projects and actions)* de la *Comissió Europea*, que proporcionà dades bàsiques a nivell nacional i europeu. La informació podia ser descarregada directament de: <http://www.ispo.cec.be/esis/HomeFacts.html>

Els indicadors més importants tinguts en compte han estat:

- infraestructura tècnica (en la mesura en què no quedés coberta per l'estudi de la demanda del propi *MSStudy*);
- infraestructura institucional: organismes polítics i privats rellevants per desenvolupar la societat de la informació, amb especial èmfasi en el desenvolupament dels mercats d’informació-e;
- accions i programes polítics especials, plans i iniciatives;
- temes legals i normatius tals com copyright, protecció de dades personals, propietat, responsabilitat, legislació multimèdia, etc.
- altres factors (p. ex. lingüístics, culturals, d’hàbit) que poden ser favorables o contraris.

A més a més de la descripció del medi empresarial, s’ha dut a terme un estudi de cas sobre un segment concret especialment rellevant en el mercat nacional. A Espanya s’ha escollit «Ús de la informació en els laboratoris farmacèutics», ja que aquests són tradicionalment uns importants usuaris o consumidors d’informació. En altres països s’ha fet: «Possibilitats de finançació externa dels serveis d’informació-e» (Bèlgica); «Un cas d’èxit: la *Société Européenne des Satellites Astra*» (Luxemburg); «Serveis d’informació-e regional/local: l’experiència de Kuusamo» (Finlàndia); «Premsa regional: conjunt de millors pràctiques per a Europa» (Regne Unit); etc.

Aquests estudis de casos han resultat valuosos per intercanviar experiència pràctica entre els diferents països.

#### **4.5. Quart nivell de recerca: entrevista a experts nacionals sobre futurs desenvolupaments, noves tendències i oportunitats**

Per:

- 1) avaluar si les tendències actuals s’alteraran (accelerant-se, ralentint-se o parant-se);
  - 2) detectar quins dels nous desenvolupaments prosperaran realment, com s’integraran en el panorama de la informació-e i de quina manera modificarien la situació general.
- En aquest context han estat d’interès els següents temes:

- R+D (de tot el món) i el seu impacte comercial en el sector de la informació;
  - nous productes i serveis electrònics, tant d'informació com generals, així com les seves aplicacions en els sectors industrials i de serveis;
  - nous enfocaments en les polítiques d'informació nacionals i internacionals;
- 3) punts forts i febles de les indústries i dels mercats d'informació europeus, identificant els colls d'ampolla que dificulten el desenvolupament del mercat i detectant avantatges competitius nacionals i paneuropees en relació als EUA, Japó i les noves regions «d'informació emergent»;
  - 4) detectar oportunitats i evitar riscos a la *Comissió Europea*, als governs nacionals i als agents dels mercats, i
  - 5) formular, on sigui possible, un consens paneuropeu per tal de, p. ex:
    - reforçar la posició de les polítiques d'informació tant de la *CE* com dels governs nacionals;
    - encoratjar els subministradors i usuaris d'informació-e a desenvolupar més els seus avantatges competitius, i
    - vèncer les barreres dels mercats nacionals dins l'AEE i contribuir a la formació d'un veritable mercat d'informació-e.

#### **4.6. Descripció, anàlisi i recomanacions**

Els informes per a cada Estat descriuen els desenvolupaments (passat, present i futur) a partir de les dades i de la informació recollida, i de la seva comparació amb el *MSSStudy I*, i resalten els punts forts i febles dels seus mercats nacionals en relació als altres països. Es poden extreure conclusions i es fan recomanacions per a la indústria, els governs nacionals i la *CE* [3].

### **5. ESBÓS D'ORGANITZACIÓ I GESTIÓ DEL PROJECTE**

Igual que en el *MSSStudy I*, les enquestes i les anàlisis han estat dutes a terme a cada país per organitzacions nacionals amb experiència en aquest camp de recerca, i han estat supervisades per un equip coordinador establert per la *Comissió Europea*.

#### **5.1. Principals tasques dels instituts nacionals**

- recollir dades mitjançant enquestes de l'oferta i la demanda dels mercats d'informació-e del seu país;
- cooperar amb l'equip coordinador per recollir dades dels proveïdors d'informació multinacionals;

- descriure el medi empresarial i de negoci dels mercats de la informació nacionals, i dur a terme un estudi de cas sobre un segment concret que tingui importància en el context nacional;
- recollir informació sobre futurs desenvolupaments en cooperació amb experts nacionals;
- anàlisi i interpretació dels resultats des del punt de vista nacional, en un informe en anglès;
- evaluació dels punts forts i febles dels mercats i de la indústria de la informació nacionals en el context d'un *benchmarking* intraeuropeu;
- publicar els resultats obtinguts a nivell nacional;
- cooperar amb l'equip de coordinació en la redacció de l'informe europeu.

## **5.2. Principals tasques de l'equip de coordinació**

- garantir la comparabilitat dels resultats de tots els estudis nacionals;
- agregar les dades a nivell de l'AEE;
- avaluar els desenvolupaments dels mercats d'informació-e des d'un punt de vista europeu;
- componer mosaics de vendes estimades a cada país dels subministradors d'informació multinacionals i americans, alguns dels quals sovint es neguen a donar xifres dels seus mercats;
- redactar l'informe europeu consultant els instituts nacionals i altres experts, presentant i analitzant els resultats agregats.

## **6. CALENDARI**

### **Fase 1: Preparació (octubre-desembre 1998)**

- conclusió dels contractes;
- reunió de posada en marxa (*kick-off meeting*) a Luxemburg amb tots els contractants nacionals, per tal de fixar les metodologies comunes i definir detalls de l'estudi, com els mostrejos de les diferents parts dels mercats;
- compilació d'adreces dels proveïdors que seran enquestats;
- discutir quins indicadors cal agafar del servidor d'*Esis* i feina addicional a fer per tal de descriure el medi empresarial dels mercats d'informació-e;
- escollir un tema o estudi de cas nacional;
- fer entre tots (instituts nacionals i equip coordinador) una llista comuna de proveïdors internacionals (multinacionals) d'informació-e;
- escollir la mida de la mostra per a l'enquesta dels usuaris (demanda).

#### **Fase 2: Recollida de dades (gener-juliol 1999)**

- recollir dades quantitatives i qualitatives de l'oferta i la demanda;
- anàlisi dels proveïdors internacionals (conjuntament amb l'equip coordinador);
- anàlisi provisional de les primeres dades resultat de les enquestes (esborrany de les «taules obligatòries» amb els resultats quantitatius més importants tant de l'oferta com de la demanda);
- equip coordinador: estudi dels mercats informació-e d'EUA, Japó, etc.;
- entrevistes amb experts.

#### **Fase 3: Anàlisi de les dades (agost-novembre 1999)**

- entrega de les taules obligatòries a l'equip coordinador;
- reunió a Luxemburg per discutir els resultats provisionals, problemes apareguts amb les taules i amb els informes en general;
- data final d'entrega dels esborrany dels informes nacionals, estudi de cas inclòs;
- data final per l'esborrany de l'informe europeu amb dades agregades, ratios, rankings, projeccions i comparacions.
- reunió final a Luxemburg. Presentació i discussió dels resultats principals del projecte.

#### **Fase 4: Informes finals (desembre 1999-octubre 2000)**

- data final d'entrega dels informes nacionals definitius;
- publicació de les dades de l'informe europeu (només les dades principals);
- data final d'entrega de l'informe europeu definitiu (amb tots els resultats quantitatius i qualitatius, anàlisis, recomanacions i perspectives de futur);
- publicació de l'informe europeu (octubre 2000) [4].

### **7. SUBVENCIÓ DE LA CE**

La *CE* ha subvencionat l'estudi espanyol amb Euro 40.000 (PTA 6.680.000) equivalent al 50% del cost total del projecte (les quantitats han estat diferents segons la mida dels països). A Espanya l'altre 50% ha estat aportat per l'*Idescat*.

### **8. RESULTATS PRINCIPALS**

En aquest article només tenim espai per comentar alguns dels trets principals. Remetem el lector als informes originals o a altres articles que han tractat aspectes parcials específics (vegeu la Bibliografia al final).

## 8.1. Oferta

Caldria distingir entre els diversos aspectes pels quals es pot subdividir la informació. Una possible divisió és entre la d'ús «professional» i la d'«entreteniment». En aquest sentit els resultats mostren un bon manteniment de la primera i un considerable increment de la producció d'aquesta última. No obstant això, i analitzant les tendències dels últims anys, els experts europeus temen que Europa aviat es veurà dominada per un oligòpoli nordamericà que posarà en perill les cultures regionals, tal com ha passat amb el cinema i la música pop.

Com ja s'ha dit, aquest estudi s'ha centrat sobretot en la informació professional.

### 8.1.1. Europa

La següent taula pot donar una idea del rànquing europeu del mercat d'informació electrònica per països. S'han exclòs els ingressos que les empreses proveïdores d'informació electrònica tenen per altres conceptes com ara formació, assessoria, publicacions impresaes, programes informàtics, etc.:

País	Vendes 1997 (M Euro)	Quota de mercat (%)
Regne Unit	2.041,8	29,9
Alemanya	1.117,8	16,3
França	746,9	10,9
Holanda	746,7	10,9
Itàlia	680,8	10,0
Espanya	305,3	4,5
Suècia	184,6	2,7
Dinamarca	180,9	2,6
Finlàndia	163,4	2,4
Noruega	151,0	2,2
Bèlgica	110,9	1,6
Grècia	108,2	1,6
Àustria	102,4	1,5
Luxemburg	73,8	1,1
Irlanda	53,1	0,8
Portugal	48,3	0,7
Islàndia	10,8	0,2
Total AEE	6.826,7	100,0

Alguns experts estimen que Suïssa, país que no és de l'Àrea Econòmica Europea, estaria entre els 250 i 300 M Euro.

La dada d'Irlanda s'ha calculat a partir del valor de *MSSstudy I* de 1994, augmentat segons el terme mitjà europeu, ja que aquest país no va participar en el *MSSstudy II*.

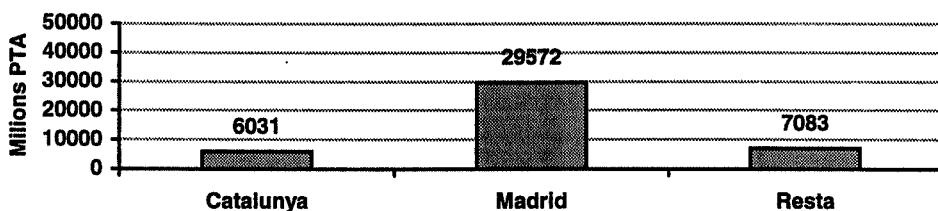
En la taula anterior es veu com en consum o vendes Espanya figura en la 6a posició sobre 17, però si aquests valors es relacionen amb el nombre d'habitants Espanya passa a ocupar un preocupant lloc 16è, seguida només per Portugal.

Totes les quantitats en euros que figuren en aquesta taula estan disminuïdes del valor de la inflació en els respectius països entre 1994 (any de referència del *MSSstudy I*) i 1997. En el cas espanyol la inflació va ser de 10,6%.

#### 8.1.2. Espanya

El cens d'empreses que produeixen informació electrònica es va obtenir a partir de fonts diverses ([5], [6], [7] i [8], entre d'altres), i finalment es varen contabilitzar les dades de 232 empreses. Anàlisis internes realitzades a l'*Idescat* a partir de les dades de *MSSstudy* han mostrat una concentració de la producció d'informació electrònica sobre tot a Madrid, i també a Catalunya. En el gràfic següent es poden veure intuïtivament les proporcions de les vendes entre Catalunya (90), Madrid (83 empreses) i la resta d'Espanya (39). A Catalunya hi ha més empreses, però són més petites.

*Ingressos dels proveïdors espanyols d'informació electrònica  
per vendes sols d'informació electrònica (total: 42.686 M PTA)*



Hem dit «intuïtivament» ja que diversos factors fan que els totals espanyols que figuren a la taula i al gràfic anteriors (305,3 M Euro i 42.686 M PTA = 256,5 M Euro, respectivament) no siguin directament comparables entre sí ja que en el gràfic:

- no hi figuren les vendes de les empreses estrangeres (importacions);
- no s'ha disminuït del valor de la inflació 1994-1997;
- s'han incorporat les vendes de 12 empreses espanyoles –ubicades a Espanya–, actualment propietat de multinacionals estrangeres, que segons la metodologia acordada pel *MSSstudy* –amb una visió paneuropea– havien de considerar-se com a estrangeres i les seves vendes com a «importacions».

## 8.2. Demanda

Tal com ja s'ha esmentat a l'apartat 4.3., la demanda potencial d'informació electrònica es va analitzar mitjançant 3.200 enquestes telefòniques, per realitzar les quals varen haver de fer-se més de 22.000 trucades. L'operació es dugué a terme el desembre de 1998 amb l'ajut d'una empresa especialitzada (*call centre*), els operadors de la qual compta-ven amb un programari *computer aided telephone interview (cati)*. L'equip de 16 entrevistadors, que reberen un curset de formació específica, va invertir-hi dues setmanes.

Seguint les normes donades per l'equip coordinador del *MSSStudy*, només es va entrevisitar i comptabilitzar l'estrat de població entre 15 i 65 anys, que és el considerat més actiu.

Mesurar l'ús de les tecnologies de la informació no és el mateix que mesurar l'ús de la informació —disposar de l'equipament necessari no vol dir que s'utilitzi amb la finalitat d'informar-se— però, al menys, en dóna una idea. En l'enquesta es varen incloure algunes preguntes directes sobre ús dels cd-roms i d'Internet per informació però les respostes no varen resultar del tot fiables per la mateixa ambigüetat de la paraula «informació».

*Penetració de TICs el desembre de 1998  
(en % sobre el total de la població)*

Pais	Ús de PC	Ús d'Internet	Ús de CD-ROM
1. Islàndia	80,0	51,7	34,8
2. Dinamarca	79,6	48,6	33,9
3. Suècia	72,5	—	—
4. Holanda	67,2	36,6	48,7
5. Finlàndia	63,0	—	—
6. Bèlgica	56,5	16,4	25,8
7. Regne Unit	55,1	34,3	22,3
8. Alemanya	52,5	20,8	24,7
9. Àustria	49,0	21,2	21,0
10. Itàlia	45,4	12,2	20,3
11. Portugal	44,4	22,5	23,3
12. Luxemburg	41,7	—	—
13. Noruega	41,0	26,5	13,5
14. França	39,4	—	—
15. Espanya	36,8	12,6	18,5
16. Grècia	33,2	5,8	5,5
<b>TOTAL</b>	<b>43,3</b>	<b>24,1</b>	<b>23,1</b>

### *8.2.1. Europa*

Degut a alguns endarreriments burocràtics de la CE en la signatura dels contractes, es va produir una desincronia de les operacions en els diversos països —des de juny de 1998 fins abril de 1999— que en aquesta matèria d'evolució tan accelerada podia produir biaixos perceptibles.

Dels molts resultats que podríem presentar en aquest curt article, hem triat els de la penetració de les TIC (tecnologies de la informació i la comunicació) per habitant que es mostren en la taula anterior. Només 12 dels 16 països varen seguir estrictament les normes que permeten la comparabilitat.

### *8.2.2. Espanya*

Gràcies al mostreig fet especialment per a Catalunya, amb 1.600 entrevistes addicionals a càrrec del pressupost de l'*Idescat*, ha estat possible comparar la situació d'aquesta comunitat autònoma amb la de la resta de l'Estat. Els resultats principals es presenten a la següent taula:

#### *Penetració de TICs el 1998*

(en %)

	Catalunya	Espanya (incl. Catalunya)
Disponibilitat de PCs a les llars	50,0	44,1
Ús individual de PCs*	45,7	36,8
Accés a Internet*	20,8	12,6
Ús de cd-rom*	26,0	18,5

\* tant a casa com a la feina, l'escola o la universitat

Quant a la demanda potencial de productes informatius en format electrònic, Catalunya està en una situació bastant més avançada que la resta de l'Estat en la disponibilitat i utilització de PCs, Internet i cd-roms, cosa que concorda amb el nombre d'empreses que serveixen informació a través de la Xarxa. No obstant això, novament, els 19 proveïdors madrilencs facturen més (2.076 M PTA) que els 27 catalans (905 M PTA), deixant a banda una resta d'Estat quasi insignificant en aquest aspecte (10 proveïdors que només facturen 40 M PTA per Internet).

## **8.3. Tendències del sector**

Aquest apartat, acomplert reunint les opinions de 29 experts espanyols que varen omplir un llarg qüestionari i posteriorment varen ser entrevistats, ha resultat molt complex,

amb una gran diversitat de matisos, impossibles de resumir aquí, malgrat no es preveure gaire més enllà del 2003. No podia ser altrament si es té en compte la gran varietat de productes i serveis, els tipus d'informació, les formes de pagament, les tècniques, les hibridacions, els ràpids canvis, les contínues fusions d'empreses, la volatilitat de la matèria objecte d'estudi (la informació), etc.

Algunes de les conclusions menys arriscades d'exposar són:

- Augmentarà molt més l'ús de la informació a través de les xarxes que no pas en el medi òptic (cd-rom, DVD, etc.). Aquest quedaria circumscribit sols a l'entorn domèstic.
- El mercat d'informació electrònica professional augmentarà d'un 8-12% anual, però disminuirà la seva quota de mercat en front de la informació per al gran públic (premsa, interès general, lleure, etc.) fins a estabilitzar-se al voltant del 25-30% del total.

La resta de conclusions varen ser constatacions i recomanacions no numèriques per a les quals remetem el lector a la bibliografia ([3], [4], [9], [10]).

#### **8.4. Estudi de cas**

Tal com s'ha dit a l'apartat 4.4., a Espanya es va escollir estudiar els hàbits d'ús de la informació en els laboratoris farmacèutics atès que constitueixen un col·lectiu que la usa extensament per sobre la mitjana del nostre país i que com a tipus empresa és relativament homogeni si salvem les diferències en dimensions, propietat nacional o multinacional, i d'altres.

Obtenir les dades no va ser fàcil: es varen enviar 130 qüestionaris als laboratoris que figuraven al *Vademecum de especialidades farmacéuticas* (publicat per *Medimedia-Medicom*) i després de moltes reclamacions telefòniques s'obtingué una resposta del 36% (malauradament bastant incompleta en molts dels apartats econòmics).

Malgrat això, l'estudi ha resultat suficientment útil com per extreure'n interessants conclusions sobre l'entorn professional, les bases de dades més utilitzades, la progressiva penetració de la informació-e, etc. El sector farmacèutic, com altres, ha experimentat grans canvis en els darrers anys degut a les compres i fusions d'empreses, i com a conseqüència de dràstiques reorganitzacions internes per tal de fer front a la competència i a les retallades d'ingressos derivades de les mesures preses per la Sanitat Pública que afecten els preus dels medicaments. Tot això ha repercutit profundament en els serveis d'informació que, generalment, han patit fortes reduccions de personal i de pressupost.

L'anàlisi concorda amb moltes de les opinions expresades pel panell d'experts (vegeu l'apartat 8.3. anterior) i reflecteix també els grans canvis de la indústria de la informació electrònica [11].

## 9. CONTINUÏTAT D'AQUEST TIPUS D'ESTUDI

La finalització del programa *Info2000* de la CE (promotor com ja s'ha dit dels *MSS-study*) va representar una ruptura d'aquesta activitat. La xarxa d'instituts nacionals que va realitzar *MSS-study II* va presentar una proposta de projecte a una convocatòria de la CE del programa *IST (Information Society Technologies)*, maig 2000, per tal de continuar les estadístiques, però amb el seu plantejament inicial no va ser acceptada i ara s'està remodelant d'acord a les indicacions rebudes de l'*Eurostat (Oficina Estadística Europea)*, per a presentar-se a la següent convocatòria.

No obstant això, en el cas d'Espanya s'espera tenir una continuïtat immediata de la medició del valor de la informació electrònica (la part de l'oferta de *MSS-study II*, que sens dubte és la més característica i especialitzada) gràcies a la participació de l'*Institut d'Estadística de Catalunya* en el projecte europeu *Eicstes (European indicators, cyberspace and the science-technology-economy system)*, 2001-2003, que lidera el *Centro de Información y Documentación Científica (Cindoc)* del *Consejo Superior de Investigaciones Científicas (Csic)*.

Una de les tasques principals que caldrà realitzar és una més actualitzada i acurada segmentació del sector, estudiant-ne les corresponents metadades (nomenclatura, definicions i mètodes de mesura) per tal d'assegurar que els resultats siguin realment comparables amb exercicis anteriors i posteriors, i amb possibles estudis que eventualment es realitzin en altres països. En els darrers anys, els instituts oficials d'estadística han vist la importància d'avaluar els continguts i han avançat considerablement en la definició de les activitats que integren l'anomenada nova economia (*Naics 2002*, discussions del *Voorburg Group*, etc.). Està previst obtenir una segmentació —que servirà després a la pràctica per medir els corresponents nínxols de mercat— bastant més detallada que la dels sistemes estadístics oficials, però evidentment es tractarà de buscar l'encaixament i la correspondència amb les classificacions estableertes per aquests.

Per últim, estem segurs que el nou programa comunitari *eContent* (2001-2004), successor d'*Info2000*, ben dotat econòmicament, cuidarà novament aquest crucial problema de les estadístiques dels continguts informatius en la societat de la informació.

## 10. REFERÈNCIES

- [1] Schwuchow, Werner; Bredemeier, Willi; Graumann, Sabine. «The markets for electronic information services in the European Economic Area. Supply, demand and information infrastructure (MSS-study I). Final Report: 1995» [en línia]. <<http://158.169.50.95:10080/imo/en/ms-study.html>>
- [2] Schwuchow, Werner; Bredemeier, Willi; Graumann, Sabine. «Manual on assessing the situation of the markets for electronic information services in the European

Economic Area (MSSStudy II)». IIE Institute for Information Economics, Hattingen / Köln; Infratest Burke, München. August 1998.

- [3] Baiget, Tomàs (1999). «Second Member States's Study of the markets for electronic information services in the European Economic Area (MSSStudy II). Spain: 1997/98» [en línia]. *Asociación Multisectorial de la Información (Asedie)*. <<http://www.asedie.es/mssstudy>>
- [4] Bredemeier, Willi; Graumann, Sabine. «Second Member States's Study of the markets for electronic information services in the European Economic Area (MSSStudy II). Final Report: 1997/98» [en línia]. <<http://www.cordis.lu/ist/ka3/publications.htm>> <<http://www.asedie.es/mssstudy/finalrep>>
- [5] Ministerio de Educación y Cultura (actualització continuada). «Libros editados en España desde 1972» [en línia]. Madrid. <<http://www.mcu.es/bases/spa/isbn/ISBN.html>>
- [6] Informa (edició anual continuada). «Sistema de análisis de balances españoles (Sabe)» [cd-rom]. Madrid: Informa, Información Económica S. A.
- [7] Dicodi [cd-rom]: «Base de datos de sociedades, consejeros y directivos» (edició anual continuada). Madrid: Equifax.
- [8] RedIRIS. Lista de proveedores de servicios Internet. <<http://www.nic.es/proveedores.html>>
- [9] Baiget, Tomàs (1999). «Tendencias del sector español de la información electrónica». *El profesional de la información*, 8, 10 (octubre), pp. 4-13.
- [10] Portela, Paloma (1999). «La información económica en España. Una revisión». *El profesional de la información*, 8, 12 (desembre), pp. 4-14.
- [11] Baiget, Tomàs (1999). «Uso de información en laboratorios farmacéuticos». *El profesional de la información*, 8, 12 (desembre), pp. 15-21.

## ENGLISH SUMMARY

### ELECTRONIC INFORMATION SERVICES MARKETS ASSESSMENT

TOMÀS BAIGET

Institut d'Estadística de Catalunya\*

*Characteristics and main results of «MSStudy II» [Study for assessing the situation of the markets for electronic information services in the European economic area (EEA)] are presented. The project, that was promoted by the Direction General XIII of the European Commission, has been carried out in Spain by the Institut d'Estadística de Catalunya (Idescat), with the cooperation of the Asociación Multisectorial de la Información (Asedie). MSStudy II, which along 1998 and 1999 has been completed simultaneously in 16 European countries, is an experimental statistics activity –although being formally listed in the Pla Estadístic de Catalunya 1997-2000– that has had the objective of obtaining disaggregated data according to the information type, kinds of distribution media (online, offline), user types, etc. The results in each country have been integrated (summer 2000) by a German experts group –the same that planned and coordinated this exercise following the initial EC mandate*

**Keywords:** Markets, electronic information, European programmes, information society, information technologies, information content.

**AMS Classification (MSC 2000):** 62P20, 91B44

---

\* Institut d'Estadística de Catalunya, Assistència Tècnica Estadística. Via Laietana, 58. 08003 Barcelona.  
Spain. E-mail: baiget@idescat.es

–Received September 1999.  
–Accepted November 2000.

## **1. BACKGROUND**

*MSSStudy II* has accomplished one of the objectives of *European Commission's Info2000* programme (1996-1999), that was to measure the development level of the electronic information market (as well as other neighbouring electronic services, like communication, e-advertising, and e-transactions) in the EEA countries (the EU plus Norway and Iceland), and to compare the results both internally and also with those of other parts of the world. Based on this exercise the *EC* could undertake the adequate actions needed to promote more qualitative and competitive services.

The study has taken into account not only the classical online hosts, newswires, financial information, etc., but also recent market new products and services like:

- multimedia products, both online and off-line (cd-rom, etc.);
- Internet;
- online services addressed directly to end consumers in private homes;
- electronic services in general, like e-mail, e-commerce, web hosting, etc., which have a great influence in the development of the information sector.

There has been a need for *MSSStudy II* of being tightly coordinated in order the participating countries used the same methodology and their results were comparable and able to be aggregated at European level.

## **2. *MSSTUDY II* OBJECTIVES**

- To offer to the electronic services market agents a strategic information as much accurate as possible;
- to give a panoramic vision of the volumes, structure and dynamics of these information markets in each country;
- to give guidelines to manage strategic aspects, as well as to assess the strengths and weaknesses;
- to identify the shortages and unbalances between offer and demand;
- to analyse economic, psychological, technical and lack of standards barriers, in the various market segments, as well as eventual monopolies and dependence situations.

Some information providers offer also other electronic services, specially in the advertising area (this is a vital activity for the viability of many publications) and the transactions (publications and other goods sales, hotel and travel reservations...), which although marginally they were interesting also to *MSSStudy II*.

### **3. GENERAL DESCRIPTION OF THE RESEARCH TASKS AND TOOLS USED**

#### **3.1. First research level: suppliers survey**

The different types of electronic information providers were surveyed (in Spain it was possible to obtain practically a census of all the existing ones), asking them their revenues, employees, benefits, etc.

#### **3.2. Second research level: demand survey**

The entire user population was screened on the basis of a representative random sample. 1.600 telephone interviews were addressed to private homes, that were carry out from a specialised call centre using a computer aided telephone interview (cati) software.

People at their homes were asked about the kind and intensity of use of electronic services both in homes, companies or organisations and schools or universities.

*Idescat* made a second run of 1.600 additional interviews in the Catalonian region only, thus making a total of 3.200.

#### **3.3. Third research level: business environment**

Due to time pressure and budget restrictions it was not possible to describe and analyse in depth all the factors involved in the national markets. The task –carried out mainly by the European coordinating team– has concentrated therefore in two aspects:

- **general markets description** from the business standpoint;
- **strengths and weaknesses** in an European benchmarking context.

Basic data at national and European level from the pre-existing *Esis project* (*European survey of information society projects and actions*) of the *European Commission* were used, thus saving considerable effort: <http://www.ispo.cec.be/esis/HomeFacts.html>

Besides the business environment description in each country one case study about a specific failure o success undertaking was carried out, specially in a market activity significant for the country. Spain chose the case «Information use in pharmaceutical laboratories», because these companies are traditionally important information users or consumers. Examples in other countries include: «External financing of electronic information services» (Belgium); «SLS with Astra, the leading satellite system for digital reception in Europe and soon beyond» (Luxemburg); «Regional electronic information service network in Kuusamo» (Finland); «UK regional press: best practices template for Europe» (UK); etc.

### **3.4. Fourth research level: national experts interviews on future trends, bottlenecks and opportunities**

Assessment on current trends (accelerating, stagnating or slowing down); identifying which new developments will really prosper; etc.

Analyses included:

- worldwide R+D and its market impact in the information sector;
- new electronic products and services, both informational and general, their applications in the other industry and service sectors;
- new focuses in national and international information policies; strengths and weaknesses (barriers, bottlenecks) of the European e-information industries and markets; and detecting competitive advantages in relation to the USA, Japan and other «information emerging» regions;
- to formulate, where possible, a pan-European consensus in order to contribute to set up of a sound electronic information market.

## **4. MAIN RESULTS**

In this article only a brief part of the results is presented. The reader is advised to refer to «10. Bibliografia» for the original reports or other articles dealing with specific topics.

### **4.1. Supplier side**

Analyses on the last years trends made by European experts show that in relation to general-interest electronic information, Europe will be dominated by a North-American oligopoly that will threaten the local regional cultures, in the same manner it is already happening with the cinema and the pop music. However, having said this finding, we remind that *MSStudy II* has focused its tasks mainly on professional information.

Although being in Catalan, the tables and graphic provided in the main part of this article are self explanatory. The European country ranking under point 8.1.1. shows the sales and market share in each country for electronic information only (excluding revenues of the information providers from other sources like printed products, consulting, training, software, etc.).

Not surprisingly, gross domestic product (GDP) and English language use are important factors related to or influencing this industry.

Spain occupies the 6<sup>th</sup> position in this table, but she would go down to the 16<sup>th</sup> place if sales per inhabitant were ranked.

All the amounts shown are decreased by the inflation rate of each country between 1994 (the year when the former *MSStudy I* was done) and 1997. In Spain this value was 10.6%.

Analysing the Spanish 232 fulfilled questionnaires (an amount that is considered to be a real census of companies producing or selling electronic information in this country) it was stated a concentration of the electronic information production in Madrid (83 companies selling 29,752 M PTA) and Catalonia (90 companies selling 6,031 M PTA –please note that the commas mark thousands–).

Some factors have to be taken into account when comparing the above European and internal Spanish figures. In these last ones:

- sales of foreign providers (imports) were not enclosed;
- inflation 1994-1997 has not been taken into account;
- sales of 12 companies located in Spain, currently owned by foreign multinational companies, have been included (according to the *MSStudy* methodology they had to be considered as foreigners and their sales treated as «imports»).

#### **4.2. Demand side**

Only people aged between 15 and 65 years was considered.

Again we ask the reader to see the table in the Catalan language part of this article. Two limitations were produced: 1) Only 12 of the 16 European countries carried out the telephone interviews with a comparable methodology. 2) Due to bureaucratic problems the interviews were done with differences of 10 months among the various countries; this time span could introduce some bias in the results.

#### **4.3. Trends**

29 Spanish experts answered a long questionnaire about the topics listed in chapter 3.4. of this English summary. Answers were put together under a unique report that was circulated again among respondents. Some of the less risky conclusions were:

- Information use through networks will increase much more than the one based on optical storage (cd-rom, DVD, etc.). The latter seems that will be confined to home environment only.
- Professional electronic information market will augment 8-12% annual, but will decrease its market share in front of the consumer information (news, general interest,

- entertainment, etc.) until to stabilise around 25-30% of the total.
- Most of the other conclusions were of qualitative nature and it would be lengthy to be explained here.

## 5. FUTURE STUDIES

Institutes that carried out *MSStudy* are still working together to put forward further coordinated work in this field.

Currently the *Institut d'Estadística de Catalunya* is participating in the European project *Eicstes* (*European indicators, cyberspace and the science-technology-economy system*), 2001-2003, coordinated by *Centro de Información y Documentación Científica (Cindoc)* of the Spanish *Consejo Superior de Investigaciones Científicas (Csic)*.

The main source of concern for measuring the electronic information markets continues to be finding a proper segmentation, producing metadata (nomenclature, definitions and measurement methods) in order to assure that the results be really comparable with previous and future exercises. In the recent years the official statistics are realising the importance of evaluating **content** and have advanced considerably in the definition of the so called new economy activities (*Naics*, *Voorburg Group*, etc.).

Finally we are sure that the new *EC* programme *eContent* (2001-2004), which supersedes *Info2000*, will care again about this crucial problem of the information society statistics.

# **Biometria**



## CÁLCULO DE PROBABILIDADES EN EL ANÁLISIS DE PERFILES GENÉTICOS COMPATIBLES

MIGUEL SÁNCHEZ GARCÍA\*

PEDRO CUESTA ALVARO\*\*

Universidad Complutense de Madrid

*Un hecho delictivo ha sido cometido por una o más personas. Para descubrir el número de presuntos delincuentes, se determina el perfil genético de la evidencia forense encontrada en el lugar del delito, mediante el estudio de marcadores de tipo STR. En el presente artículo se desarrollan diversos algoritmos. En el primero se considera la probabilidad de que n personas elegidas aleatoriamente de la población de referencia contengan un perfil genético compatible, en el segundo se calcula la evidencia del material genético y en el tercero se calcula el mínimo número de presuntos culpables seleccionado de un conjunto compatible. Finalmente, se analizan las evidencias en diversas situaciones particulares.*

**Calculation of evidence in the analysis of genetic compatible profiles**

**Palabras clave:** STR, alelo, principio de inclusión-exclusión, multigrafo, problema de recubrimiento

**Clasificación AMS (MSC 2000):** 60C05,92D20

\* Departamento de Estadística. Facultad de Medicina. Universidad Complutense de Madrid. 28040 Madrid.  
Tel.: 913941666. E-mail: mesegar@eucmos.sim.ucm.es

\*\*Centro de Proceso de Datos. Servicios Informáticos. Avda. Complutense s/n. 28040 Madrid.  
Tel.: 913944755. E-mail: pedro@sim.ucm.es

– Recibido en marzo de 1998.  
– Aceptado en octubre de 2000.

## 1. INTRODUCCIÓN

Se supone que se ha cometido un hecho delictivo sobre una o más personas, que se denotarán con el nombre genérico de *víctima*. Para extraer más información sobre el suceso, se analiza el material genético perteneciente a uno o más STR (iniciales de Short Tandem Repeat, según se expone en Foreman y otros (1997)), encontrado en el lugar donde se cometió el delito. Para cada STR, una persona aporta dos bandas alélicas  $A_i, A_j$ , que pueden ser iguales o distintas.

Suponiendo que se analiza un único STR, sea  $\mathbf{B} = \{A_1, A_2, \dots, A_q\}$  el conjunto de bandas alélicas diferentes que no pertenecen a la víctima y  $\mathbf{V} = \{A_{q+1}, \dots, A_k\}$  el conjunto de bandas alélicas pertenecientes a la víctima.  $\mathbf{B} \cup \mathbf{V}$  es el conjunto de las  $k$  bandas alélicas diferentes encontradas en el lugar del suceso.

Es claro que cuando una persona tiene una banda alélica  $A_s \notin \mathbf{B} \cup \mathbf{V}$  debe descartarse como presunto culpable. Además si  $q = 2s$ , el número de presuntos culpables debe ser al menos  $s$ ; mientras que si  $q = 2s + 1$ , dicho número es al menos  $s + 1$ , debido a que cada uno aporta como máximo dos bandas alélicas.

En el artículo se supone que se cumple la ley de equilibrio de Hardy-Weinberg (que se explica en el apartado 3 y en Louis y Dempster (1987)) y que son conocidas las probabilidades de las bandas alélicas. Con estas hipótesis se proporcionan métodos de cálculo para resolver los tres problemas siguientes:

- 1) Calcular la probabilidad de que  $n$  personas seleccionadas aleatoriamente de una población, supuestamente infinita, puedan ser presuntas culpables por presentar compatibilidad con el material genético encontrado.
- 2) Calcular la probabilidad de que un conjunto de cualquier número de personas sea compatible. El inverso de esta probabilidad se llama *evidencia*.
- 3) Hallar un subconjunto compatible, de cardinal mínimo, de un conjunto compatible de  $n$  personas.

Respecto al cálculo de las probabilidades de compatibilidad, se realizan los oportunos comentarios para el caso de que no se cumpla la ley de Hardy-Weinberg, se analicen varios STR o se desconozcan las probabilidades alélicas y deban ser estimadas por las correspondientes frecuencias.

El artículo se estructura en seis apartados. En el segundo se formulan los tres problemas, que se resuelven en los apartados tercero, cuarto y quinto. En el sexto, se calculan y analizan las probabilidades relativas al cálculo de evidencias en diversas situaciones. Finalmente, se comentan las potenciales aplicaciones forenses de estos cálculos.

## 2. PLANTEAMIENTO DE LOS PROBLEMAS

Según lo descrito en la introducción,  $n$  personas seleccionadas aleatoriamente de la población de referencia aportan  $2n$  bandas alélicas

$$\mathbf{L}_n = \{D_1, D_2, \dots, D_{2n-1}, D_{2n}\}$$

$\mathbf{L}_n$  puede tener elementos repetidos, por lo que se utiliza la denominación de lista en lugar de la de conjunto.

Para que  $\mathbf{L}_n$  sea compatible con la información genética recogida en  $\mathbf{B}$  y  $\mathbf{V}$ , se deben cumplir las dos condiciones siguientes

- c1) Todo  $D_j$  de la lista  $\mathbf{L}_n$ , es tal que  $D_j \in \mathbf{B} \cup \mathbf{V}$
- c2) Para todo  $A_i \in \mathbf{B} = \{A_1, \dots, A_q\}$ , existe un  $D_j$  de  $\mathbf{L}_n$  tal que  $A_i = D_j$ .

Si  $\mathbf{L}_n^*$  es el conjunto que se obtiene cuando se quitan de  $\mathbf{L}_n$  los elementos repetidos, las condiciones previas de compatibilidad equivalen a  $\mathbf{B} \subset \mathbf{L}_n^* \subset \mathbf{B} \cup \mathbf{V}$ .

**Problema P1].** Sea  $\mathbf{D}_n$  el suceso de que  $n$  personas, seleccionadas aleatoriamente de la población de referencia, aporten un material genético compatible con  $\mathbf{B}$  y  $\mathbf{V}$ . Calcular  $P(\mathbf{D}_n)$ .

En el apartado tercero se presentan los algoritmos para el cálculo de la probabilidad de  $\mathbf{D}_n$ , para cada  $n$ . No obstante, muchos problemas genéticos sobre Medicina Forense consisten en calcular la evidencia del material genético encontrado en el lugar del suceso. Cuando sólo se dispone de la información aportada por  $\mathbf{B}$  y  $\mathbf{V}$ , la *evidencia* es el inverso de la probabilidad de la unión de los  $\mathbf{D}_n$ ; esto es:

$$1 / P\left(\bigcup_{n=1}^{\infty} \mathbf{D}_n\right)$$

lo que lleva a formular el segundo problema.

**Problema P2].** Calcular la probabilidad del suceso  $\bigcup_{n=1}^{\infty} \mathbf{D}_n$

Puesto que los sucesos  $\mathbf{D}_n$  no son disjuntos, el cálculo, que se desarrolla en el apartado cuarto, es

$$P\left(\bigcup_{n=1}^{\infty} \mathbf{D}_n\right) = P(\mathbf{D}_1) + \sum_{n=2}^{\infty} P\left(\mathbf{D}_n \cap \left(\bigcup_{j=1}^{n-1} \mathbf{D}_j\right)^c\right)$$

El hecho de que  $n$  personas sean compatibles con el material genético encontrado en el lugar del delito, no quiere decir que no haya un subconjunto de estas personas que también sea compatible con el mismo suceso. En esta observación está la génesis del tercer problema, que se abordará en el quinto apartado de este trabajo, y que se enuncia en los siguientes términos.

**Problema P3].** Sea un conjunto de  $n$  personas  $\{I_1, I_2, \dots, I_n\}$  elegidas aleatoriamente de la población de referencia, compatible con  $\mathbf{B}$  y  $\mathbf{V}$ . Hallar un subconjunto  $\mathbf{F} \subset \{I_1, I_2, \dots, I_n\}$ , tal que  $\mathbf{F}$  también sea compatible y tenga cardinal mínimo.

Es obvio que  $P3$  siempre tiene solución, aunque puede no ser única.

### 3. ALGORITMO PARA CALCULAR LAS PROBABILIDADES DE LOS SUCESOS $\mathbf{D}_n$

Sobre cada persona de la población de referencia  $\Omega$ , se observan dos bandas alélicas  $\{A_i, A_j\}$ , que forman su sistema STR. Por tanto, sobre  $\Omega$  se pueden definir dos variables  $\xi_1$  y  $\xi_2$  tales que  $\xi_1(\omega) = A_i$  y  $\xi_2(\omega) = A_j$ . Si  $G(\omega) = (A_i, A_j)$ , la ley de Hardy-Weinberg dice que las variables  $\xi_1$  y  $\xi_2$  son probabilísticamente independientes, esto es

$$P[\omega|G(\omega) = (A_i, A_j)] = P[\omega|\xi_1(\omega) = A_i] \cdot P[\omega|\xi_2(\omega) = A_j]$$

Esta ley facilita el cálculo de probabilidades, pues sólo es necesario conocer las probabilidades de los alelos, cuyo número es menor que el de combinaciones genotípicas del STR.

Se supone que las bandas alélicas del STR analizado son  $\{AL_1, AL_2, \dots, AL_m\}$  y que  $q_j = P(AL_j)$  son las correspondientes probabilidades.

Sea

$$\begin{aligned} \mathbf{B} &= \{A_1, A_2, \dots, A_q\} = \{AL_{i_1}, AL_{i_2}, \dots, AL_{i_q}\} \quad \text{y} \\ \mathbf{V} &= \{A_{q+1}, \dots, A_k\} = \{AL_{i_{q+1}}, \dots, AL_{i_k}\}. \end{aligned}$$

Se denota por  $p_j = P(A_j) = P(AL_{i_j}) = q_{i_j}$  y  $p_v = P(\mathbf{V}) = \sum_{s=q+1}^k q_{i_s}$

Llamando  $\mathbf{T} = \mathbf{B} \cup \mathbf{V}$  y  $\mathbf{S}_j = \mathbf{T} - \{A_j\}$ , para  $j = 1, 2, \dots, q$ ; como consecuencia del principio de inclusión-exclusión, Feller (1973 Cap. IV) se verifica la siguiente igualdad:

(1)

$$\begin{aligned}
P(\mathbf{D}_n) &= P\left(\mathbf{T}^{2n} - \bigcup_{j=1}^q \mathbf{S}_j^{2n}\right) = P(\mathbf{T}^{2n}) - P\left(\bigcup_{j=1}^q \mathbf{S}_j^{2n}\right) = P(\mathbf{T}^{2n}) - \sum_{j=1}^q P(\mathbf{S}_j^{2n}) + \\
&+ \sum_{1 \leq i < j \leq q} P(\mathbf{S}_i^{2n} \cap \mathbf{S}_j^{2n}) - \sum_{1 \leq i < j < r \leq q} P(\mathbf{S}_i^{2n} \cap \mathbf{S}_j^{2n} \cap \mathbf{S}_r^{2n}) + \cdots + \\
&+ (-1)^{q-1} \sum_{j=1}^q P\left(\bigcap_{i \neq j}^q \mathbf{S}_i^{2n}\right) + (-1)^q P(\mathbf{S}_1^{2n} \cap \cdots \cap \mathbf{S}_j^{2n} \cap \cdots \cap \mathbf{S}_q^{2n})
\end{aligned}$$

Para facilitar el cálculo de probabilidades dadas por la fórmula (1), se tiene en cuenta la independencia de las bandas alélicas, tanto para una misma persona como para personas distintas:

- i)  $P(\mathbf{T}^{2n}) = [P(\mathbf{T})]^{2n} = [p_1 + p_2 + \cdots + p_q + p_v]^{2n}$
- ii)  $P(\mathbf{S}_{i_1}^{2n} \cap \mathbf{S}_{i_2}^{2n} \cap \cdots \cap \mathbf{S}_{i_r}^{2n}) = [P(\mathbf{S}_{i_1} \cap \mathbf{S}_{i_2} \cap \cdots \cap \mathbf{S}_{i_r})]^{2n}$

Se realiza el cálculo de la probabilidad de  $\mathbf{D}_n$  en dos etapas: En la primera se calculan las probabilidades de los sucesos:

$$\mathbf{S}_{i_1, i_2, \dots, i_r} = \mathbf{S}_{i_1} \cap \mathbf{S}_{i_2} \cap \cdots \cap \mathbf{S}_{i_r} = \mathbf{T} - \{A_{i_1}, A_{i_2}, \dots, A_{i_r}\}$$

y en la segunda se calcula, iterativamente, la probabilidad de  $\mathbf{D}_n$ .

Cualquier suceso  $\mathbf{S}_{i_1, i_2, \dots, i_r}$  se identifica biunívocamente con el número binario de  $q$  cifras que tiene unos en las posiciones  $i_1, i_2, \dots, i_r$  y ceros en las restantes posiciones. Esta identificación facilita el diseño del algoritmo siguiente, que calcula las probabilidades de los sucesos  $\mathbf{S}_{i_1, i_2, \dots, i_r}$ .

### 3.1. Algoritmo A1 de aplicación a las bandas alélicas B y V

Dados  $p(1) = P(A_1), \dots, p(q) = P(A_q); p(q+1) = P(\mathbf{V}) = P\{A_{q+1}, \dots, A_k\}$ .

Iniciarizar  $\mathbf{b}(1 : q) = 0$ ;

Para  $i = 1$  hasta  $2^q - 1$

    Para  $j = 1$  hasta  $q$

        Si  $b(j) = 0$ , poner  $b(j) = 1$  y Salir  $j$

        En otro caso  $b(j) = 0$

    Siguiente  $j$

$$s(i) = s(0) - \sum_{j=1}^q b(j) \cdot p(j)$$

Siguiente i

En  $\mathbf{s}(0 : 2^q - 1)$  se almacena las probabilidades de los sucesos  $\mathbf{S}_{i_1, i_2, \dots, i_r}$ .

### 3.2. Algoritmo A2 de aplicación a las bandas alélicas B y V

Se calculan las probabilidades de las sucesivas potencias cartesianas de los sucesos  $S_{i_1, i_2, \dots, i_r}$ , para evaluar las probabilidades de los sucesos  $D_n$  por la fórmula (1); para valores de  $n$  desde 1 a  $nbio$ .

Dados  $q; nbio; s(0 : 2^q - 1)$

Para  $i = 0$  hasta  $2^q - 1$ ; poner  $sc(1 - i) = s(i) \cdot s(i)$ ; Siguinte  $i$

Poner  $pD(1) = sc(1, 0)$

Para  $n = 2$  hasta  $nbio$

    Para  $i = 0$  hasta  $2^q - 1$ ; poner  $sc(n, i) = sc(n(1, i)) \cdot sc(1, i)$ ; Siguinte  $i$

$pD(n) = sc(n, 0)$

    Siguinte  $n$

Para  $n = 1$  hasta  $nbio$

$\mathbf{b}(1 : q) = 0$

    Para  $i = 1$  hasta  $2^q - 1$

        Para  $j = 1$  hasta  $q$

            Si  $b(j) = 0$ , poner  $b(j) = 1$  y Salir  $j$

            En otro caso  $b(j) = 0$

            Siguinte  $j$

$pD(n) = pD(n) + (-1)^{\text{suma}(\mathbf{b})} \cdot sc(n, i)$

        Siguinte  $i$

    Escribe  $n, pD(n)$

    Siguinte  $n$

*Nota:*

El entero  $nbio$  representa el número máximo de presuntos delincuentes, hasta el que se calculan las probabilidades de los sucesos  $D_n$ .

En  $s(i)$  se almacena la probabilidad del suceso  $S_{i_1, i_2, \dots, i_r}$  asociado con el número  $i$  representado en binario.

En  $sc(n, i)$  se almacena  $s(i)^{2n}$ .

$\mathbf{b}(1 : q)$  es una referencia a los elementos  $b(j)$ ,  $j = 1, \dots, q$ . En  $\mathbf{b}(1 : q)$  se almacena iterativamente los desarrollos binarios de los números que van desde 0 hasta  $2^q - 1$ .

La expresión  $\text{suma}(\mathbf{b})$  es el número de unos en  $\mathbf{b}(1 : q)$ .

Finalmente en  $pD(n)$  se almacena la probabilidad del suceso  $D_n$ .

## OBSERVACIONES

- a) Cuando no se puede admitir la ley de Hardy-Weinberg, se deben conocer las probabilidades conjuntas de las dos bandas alélicas ( $A_i, A_j$ ). El cálculo de la probabilidad de  $\mathbf{D}_n$  se realiza de forma idéntica al descrito, sustituyendo en la fórmula (1) el conjunto  $\mathbf{T}$  por los pares de bandas alélicas pertenecientes ambas a  $\mathbf{B} \cup \mathbf{V}$ , cada suceso  $S_j$ ,  $j \in \mathbf{B}$ , por los pares de bandas alélicas que pertenecen a  $\mathbf{B} \cup \mathbf{V} - j$ , la potencia  $2n$  por  $n$  y las probabilidades de las bandas alélicas por las de los pares de bandas.
- b) En el caso de medir varios STR, si éstos son independientes se calcularía la probabilidad por la regla del producto, mientras que si son dependientes se utilizaría la regla del producto condicional.
- c) Sea  $\hat{p}_i$  la estimación por máxima verosimilitud de las probabilidades de las bandas alélicas, cuando éstas se estiman por las frecuencias observadas en una muestra de la población de referencia.

Al sustituir en el cálculo de  $P(\mathbf{D}_n)$ ,  $p_i$  por  $\hat{p}_i$  se obtiene el estimador  $\hat{P}(\mathbf{D}_n)$ .

Si  $(a_i, b_i)$  son intervalos de confianza de  $\hat{p}_i$ , al tomar muestras aleatoriamente tal que  $(p_1^0, p_2^0, \dots, p_k^0) \in \prod_{i=1}^k (a_i, b_i)$ , se obtienen distintos valores de  $\hat{P}(\mathbf{D}_n)$ , con los que se pueden construir intervalos de confianza para  $P(\mathbf{D}_n)$ .

## EJEMPLO

En la tabla 3.1 se muestran las probabilidades  $P(\mathbf{D}_n)$ ,  $n = 1, \dots, 10$ ; en diferentes supuestos.  $P_B$  es la suma de las probabilidades de los alelos no pertenecientes a la víctima; probabilidades que se muestran en las columnas  $p_j$ ,  $j = 1, \dots, q$ ; para diferentes valores de  $q$ .

El valor  $m$  que maximiza  $P(\mathbf{D}_n)$  es el número de personas con mayor probabilidad de genotipos compatibles. En el primer supuesto de la tabla es  $m = 1$ , en el segundo  $m = 2$  y en el tercero  $m = 3$ .

## 4. ALGORITMO PARA EL CÁLCULO DE EVIDENCIAS

Puesto que los sucesos  $\mathbf{D}_n$  no son incompatibles, no se puede aplicar la fórmula de sucesos disjuntos para calcular la probabilidad de  $\mathbf{E} = \bigcup_{n=1}^{\infty} \mathbf{D}_n$ . La descomposición en sucesos disjuntos se consigue por el procedimiento:

$$\mathbf{E} = \bigcup_{n=1}^{\infty} \mathbf{D}_n = \mathbf{D}_1 \cup \left[ \bigcup_{n=2}^{\infty} \left( \mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c \right) \right]$$

**Tabla 3.1.** Valores de  $P(\mathbf{D}_n)$  en diversos supuestos

$p_V$	$p_B$	$q$					$P(\mathbf{D}_n)$									
			$p_1$	$p_2$	$p_3$	$p_4$	1	2	3	4	5	6	7	8	9	10
0,10	0,35	2	0,205	0,145			0,0595	0,0289	0,0073	0,0016	0,0003	0,0001	0,0000	0,0000	0,0000	0,0000
0,10	0,35	3	0,072	0,105	0,174		0,0000	0,0086	0,0038	0,0011	0,0003	0,0001	0,0000	0,0000	0,0000	0,0000
0,10	0,35	4	0,075	0,062	0,092	0,122	0,0000	0,0012	0,0015	0,0006	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000
0,10	0,45	2	0,176	0,274			0,0965	0,0663	0,0245	0,0080	0,0025	0,0008	0,0002	0,0001	0,0000	0,0000
0,10	0,45	3	0,148	0,142	0,161		0,0000	0,0262	0,0161	0,0064	0,0022	0,0007	0,0002	0,0001	0,0000	0,0000
0,10	0,45	4	0,110	0,143	0,091	0,106	0,0000	0,0036	0,0060	0,0036	0,0015	0,0006	0,0002	0,0001	0,0000	0,0000
0,10	0,60	2	0,235	0,365			0,1716	0,1809	0,1061	0,0553	0,0278	0,0137	0,0068	0,0033	0,0016	0,0008
0,10	0,60	3	0,293	0,199	0,108		0,0000	0,0603	0,0586	0,0384	0,0219	0,0117	0,0061	0,0031	0,0015	0,0008
0,10	0,60	4	0,164	0,131	0,148	0,157	0,0000	0,0120	0,0301	0,0277	0,0185	0,0108	0,0058	0,0030	0,0015	0,0008
0,20	0,30	2	0,099	0,201			0,0397	0,0302	0,0108	0,0032	0,0009	0,0002	0,0001	0,0000	0,0000	0,0000
0,20	0,30	3	0,120	0,105	0,075		0,0000	0,0079	0,0052	0,0020	0,0006	0,0002	0,0001	0,0000	0,0000	0,0000
0,20	0,30	4	0,083	0,094	0,054	0,069	0,0000	0,0007	0,0013	0,0008	0,0003	0,0001	0,0000	0,0000	0,0000	0,0000
0,20	0,40	2	0,185	0,215			0,0796	0,0796	0,0384	0,0154	0,0058	0,0021	0,0008	0,0003	0,0001	0,0000
0,20	0,40	3	0,136	0,108	0,157		0,0000	0,0220	0,0191	0,0102	0,0045	0,0018	0,0007	0,0003	0,0001	0,0000
0,20	0,40	4	0,070	0,087	0,134	0,109	0,0000	0,0021	0,0052	0,0043	0,0025	0,0012	0,0005	0,0002	0,0001	0,0000
0,20	0,60	1	0,600				0,6000	0,4080	0,2621	0,1678	0,1074	0,0687	0,0440	0,0281	0,0180	0,0115
0,20	0,60	2	0,375	0,225			0,1686	0,2691	0,2201	0,1547	0,1032	0,0674	0,0435	0,0280	0,0180	0,0115
0,20	0,60	3	0,274	0,120	0,206		0,0000	0,0812	0,1148	0,1040	0,0798	0,0567	0,0387	0,0258	0,0169	0,0110
0,20	0,60	4	0,098	0,177	0,155	0,170	0,0000	0,0110	0,0424	0,0572	0,0543	0,0439	0,0324	0,0228	0,0155	0,0104

De esta forma los sucesos  $\mathbf{D}_1$  y  $\mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c$ , para  $n \geq 2$ , son disjuntos.

Por tanto, el cálculo de  $P(\mathbf{E})$ , cuyo inverso es la evidencia de  $\mathbf{B}$  y  $\mathbf{V}$ , se reduce al cálculo iterativo de las probabilidades de los sucesos  $\mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c$  y a su posterior suma.

Para el cálculo de estas probabilidades se debe tener en cuenta que la persona seleccionada en el lugar  $n$ -ésimo debe aportar las bandas alélicas de  $\mathbf{B}$  que no pertenezcan a las primeras  $n - 1$  personas seleccionadas. Como una persona aporta como máximo dos bandas alélicas, en  $\mathbf{L}_{n-1}$  tienen que estar todos los elementos de  $\mathbf{B}$ , salvo uno o dos como máximo. El uno o dos que faltan deben ser aportados por la  $n$ -ésima persona.

Se denota por  $P(n, -A_i)$ ,  $1 \leq i \leq q$  la probabilidad de que  $n$  personas, elegidas aleatoriamente de la población, sean compatibles con las bandas alélicas  $\mathbf{B} - \{A_i\}$  y  $\mathbf{V}$  y por  $P(n, -\{A_i, A_j\})$ ,  $1 \leq i < j \leq q$  la probabilidad de que sean compatibles con las bandas alélicas  $\mathbf{B} - \{A_i, A_j\}$  y  $\mathbf{V}$ .

Las probabilidades de los sucesos  $\mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c$  se calculan por la fórmula:

(2)

$$pI(n) = P\left(\mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c\right) = \sum_{i=1}^q P(n-1, -A_i) [2P(A_i)P(\mathbf{B} \cup \mathbf{V}) - P(A_i)P(A_i)] + \\ + \sum_{1 \leq i < j \leq q} P(n-1, -\{A_i, A_j\}) [2P(A_i)P(A_j)]$$

Los cálculos de  $P(n-1, -A_i)$  y  $P(n-1, -\{A_i, A_j\})$  se realizan por los algoritmos A1 y A2 aplicados a las bandas alélicas  $\mathbf{B} - \{A_i\}$ ,  $\mathbf{V}$  y  $\mathbf{B} - \{A_i, A_j\}$ ,  $\mathbf{V}$ , respectivamente.

Con esta terminología el algoritmo para calcular evidencias se diseña como sigue.

### Algoritmo A3

Dado  $nmax$ , valor máximo de  $n$  para el que se calculan las probabilidades  $pI(n)$ .

Iniciarizar  $\mathbf{pI}(1 : nmax) = 0$

Calcular  $pI(1) = pD(1)$  aplicando A1 y A2

Para  $n = 2$  hasta  $nmax$

    Para  $i = 1$  hasta  $q$

        Calcular  $P(n-1, -A_i)$  aplicando los algoritmos A1 y A2 para

$nbio = n - 1$  y las bandas alélicas  $\mathbf{B} - \{A_i\}$  y  $\mathbf{V}$ .

$$pI(n) = pI(n) + P(n - 1, -A_i) \cdot [2P(A_i)P(\mathbf{B}(\mathbf{V}) - P(A_i)P(A_i))]$$

Siguiente  $i$

Para  $i = 1$  hasta  $q - 1$

Para  $j = i + 1$  hasta  $q$

Calcular  $P(n - 1, -\{A_i, A_j\})$  aplicando los algoritmos A1 y A2 para  $nbio = n - 1$  y las bandas alélicas  $\mathbf{B} - \{A_i, A_j\}$  y  $\mathbf{V}$ .

$$pI(n) = pI(n) + P(n - 1, -\{A_i, A_j\}) \cdot [2P(A_i)P(A_j)]$$

Siguiente  $j$

Siguiente  $i$

Escribe  $n, pI(n)$

Siguiente  $n$

**Nota:**

Se pueden evaluar y almacenar, por los algoritmos A1 y A2, las probabilidades:

$$P(n, -A_i), n = 1, 2, \dots, nmax, \quad \text{fijando cada banda alélica } A_i.$$

$$P(n, -\{A_i, A_j\}), n = 1, 2, \dots, nmax, \quad \text{fijando cada par } A_i, A_j, \text{ con } i < j.$$

Y después evaluar  $pI(n)$ , para  $n = 2, \dots, nmax$  por la fórmula (2).

Con esta forma de proceder se necesita más memoria para guardar datos intermedios, aunque se gana rapidez computacional.

## 5. PLANTEAMIENTO Y RESOLUCIÓN DEL PROBLEMA P3

El problema que consiste en calcular el mínimo número de personas, tal que el material genético aportado por ellas sea compatible, se formula como sigue:

**Problema P3].** Sea un conjunto de  $n$  personas  $\{I_1, I_2, \dots, I_n\}$  elegidas aleatoriamente de la población de referencia, cuyo material genético es compatible con el encontrado en el lugar del delito. Hallar un subconjunto  $\mathbf{F} \subset \{I_1, I_2, \dots, I_n\}$ , tal que  $\mathbf{F}$  también sea compatible y tenga cardinal mínimo.

El problema se resuelve hallando  $\mathbf{F} \subset \{I_1, I_2, \dots, I_n\}$  tal que  $|\mathbf{F}|$  sea mínimo y se cumpla la condición de compatibilidad c2), ya que la c1) se cumple obviamente.

Para facilitar su resolución se plantea P3 como un problema de recubrimiento. Para ello se utilizan  $n$  variables binarias; de tal forma que si  $x_j = 1$  entonces  $I_j \in \mathbf{F}$ , mientras que

si  $x_j = 0$  entonces  $I_j \notin \mathbf{F}$ . La matriz  $\mathbf{A}$  del problema de recubrimiento es de dimensión  $q \times n$  y se construye por el siguiente procedimiento.

Para cada  $i$  y cada  $j$ ,  $1 \leq i \leq q$ ,  $1 \leq j \leq n$ , el elemento  $a_{ij}$  de  $\mathbf{A}$  es:

$$a_{ij} = \begin{cases} 1 & \text{si } A_i \in \{D_{2j-1}, D_{2j}\} \\ 0 & \text{si } A_i \notin \{D_{2j-1}, D_{2j}\} \end{cases}$$

siendo  $A_i \in \mathbf{B}$  y  $\{D_{2j-1}, D_{2j}\}$  las bandas alélicas de la persona  $I_j$ .

Con estos datos, el problema de recubrimiento equivalente a  $P3$ , se formula como:

$$\begin{aligned} PR1] \quad & \min x_1 + x_2 + \cdots + x_n && \text{con } x_j \in \{0, 1\} \\ & \text{Sujeto a } a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq 1 && i = 1, 2, \dots, q \end{aligned}$$

Cada solución de  $PR1$  es compatible con  $\mathbf{B}$  y  $\mathbf{V}$  y tiene cardinal mínimo, por lo que es solución de  $P3$ . Recíprocamente cada solución de  $P3$  es una solución de  $PR1$ , de lo que se deduce que ambos problemas son equivalentes.

### 5.1. Planteamiento del problema de recubrimiento por un grafo

La información suministrada por  $\mathbf{B}$ ,  $\mathbf{V}$  y  $\mathbf{L}_n = \{D_1, D_2, \dots, D_{2n-1}, D_{2n}\}$ , bandas alélicas aportadas por  $\{I_1, I_2, \dots, I_n\}$  posibilita la construcción del multigrafo  $\mathbf{G}^* = (\mathbf{h}, \mathbf{a}^*)$ , donde  $\mathbf{h} = \{1, 2, \dots, q+1\}$  son los  $q+1$  vértices de  $\mathbf{G}^*$ . El vértice  $q+1$  se identifica con las bandas alélicas de  $\mathbf{V} = \{A_{q+1}, \dots, A_k\}$  y cada vértice restante  $i$  se identifica con la banda alélica  $A_i$ . Cada arista se corresponde con el par de bandas alélicas de cada persona de  $\{I_1, I_2, \dots, I_n\}$  mediante el siguiente proceso.

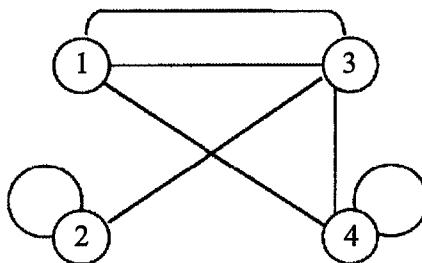
La arista asociada con las bandas alélicas  $\{D_{2j-1}, D_{2j}\}$  de  $I_j$  es  $(i, s)$  si  $D_{2j-1} = A_i$  y  $D_{2j} = A_s$ , para  $1 \leq i, s \leq q$ ; es  $(i, q+1)$  si  $D_{2j-1} = A_i$  y  $D_{2j} \in \mathbf{V}$  ó  $D_{2j} = A_i$  y  $D_{2j-1} \in \mathbf{V}$ ,  $1 \leq i \leq q$ ; y es el bucle  $(q+1, q+1)$  si  $D_{2j-1} \in \mathbf{V}$  y  $D_{2j} \in \mathbf{V}$ .

#### Ejemplo

Sea  $\mathbf{B} = \{A_1, A_2, A_3\}$ ,  $\mathbf{V} = \{A_4, A_5\}$  y siete personas  $\{I_1, I_2, \dots, I_7\}$  que aportan el material genético

$$\mathbf{L}_7 = \{A_4, A_1; A_2, A_2; A_4, A_4; A_1, A_3; A_2, A_3; A_3, A_4; A_1, A_3\}$$

El multigrafo asociado  $\mathbf{G}^* = (\mathbf{h}, \mathbf{a}^*)$  tiene cuatro vértices; esto es  $\mathbf{h} = \{1, 2, 3, 4\}$  y 7 aristas  $\mathbf{a}^* = \{(1, 4)(2, 2)(4, 4)(1, 3)(2, 3)(3, 4)(1, 3)\}$ . Este multigrafo se representa en la figura 5.1

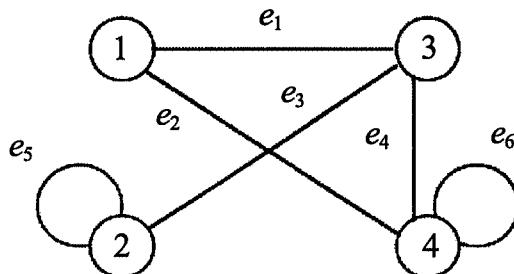


**Figura 5.1.**

En términos del multigrafo  $\mathbf{G}^*$ , la solución del problema *PR1* consiste en hallar el mínimo número de aristas que recubran los vértices  $\{1, 2, 3\}$  del grafo  $\mathbf{G}^*$ .

## 5.2. Resolución del problema

Es claro que, para hallar la solución del problema *PR1*, las aristas repetidas son redundantes, pues recubren los mismos vértices (alelos) repetidamente. Por ello se deben suprimir, en un primer paso, todas las aristas y bucles que estén repetidos. Aplicando este proceso al grafo de la figura 5.1 se obtiene el grafo de la figura 5.2.



**Figura 5.2.**

Eliminadas las aristas repetidas del multigrafo  $\mathbf{G}^* = (\mathbf{h}, \mathbf{a}^*)$  queda el unigrafo  $\mathbf{G} = (\mathbf{h}, \mathbf{a})$ , con  $\mathbf{a} \subset \mathbf{a}^*$ .

Por construcción del grafo  $\mathbf{G}$ , cada uno de sus vértices está recubierto por al menos una arista de  $\mathbf{a}$ , y por tanto tiene solución el problema de recubrimiento de los vértices  $\mathbf{h}$  de  $\mathbf{G}$  por sus aristas.

Si  $\mathbf{G}$  no es un grafo conexo, la resolución del problema de mínimo recubrimiento de los vértices por aristas se realiza resolviendo el problema para cada una de las componentes conexas de  $\mathbf{G}$ , por lo que, sin perder generalidad, se puede suponer que el grafo  $\mathbf{G}$  es conexo.

El problema de recubrimiento de los vértices de un grafo por sus aristas, se resuelve por programación lineal continua, ya que se ha logrado caracterizar el politopo del problema de recubrimiento entero asociado. Esta caracterización está desarrollada en Nemhauser y Wolsey (1999). Debido a su sencillez y brevedad se expone a continuación.

Si  $\mathbf{U} \subset \mathbf{h}$  es un subconjunto de vértices del grafo  $\mathbf{G}$  sean

$$\begin{aligned} E(\mathbf{U}) &= \{a \in \mathbf{a} / \text{Si } a = (v_i, v_j) \quad \text{entonces } v_i, v_j \in \mathbf{U}\} \\ \delta(\mathbf{U}) &= \{a \in \mathbf{a} / \text{Si } a = (v_i, v_j) \quad \text{entonces } v_i \in \mathbf{U} \text{ y } v_j \in \mathbf{h} - \mathbf{U} \text{ ó viceversa}\} \end{aligned}$$

Es claro que para recubrir un conjunto  $\mathbf{U}$  de vértices de cardinal impar, se necesitan al menos  $\left\lfloor \frac{|\mathbf{U}|}{2} \right\rfloor + 1$  aristas. De esta forma, son desigualdades válidas para el politopo recubrimiento las siguientes.

Para todo subconjunto de vértices  $\mathbf{U}$  de cardinal impar:

$$\sum_{a \in E(\mathbf{U})} x_a + \sum_{a \in \delta(\mathbf{U})} x_a \geq \left\lfloor \frac{|\mathbf{U}|}{2} \right\rfloor + 1$$

Estas desigualdades, unidas a las restricciones de grado y no negatividad de las variables, definen el politopo de recubrimiento, que es la envoltura convexa de los recubrimientos por aristas. Estas ideas se sintetizan en el teorema siguiente.

### Teorema 5.2 (Nemhauser-Wolsey)

*La envoltura convexa del problema del recubrimiento de los vértices de un grafo  $\mathbf{G} = (\mathbf{h}, \mathbf{a})$  por sus aristas, se caracteriza por las desigualdades:*

$$\begin{aligned} \sum_{e \in \delta(v)} x_e &\geq 1 \quad \forall v \in \mathbf{h} \\ \sum_{e \in E(\mathbf{U}) \cup \delta(\mathbf{U})} x_e &\geq \left\lfloor \frac{|\mathbf{U}|}{2} \right\rfloor + 1 \\ 0 \leq x_e &\leq 1 \quad \forall e \in \mathbf{a} \end{aligned}$$

para todos los subconjuntos  $\mathbf{U}$  de vértices de cardinal impar.

□

Se finaliza este apartado hallando la solución del correspondiente problema del grafo de la figura 5.2. En este ejemplo sólo hay una faceta de cardinal impar, que corresponde a  $\mathbf{U} = \{1, 2, 3\}$ .

Para la formulación del politopo de recubrimiento, se renumera las aristas, llamando  $e_1 \equiv (1, 3)$ ,  $e_2 \equiv (1, 4)$ ,  $e_3 \equiv (2, 3)$ ,  $e_4 \equiv (3, 4)$ ,  $e_5 \equiv (2, 2)$  y  $e_6 \equiv (4, 4)$ . Con esta notación la formulación del problema es:

$$\begin{aligned} \text{Mín } & x_{e_1} + x_{e_2} + x_{e_3} + x_{e_4} + x_{e_5} + x_{e_6} \\ & x_{e_1} + x_{e_2} \geq 1 \\ & x_{e_3} + x_{e_5} \geq 1 & 0 \leq x_{e_j} \leq 1 & 1 \leq j \leq 6 \\ & x_{e_1} + x_{e_3} + x_{e_4} \geq 1 \\ & x_{e_1} + x_{e_2} + x_{e_3} + x_{e_4} + x_{e_5} \geq \left\lfloor \frac{3}{2} \right\rfloor + 1 = 2 \end{aligned}$$

El problema tiene varias soluciones, cada una de ellas formada por dos aristas. Son soluciones  $\{e_1, e_5\}$ ,  $\{e_1, e_3\}$ ,  $\{e_2, e_3\}$ .

Subconjuntos compatibles de personas de tamaño mínimo son  $\{I_4, I_2\}$ ,  $\{I_7, I_2\}$ ,  $\{I_4, I_5\}$ ,  $\{I_7, I_5\}$  y  $\{I_1, I_5\}$ .

## 6. CÁLCULO DE EVIDENCIAS EN DIVERSOS SUPUESTOS PRÁCTICOS

En la tabla 6.1 se calcula el valor de  $P(\mathbf{E})$ , inverso de la evidencia, para diferentes valores de  $P_V$ ,  $P_B$  y  $q$ , número de alelos de  $\mathbf{B}$ . Se asignan valores aleatorios a las probabilidades de los  $q$  alelos de  $\mathbf{B}$  de tal forma que  $\sum_{j=1}^q p_j = P_B$ .

Se supone que éstas son las probabilidades alélicas conocidas en la población, que se cumple la ley de Hardy-Weinberg y que se analiza un único STR.

Se ha calculado  $P(\mathbf{E})$ , para un valor de  $n$  máximo  $n_{max} = 50$ . Si el número de alelos crece, es necesario aumentar  $n_{max}$  para estimar  $P(\mathbf{E})$  con más precisión.

Puede observarse lo siguiente:

1. Fijado  $q$ , al aumentar  $P_B$  aumenta  $P(\mathbf{E})$ , disminuyendo la evidencia.
2. Fijado  $P_B$  y  $q$  al aumentar  $P_V$  aumenta  $P(\mathbf{E})$ , disminuyendo la evidencia, aunque no exageradamente. Como es lógico,  $P(\mathbf{E})$  aumenta con la suma  $P_V + P_B$ , llegando a ser  $P(\mathbf{E}) = 1$  cuando  $P_V + P_B = 1$ .

**Tabla 6.1.** Valores de  $P(\mathbf{E})$  en diversos supuestos

$P_V$	$P_B$	$P_V + P_B$	$q$	$P(\mathbf{E})$	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$
0.10	0.05	0.15	1	0.01262626	0.050							
0.10	0.05	0.15	2	0.00130111	0.021	0.029						
0.10	0.05	0.15	3	0.00001279	0.011	0.024	0.015					
0.10	0.05	0.15	4	0.00000061	0.007	0.015	0.013	0.015				
0.10	0.05	0.15	5	0.00000001	0.009	0.007	0.009	0.012	0.013			
0.10	0.05	0.15	6	0.00000000	0.008	0.006	0.010	0.011	0.007	0.008		
0.10	0.05	0.15	7	0.00000000	0.008	0.005	0.009	0.005	0.005	0.009	0.009	
0.10	0.05	0.15	8	0.00000000	0.003	0.008	0.007	0.005	0.006	0.008	0.009	0.004
0.10	0.20	0.30	1	0.08080808	0.200							
0.10	0.20	0.30	2	0.01879462	0.144	0.056						
0.10	0.20	0.30	3	0.00144784	0.046	0.064	0.090					
0.10	0.20	0.30	4	0.00020712	0.058	0.068	0.037	0.037				
0.10	0.20	0.30	5	0.00001764	0.033	0.047	0.027	0.048	0.045			
0.10	0.20	0.30	6	0.00000191	0.030	0.045	0.044	0.029	0.030	0.022		
0.10	0.20	0.30	7	0.00000015	0.039	0.034	0.023	0.037	0.031	0.020	0.017	
0.10	0.20	0.30	8	0.00000001	0.015	0.022	0.026	0.035	0.014	0.032	0.023	0.033
0.10	0.35	0.45	1	0.19444444	0.350							
0.10	0.35	0.45	2	0.07476469	0.221	0.129						
0.10	0.35	0.45	3	0.01119406	0.115	0.164	0.070					
0.10	0.35	0.45	4	0.00315265	0.090	0.097	0.083	0.079				
0.10	0.35	0.45	5	0.00040041	0.097	0.040	0.097	0.040	0.077			
0.10	0.35	0.45	6	0.00009305	0.080	0.080	0.038	0.051	0.061	0.040		
0.10	0.35	0.45	7	0.00001786	0.049	0.027	0.047	0.060	0.041	0.062	0.064	
0.10	0.35	0.45	8	0.00000307	0.023	0.034	0.049	0.034	0.061	0.044	0.061	0.045
0.10	0.50	0.60	1	0.35353534	0.500							
0.10	0.50	0.60	2	0.18294196	0.207	0.293						
0.10	0.50	0.60	3	0.05507759	0.159	0.141	0.200					
0.10	0.50	0.60	4	0.01630503	0.074	0.102	0.135	0.188				
0.10	0.50	0.60	5	0.00533386	0.128	0.078	0.129	0.100	0.065			
0.10	0.50	0.60	6	0.00192446	0.080	0.082	0.092	0.095	0.089	0.063		
0.10	0.50	0.60	7	0.00046448	0.045	0.066	0.109	0.060	0.050	0.073	0.097	
0.10	0.50	0.60	8	0.00013885	0.054	0.036	0.050	0.097	0.077	0.053	0.082	0.051
0.10	0.65	0.75	1	0.55808079	0.650							
0.10	0.65	0.75	2	0.37073513	0.288	0.362						
0.10	0.65	0.75	3	0.16721177	0.139	0.252	0.259					
0.10	0.65	0.75	4	0.08777120	0.136	0.201	0.188	0.125				
0.10	0.65	0.75	5	0.04187640	0.105	0.159	0.158	0.128	0.099			
0.10	0.65	0.75	6	0.01409525	0.077	0.165	0.063	0.057	0.146	0.140		
0.10	0.65	0.75	7	0.00898127	0.082	0.117	0.073	0.071	0.085	0.132	0.090	
0.10	0.65	0.75	8	0.00461954	0.076	0.086	0.080	0.103	0.066	0.076	0.079	0.085

**Tabla 6.1.** Valores de  $P(\mathbf{E})$  en diversos supuestos (Cont.)

$P_V$	$P_B$	$P_V + P_B$	$q$	$P(\mathbf{E})$	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$
0.20	0.05	0.25	1	0.02343750	0.050							
0.20	0.05	0.25	2	0.00149686	0.031	0.019						
0.20	0.05	0.25	3	0.00002957	0.014	0.018	0.019					
0.20	0.05	0.25	4	0.00000096	0.018	0.009	0.013	0.010				
0.20	0.05	0.25	5	0.00000002	0.010	0.005	0.012	0.013	0.009			
0.20	0.05	0.25	6	0.00000000	0.009	0.006	0.011	0.007	0.007	0.010		
0.20	0.05	0.25	7	0.00000000	0.010	0.010	0.004	0.010	0.006	0.006	0.005	
0.20	0.05	0.25	8	0.00000000	0.004	0.006	0.005	0.007	0.007	0.006	0.007	0.007
0.20	0.20	0.40	1	0.12500000	0.200							
0.20	0.20	0.40	2	0.02547343	0.134	0.066						
0.20	0.20	0.40	3	0.00255608	0.049	0.096	0.055					
0.20	0.20	0.40	4	0.00030535	0.035	0.077	0.058	0.031				
0.20	0.20	0.40	5	0.00003621	0.027	0.032	0.057	0.034	0.050			
0.20	0.20	0.40	6	0.00000363	0.024	0.031	0.048	0.047	0.032	0.017		
0.20	0.20	0.40	7	0.00000047	0.034	0.027	0.031	0.032	0.015	0.030	0.032	
0.20	0.20	0.40	8	0.00000004	0.019	0.027	0.020	0.038	0.016	0.038	0.013	0.029
0.20	0.35	0.55	1	0.27343749	0.350							
0.20	0.35	0.55	2	0.09856715	0.136	0.214						
0.20	0.35	0.55	3	0.02251314	0.102	0.133	0.115					
0.20	0.35	0.55	4	0.00543509	0.098	0.097	0.062	0.093				
0.20	0.35	0.55	5	0.00104799	0.087	0.037	0.068	0.067	0.090			
0.20	0.35	0.55	6	0.00021431	0.088	0.036	0.052	0.035	0.059	0.079		
0.20	0.35	0.55	7	0.00005886	0.037	0.036	0.059	0.059	0.051	0.062	0.046	
0.20	0.35	0.55	8	0.00001011	0.049	0.032	0.060	0.051	0.033	0.053	0.052	0.021
0.20	0.50	0.70	1	0.46874999	0.500							
0.20	0.50	0.70	2	0.24653688	0.292	0.208						
0.20	0.50	0.70	3	0.08907496	0.214	0.177	0.109					
0.20	0.50	0.70	4	0.03167564	0.163	0.172	0.092	0.073				
0.20	0.50	0.70	5	0.01271238	0.138	0.118	0.104	0.078	0.062			
0.20	0.50	0.70	6	0.00481702	0.050	0.104	0.085	0.111	0.072	0.077		
0.20	0.50	0.70	7	0.00173363	0.053	0.071	0.072	0.045	0.087	0.067	0.105	
0.20	0.50	0.70	8	0.00070759	0.052	0.071	0.044	0.062	0.071	0.060	0.059	0.081
0.20	0.65	0.85	1	0.71093748	0.650							
0.20	0.65	0.85	2	0.48876294	0.215	0.435						
0.20	0.65	0.85	3	0.30504636	0.142	0.225	0.283					
0.20	0.65	0.85	4	0.17157925	0.082	0.211	0.153	0.203				
0.20	0.65	0.85	5	0.09504163	0.190	0.075	0.084	0.193	0.107			
0.20	0.65	0.85	6	0.05957792	0.083	0.054	0.113	0.127	0.133	0.140		
0.20	0.65	0.85	7	0.03566955	0.055	0.108	0.116	0.120	0.060	0.089	0.103	
0.20	0.65	0.85	8	0.02045982	0.065	0.078	0.105	0.076	0.043	0.119	0.081	0.082

3. Fijado  $P_B$ , destaca el fuerte decrecimiento de  $P(E)$ , al aumentar el número de alelos  $q$ ; como consecuencia el aumento de la evidencia es grande.
4. Incluso para probabilidades grandes,  $P_V = 0.2$  y  $P_B = 0.65$ , si el número de bandas alélicas es superior a 4, la probabilidad  $P(E)$  es pequeña y por tanto la evidencia es grande.
5. Como se infiere de 4., supuesto la independencia de los sistemas, sería necesario observar pocos sistemas polimórficos para obtener evidencias mayores que un millón, cifra superior a la evidencia que las leyes de cualquier país fijan, para que sea tenida en cuenta a efectos legales.

## 7. CONSIDERACIONES SOBRE APLICACIONES PRÁCTICAS

El artículo tiene una gran relevancia para las aplicaciones prácticas, debido a que permite calcular, de forma rápida, la evidencia del material genético encontrado en el lugar donde se cometió el hecho delictivo; bien para un solo sistema STR, o para más de uno; utilizando, en este último supuesto, la regla de la probabilidad producto condicional o independiente, según los casos. Además, debido al carácter general de la formulación de los problemas, sus posibilidades de aplicación son muy amplias.

Como consecuencia de los cálculos, cuando la evidencia es pequeña, el material genético tiene escasa relevancia en el esclarecimiento del caso, y se puede poner en duda la realización de pruebas genéticas a los presuntos culpables. Sin embargo, ante una gran evidencia del material genético, sería recomendable la realización de dichas pruebas.

La utilización, en el cálculo de la evidencia, de las probabilidades alélicas estimadas, en función de las frecuencias de los datos muestrales disponibles, no supone ningún inconveniente en la práctica, puesto que se suele mantener un cierto orden entre las magnitudes de la verdadera evidencia y de la evidencia estimada. Por ello, los métodos de cálculo expuestos en el artículo son también de gran utilidad práctica, en este supuesto.

La mayor, y principal, ventaja de la aplicación de los métodos de la genética a casos delictivos es la de decidir sobre la inocencia de posibles sospechosos, cuando su ADN no coincide con el hallado en el lugar del delito. Salvo en el caso de mutaciones o errores en las medidas, la decisión sobre la inocencia sería, en estos casos, segura.

Uno de los mayores inconvenientes es precisar el espacio poblacional de referencia, en el que se encuentran los verdaderos culpables.

También puede ser un verdadero inconveniente la utilización de la independencia entre sistemas en el cálculo de probabilidades. Este hecho suele dar lugar, muy frecuentemente,

mente, a una excesiva sobreestimación de la verdadera evidencia. Por otra parte, no se suele disponer, en las bases de datos genéticas, de información suficiente para calcular, con cierta precisión, las probabilidades que se necesitan para estimar el producto condicional. En nuestra opinión, estos son los mayores inconvenientes con que podemos encontrarnos en la práctica.

#### **AGRADECIMIENTOS**

Los autores están muy agradecidos al Dr. Ruiz de la Cuesta, así como al Dr. Arroyo, ambos pertenecientes al Departamento de Medicina Legal de la Universidad Complutense de Madrid, pues el primer problema surgió de las conversaciones que los autores mantuvieron con ellos. También están muy agradecidos a los valiosos comentarios de dos revisores anónimos, que han servido para mejorar notablemente la redacción de este artículo.

#### **BIBLIOGRAFÍA**

- Feller, W. (1973). *Introducción a la Teoría de las Probabilidades y sus Aplicaciones*. Limusa-Wiley.
- Foreman, L.A.; Smith, A.F.M. and Evett, I.W. (1997). «A Bayesian Approach to validating STR Multiplex Databases for use in Forensic Casework». *Int. J. Legal Med.*, 110, 244-250.
- Louis, E.J. and Dempster, E.R. (1987). «An Exact Test for Hardy-Weinberg and Multiple Alleles». *Biometrics*, 43, 805-811.
- Nemhauser, G.L. and Wolsey, L.A. (1999). *Integer and Combinatorial Optimization*. John Wiley.

## **ENGLISH SUMMARY**

### **CALCULATION OF EVIDENCE IN THE ANALYSIS OF GENETIC COMPATIBLE PROFILES**

MIGUEL SÁNCHEZ GARCÍA\*

PEDRO CUESTA ALVARO\*\*

University Complutense of Madrid

*One or more people have committed an offence. To discover the number of presumed delinquents, the genetic profile of the forensic evidence found in the place of the offence has been analysed, by means of the study of STR markers systems. Several algorithms are developed for that purpose in this paper. In the first one we consider the probability that  $n$  individuals, chosen at random from the reference population, contain a compatible genetic profile. In the second, the evidence of the genetic material is calculated, and in the third, the minimum number of presumed guilties, selected from a compatible group, is found. Finally we show and analyse evidences in several particular situations.*

**Keywords:** STR, allele, inclusion-exclusion principle, multigraph, set covering problem

**AMS Classification (MSC 2000):** 60C05,92D20

---

\* Departamento de Estadística. Facultad de Medicina. Universidad Complutense de Madrid. 28040 Madrid.  
Tel.: 913941666. E-mail: mesegar@eucmos.sim.ucm.es

\*\*Centro de Proceso de Datos. Servicios Informáticos. Avda. Complutense s/n. 28040 Madrid.  
Tel.: 913944755. E-mail: pedro@sim.ucm.es

– Received March 1998.  
– Accepted October 2000.

A crime has been committed against one or more persons, who will be referred to as *victim*. The genetic profile belonging to one or more STR (Short Tandem Repeat), found in the crime scene, is analyzed. For each STR, a person supplies two alleles  $A_i, A_j$ , that can be either equal or distinct.

For a single STR, let  $\mathbf{V} = \{A_{q+1}, \dots, A_k\}$  the set of different alleles that belong to the victim and  $\mathbf{B} = \{A_1, A_2, \dots, A_q\}$  the alleles that do not.  $\mathbf{B} \cup \mathbf{V}$  is the set of  $k$  different alleles found in the scene of the crime.

A group of  $n$  persons, randomly selected from a population, supplies a list of  $2n$  alleles,  $\mathbf{L}_n = \{D_1, D_2, \dots, D_{2n-1}, D_{2n}\}$ .  $\mathbf{L}_n$  is *compatible* with the genetic information picked up in  $\mathbf{B}$  and  $\mathbf{V}$ , when the following two conditions are matched.

- c1) Any  $D_j$  of the list  $\mathbf{L}_n$ , is such as  $D_j \in \mathbf{B} \cup \mathbf{V}$
- c2) For any  $A_i \in \mathbf{B} = \{A_1, \dots, A_q\}$ , there is  $D_j$  of  $\mathbf{L}_n$  such as  $A_i = D_j$

If  $\mathbf{L}_n^*$  is the set obtained removing from  $\mathbf{L}_n$  the repeated elements, the previous compatibility conditions are equivalent to  $\mathbf{B} \subset \mathbf{L}_n^* \subset \mathbf{B} \cup 'V'$ .

In the present article the following three problems are formulated and solved.

**Problem P1].** Let  $\mathbf{D}_n$  be the event that  $n$  persons, randomly selected from a population, provide a genetic profile that is compatible with  $\mathbf{B}$  and  $\mathbf{V}$ . Calculate  $P(\mathbf{D}_n)$ .

**Problem P2].** Calculate the probability of the event  $\bigcup_{n=1}^{\infty} \mathbf{D}_n$ , whose inverse is the evidence.

**Problem P3].** Let  $\{I_1, I_2, \dots, I_n\}$  be a set of  $n$  persons, whose genetic information is compatible with  $\mathbf{B}$  and  $\mathbf{V}$ . Find a subset  $\mathbf{F} \subset \{I_1, I_2, \dots, I_n\}$ , such as  $\mathbf{F}$  is the smallest compatible subset.

To solve P1 the Probability Theory Inclusion-Exclusion principle is used. Denoting  $\mathbf{T} = \mathbf{B} \cup \mathbf{V}$  and  $\mathbf{S}_j = \mathbf{T} - \{A_j\}$ , for  $j = 1, 2, \dots, q$ ; the calculations are performed through the formula:

$$\begin{aligned}
P(\mathbf{D}_n) &= P\left(\mathbf{T}^{2n} - \bigcup_{j=1}^q \mathbf{S}_j^{2n}\right) = P(\mathbf{T}^{2n}) - P\left(\bigcup_{j=1}^q \mathbf{S}_j^{2n}\right) = P(\mathbf{T}^{2n}) - \sum_{j=1}^q P(\mathbf{S}_j^{2n}) + \\
&+ \sum_{1 \leq i < j \leq q} P(\mathbf{S}_i^{2n} \cap \mathbf{S}_j^{2n}) - \sum_{1 \leq i < j < r \leq q} P(\mathbf{S}_i^{2n} \cap \mathbf{S}_j^{2n} \cap \mathbf{S}_r^{2n}) + \dots + \\
&+ (-1)^{q-1} \sum_{j=1}^q P\left(\bigcap_{i \neq j}^q \mathbf{S}_i^{2n}\right) + (-1)^q P(\mathbf{S}_1^{2n} \cap \dots \cap \mathbf{S}_i^{2n} \cap \dots \cap \mathbf{S}_q^{2n})
\end{aligned}$$

In order to make calculations easy, two algorithms have been developed to evaluate  $P(\mathbf{D}_n)$  in a fast way, taking into account the existing correspondence between the terms of the previous formula and the binary decomposition of the numbers lower than  $2^q$ .

The problem  $P2$  can be solved bearing in mind that

$$\mathbf{E} = \bigcup_{n=1}^{\infty} \mathbf{D}_n = \mathbf{D}_1 \cup \left[ \bigcup_{n=2}^{\infty} \left( \mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c \right) \right]$$

This way the events  $\mathbf{D}_1$  and  $\mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c$ , for  $n \geq 2$ , are disjoint.

The probabilities are calculated through the formula

$$\begin{aligned}
P\left(\mathbf{D}_n \cap \left( \bigcup_{j=1}^{n-1} \mathbf{D}_j \right)^c\right) &= \sum_{i=1}^q P(n-1, -A_i) [2P(A_i)P(\mathbf{B} \cup \mathbf{V}) - P(A_i)P(A_i)] + \\
&+ \sum_{1 \leq i < j \leq q} P(n-1, -\{A_i, A_j\}) [2P(A_i)P(A_j)]
\end{aligned}$$

$P(n, -A_i)$ ,  $1 \leq i \leq q$  is the probability that  $n$  persons, chosen at random from the population, are compatible with the alleles  $\mathbf{B} - \{A_i\}$  and  $\mathbf{V}$ .

$P(n, -\{A_i, A_j\})$ ,  $1 \leq i < j \leq q$  is the probability that  $n$  persons are compatible with the alleles  $\mathbf{B} - \{A_i, A_j\}$  and  $\mathbf{V}$ .

The calculations of these probabilities are performed by the algorithms for the first problem, respectively applied to  $\mathbf{B} - \{A_i\}$ ,  $\mathbf{V}$  and  $\mathbf{B} - \{A_i, A_j\}$ ,  $\mathbf{V}$ .

Problem  $P3$  is modeled by a multigraph. Vertices are identified with the alleles of  $\mathbf{B}$ , to which a new vertex identifying the alleles of  $\mathbf{V}$  is added. Each person, with alleles  $A_i, A_j$ , has correspondence with an edge in the multigraph. The goal is to solve a set covering problem, stated as finding the minimum number of edges that covers all hypergraph vertices.

The problem is solved by set covering problem polytopes characterization, developed by Nemhauser and Wolsey.

A computer program has been developed for the algorithms of the problems  $P1$  and  $P2$ . Several tables of results are shown in the article, together with examples of each problem and comments about their practical advantages and disadvantages.

## **Secció Docent i Problemes**



## **SECCIÓ DOCENT I PROBLEMES**

La «Secció docent i problemes» té l'objectiu de publicar articles de caire docent, difícilment publicables en revistes de recerca. A cada número de *Qüestiió* s'inclouen d'un a tres problemes i les solucions es donen en el número següent.

Els lectors poden 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.



## SOLUCIÓ AL PROBLEMA PROPOSAT AL VOLUM 24 N. 2

### PROBLEMA N. 86

Anem a proposar una funció de distribució bivariant tal que el coeficient de correlació de Spearman sigui  $\rho_s = 0$ , però que les variables aleatòries siguin estocàsticament dependents.

Considerem el polinomi

$$K(x) = 2x^3 - 3x^2 + x,$$

que verifica

$$K(0) = K(1) = 0.$$

Definim la funció de distribució de les variables aleatòries  $U, V$  com segueix:

$$F(u, v) = u \cdot v + \theta K(u) K(v) \quad 0 \leq u, v \leq 1,$$

essent  $0 \leq \theta \leq 2$  un paràmetre. Es verifica

$$F(0, 0) = F(u, 0) = F(0, v) = 0$$

$$F(u, 1) = u$$

$$F(1, v) = v$$

$$F(1, 1) = 1$$

$$\frac{\partial^2}{\partial u \partial v} F(u, v) = 1 + \theta (6u^2 - 6u + 1)(6v^2 - 6v + 1) \geq 0$$

Per tant  $F(u, v)$  és una funció de distribució amb marginals uniformes sobre  $(0, 1)$ . La funció de distribució d' $U$  és  $F_U(u) = u$ , i la de  $V$  és  $F_V(v) = v$ . Es verifica:

$$\int_0^1 K(u) du = \left[ 2\frac{u^4}{4} - 3\frac{u^3}{3} + \frac{u^2}{2} \right]_0^1 = 0$$

El coeficient de correlació de Spearman entre  $U$  i  $V$  és el coeficient de correlació ordinari entre  $F_U(U) = U$  i  $F_V(V) = V$ , és a dir

$$\begin{aligned} \rho_s &= 12 \int_0^1 \int_0^1 (F(u, v) - u \cdot v) du dv \\ &= 12\theta \int_0^1 \int_0^1 K(u) \cdot K(v) du dv = 12\theta \int_0^1 K(u) du \cdot \int_0^1 K(v) dv = 0. \end{aligned}$$

Llavors  $\rho_s = 0$ , qualsevol que sigui  $\theta$ . D'altra banda,  $U$  i  $V$  són estocàsticament dependents, excepte en el cas  $\theta = 0$  en què  $F(u, v) = u \cdot v$  i hi ha independència estocàstica. Per tant, dues variables aleatòries poden ser dependents i no obstant  $\rho_s = 0$ .

C.M. Cuadras  
Universitat de Barcelona

## PROBLEMA PROPOSAT

### PROBLEMA N. 87

En la llamada lotería genovesa, los jugadores compraban billetes que contenían números del 1 al 90. Se podían comprar billetes con 1, 2, 3, 4 ó 5 números. El sorteo consistía en extraer 5 números. Ganaban los jugadores en cuyos billetes se hallaban todos los números que habían sido extraídos. Por ejemplo, si un jugador tenía un billete con tres números: 8, 21, 49 y había sido extraídos los números 3, 8, 21, 49 y 63, el jugador ganaba; pero si los números extraídos eran el 3, 7, 21, 49 y 63, el jugador perdía, pues el 8 no se encuentra entre los números extraídos.

La distribución de los premios a los posibles ganadores era:

billetes con 1 número:	15 veces el importe del billete
billetes con 2 números:	270 « «
billetes con 3 números:	5.500 « «
billetes con 4 números:	75.000 « «
billetes con 5 números:	1.000.000 « «

Muchos jugadores jugaban a 2 y a 3 números, pero muy pocos ganaban, y nadie consiguió premio comprando billetes de 4 y 5 números.

Comprobar este hecho calculando la probabilidad de ganar en cada una de las apuestas con 1, 2, 3, 4 y 5 números.

C.M. Cuadras  
Universitat de Barcelona



# **Comentaris de llibres**



***Problemas de Probabilidades y Estadística***, segunda edición

**Volumen I: Probabilidades**, 415 páginas + tablas + figuras

**Volumen II: Inferencia Estadística**, 454 páginas + tablas + figuras

**Carles M. Cuadras**

Ediciones Universitarias de Barcelona, EUB, Barcelona, 1999 (Volumen I) y 2000 (Volumen II)

La primera edición de *Problemas de Probabilidades y Estadística* fue publicada en 1971 por el Laboratorio de Cálculo de la Universidad de Barcelona, en un único volumen. Las siguientes ediciones fueron publicadas por dos editoriales distintas hasta la primera edición EUB. Esta segunda edición EUB es una reedición completamente revisada y ampliada de la primera.

El Volumen I, íntegramente dedicado a la Probabilidad, contiene 15 capítulos, que tratan de los principales temas de la probabilidad y de las variables aleatorias, incluyendo las cadenas de Markov. Contiene un amplio resumen de teoría, casi 200 problemas resueltos y cerca de 300 enunciados de problemas. Respecto de las ediciones anteriores, destaca la inclusión del tema de las distribuciones con marginales dadas, las correlaciones máxima y mínima, la función generatriz de momentos y un número significativo de nuevos problemas.

El Volumen II, íntegramente dedicado a la Estadística, se estructura en 14 capítulos que tratan temas clásicos de la Estadística como muestreo, estimación, contraste de hipótesis, análisis de la varianza y estadística no paramétrica y contiene unos 140 problemas resueltos y casi 200 enunciados. Esta edición presenta algunas novedades respecto de la anterior: una referencia al metaanálisis, al tratamiento estadístico del caos y a la inferencia multivariante, así como una tabla con los principales estimadores y un índice analítico.

La obra es muy completa, está bien presentada y puede ser de gran utilidad para estudiantes universitarios de carreras científicas, técnicas, económicas y empresariales.

J.M. Oller

## *Las Matemáticas del Siglo XX. Una mirada en 101 artículos*

Editorial: Sociedad Canaria Isaac Newton de Profesores de Matemáticas  
NIVOLA, libros y ediciones, S.L., 2000

Editor-coordinador: Antonio Martinón

Como su título indica, esta obra publicada en el Año Mundial de las Matemáticas contiene, a lo largo de 101 artículos, una amplia visión de las matemáticas del siglo XX. Junto a artículos y referencias biográficas de los matemáticos más importantes del pasado siglo (Banach, Cantor, Cartan, Dieudonné, Eilenberg, Erdős, Fréchet, Gödel, Hardy, Hilbert, Klein, Kolmogorov, Levy, Mandelbrot, von Neumann, Noether, Poincaré, Polya, Ramanujan, Riesz, Russell, Schwartz, Selberg, Thom, Turing, Volterra, Whitehead, Wiener, Wiles, etc.) y de físicos teóricos cuyos resultados influyeron o estuvieron influídos por las matemáticas (Bohr, de Broglie, Dirac, Einstein, Eisenberg, Planck, Schrödinger), se incluyen numerosos artículos temáticos sobre los aspectos más interesantes de las matemáticas (los 23 problemas de Hilbert, incluyendo la hipótesis del continuo, algunas paradojas famosas, las consecuencias del axioma de la elección, las conjeturas de Poincaré y Kepler, las onditas, el teorema de Gödel, etc.). Se dedican también tres artículos a glosar la obra de destacados matemáticos españoles (Puig Adam, Rey Pastor, Santaló), incluyendo interesantes notas bibliográficas.

Casi todos los temas en que las matemáticas han intervenido son tratados en diferentes artículos por un total de 106 autores (arquitectura, economía, mecánica cuántica, teoría de la relatividad, estudio del clima, paradojas en política, estudio del caos).

Mención especial merecen los artículos dedicados a la didáctica de las matemáticas (resolución de problemas, formación de profesores, libros de texto, las matemáticas como disciplina científica, el papel de los ordenadores). Se dedican también algunos artículos a las sociedades profesionales y académicas de matemáticas, su estructura y funciones, con especial referencia a las sociedades españolas. Otros artículos explican la historia del Courant Institute, del colectivo Bourbaki y de la medalla Fields.

Los artículos de biometría, estadística y temas relacionados son seis, que abordan diversos aspectos importantes (análisis estocástico en economía, axiomática de la probabilidad, análisis de datos, biología matemática, diseños experimentales, lógica borrosa, teoría de la información), con una referencia a los autores más importantes (Fisher, Mahalanobis, Pearson, Shannon, Tukey, Wald, Zadeh). Se incluye también un artículo dedicado a la educación en estadística, que debe proporcionar una forma de razonar adaptada a la sociedad de la información. Se echa en falta una mención a Rao, sobre todo teniendo en cuenta que uno de los artículos se fundamenta en la primera parte del

capítulo 3 de la versión española de «Estadística y Verdad», de C.R. Rao, editado por PPU, Barcelona, 1994.

Se dedican seis artículos a la investigación operativa (optimización combinatoria, decisión, teoría de juegos, teoría de la elección social, algoritmos genéticos y grafos), con referencias a los principales autores (Edmonds, Ford, Fulkerson, Holland).

Los ordenadores y la informática son motivo de varios artículos que destacan su enorme influencia en las matemáticas (problema de los cuatro colores, teoría de sistemas, cibernetica, máquina de Turing, álgebra computacional, cálculo numérico).

Contiene dos artículos dedicados a las mujeres matemáticas (Noether, Robinson) con notas biográficas y comentarios sobre sus principales aportaciones, lo que constituye otro acierto de la obra, y muchos otros temas que influyeron en el siglo XX y que siguen de actualidad (bomba atómica, nanoelectrónica, psicología, fractales, viajes espaciales, etc.), incluyendo muchas direcciones web.

El artículo 101, y último, destaca la generalidad e importancia de las matemáticas, comenta algunos problemas pendientes y hace consideraciones sociológicas sobre el compromiso de esta disciplina en lograr un mundo mejor.

En resumen, se trata de una obra objetiva, amena y completa, muy recomendada para matemáticos y científicos, profesores y alumnos, interesados en la historia reciente, la belleza y la actualidad de las matemáticas.

C.M. Cuadras  
Universitat de Barcelona



## **Ressenyes d'activitats institucionals**



# Sociedad Española de Biometría



Región Española  
de la Sociedad Internacional de Biometría IBS  
Spanish Region  
of the International Biometric Society

<http://www.iata.csic.es/ibsresp>

La Sociedad Española de Biometría/Región Española de la Sociedad Internacional de Biometría (abreviadamente SEB o REsp) tiene como objetivos promover, impulsar y difundir el desarrollo y la aplicación de los métodos matemáticos y estadísticos a la biología, medicina, psicología, farmacología, agricultura y otras ciencias afines (ciencias relacionadas con los seres vivos). Cualquier profesional o alumno de estas disciplinas puede ser miembro de la SEB.

## Consejo Directivo

<i>Presidenta:</i>	Guadalupe Gómez Melis (Biología)
<i>Vicepresidente:</i>	María Jesús Bayarri García (Medicina)
<i>Secretario y Tesorero:</i>	Fernando López Santoveña (Agronomía)
<i>Vocal en calidad de</i>	
<i>Miembro del Consejo de la IBS:</i>	Emilio A. Carbonell Guevara (Agronomía)
<i>Vocales:</i>	Juan Luis Chorro Gascó (Psicología) Juan Ferrández Ferragud (Biología) Purificación Galindo Villardón (Medicina) Eduardo García Cueto (Psicología) José Luis González Andújar (Agronomía) Alex Sánchez Plá (Biología)
<i>Corresponsal de la REsp en el</i> <i>«Biometric Bulletin» de la IBS:</i>	María Luz Calle Rosingana

## La VIII CONFERENCIA ESPAÑOLA DE BIOMETRÍA

se celebrará en Pamplona

los días 28, 29 y 30 de marzo 2001.

La información preliminar puede ser consultada en

<http://www.unavarra.es/directo/congresos/apoyo/biometria.htm>



**The TES Institute  
Training of European Statisticians**

**GENERAL INTRODUCTION**

The TES Institute is a non-profit making association of 17 European National Statistical Institutes and the University Centre of Luxembourg. It offers a post-graduate vocational training programme for statisticians and other groups working in the statistical environment. The TES Institute today boasts a solid foundation of statistical training experience, built up since the beginning of the nineties.

Our mission is to provide world-wide cutting edge lifelong international training for professional statisticians. The training emphasises the inter country know-how transfer and focuses on the dissemination of best practices in terms of providing statistics compliant with both European and International Statistical System standards.

The training programmes of the TES Institute are designed for both public and private sector statisticians in the broadest sense of the word i.e. university graduates of any discipline involved in statistical work within Statistical Institutes, European Institutions, Government Agencies, National or private Banks, Enterprises, ...

Training methods are problem-oriented and based on a twofold track approach combining both teaching of theoretical concepts and practical sessions using real life cases. Moreover, the «learning-by-doing» process provides an opportunity to respond in an effective way to recent scientific and technological developments as well as new requirements from the labour market.

The training programmes offered by the TES Institute provide both theoretical and practical background but the courses have all a very strong applied character.

These programmes also offer participants the opportunity to meet colleagues from all over Europe and other countries since the TES Institute has extended its activities to the Central European, Mediterranean Basin and TACIS countries.

The above characteristics represent the basic conditions to acquire sharper competence in their work environment and highlight the European dimension of their activity.

After ten years of existence, the programme became entire part of the statistical world. For the time being, around 500 participants coming from more than 30 countries are trained every academic year. Such an interest is mainly due to the large number of courses on offer. Indeed, the TES portfolio comprises more than 80 courses of short duration all at post-graduate level.

The TES Institute offers a coherent set of interrelated courses in the following domains:

- Official Statistics: General Issues (OSG)
- Official Economic Statistics (ECO)
- Official Social Statistics (SOC)
- Data Collection and Survey Methodology (DAT)
- Applied Statistical Analysis (ASA)
- Publication and Dissemination of Statistics (PDS)
- Management in a Statistical Institute (MSI)
- Statistical Information Systems (SIS)

To reach the widest audience possible, the TES Institute currently offers courses in English, French, German, Arabic and Russian.

After a few years of co-operation with the Central European countries, the TES Institute has recently extended the co-operation to the MEDSTAT and TACIS region. Such an internationalisation is the direct result of the growing importance of training as a part of the current technological and intellectual development. Therefore, as far as statistics and economics are concerned, it is of the utmost importance to extend the best national practices to an international level.

It is obvious that the TES programmes should be considered as a complement and not a substitute to the training provided at national level.

In brief, one may say that by offering training opportunities which are complementary to the ones provided at national level, the TES Institute is offering a new approach of the subsidiarity concept.

## **TRAINING**

The Core Programme 2001 offering 26 courses will run from January to November 2001. The overview of the courses can be found at the end of the present paper.

Beside the execution of the subsidised annual vocational training programme –which is only one of our many activities– the TES Institute is active through the PHARE programme for Central European Countries, through the MEDSTAT programme in the countries of the Mediterranean Basin and through the TACIS programme in a number of CIS countries.

## **CONSULTING**

In addition to the vocational training activities, we will continue to develop and extend our consulting activities in the above mentioned regions and elsewhere in order to maintain and solidify our position in the market of professional training and staff development for statisticians. These consulting activities consist in the training of trainers, the setting up of training centres, reform of training policy and curricula. For example, the TES Institute has also been involved in consulting activities with Ukraine and the Russian Federation for the set-up of training centres and for drafting recommendations for the improvement of the vocational training system for the Russian Statistical Agency.

## **RESEARCH**

In various Research Institutes, Universities and National Statistical Institutes of the Member States of the European Union and other countries, research is taking place covering statistical methodology and statistical techniques such as data collection methods, data transmission, storage and warehousing, statistical data analysis, etc. TES is confident that, being closely involved in dissemination of the results of activities in the field of research and technological development, will not only help the Commission in its future policy on R&D but will also contribute to a better communication between the various players in statistical research. As a result, the TES Institute has decided to play an active role in European research projects in various fields of statistics.

Within this framework, the TES Institute is participating in a consortium, in the 5th Framework Programme of the Commission, focusing on the creation and the development of on-line training courses (VL-CATS Project). On the basis of possibilities offered by existing tools, the VL-CATS project will provide access over the Internet, to teaching and educational material with a special focus on Official Statistics. Through this project, experts will be given the opportunity to share and disseminate their statistical knowled-

ge on a wider scale. Along with EU and EFTA countries, statisticians from accessing countries and other Central European countries will also be given the opportunity to produce statistics compliant with the European Statistical System standards.

The main objectives of VL-CATS Project are:

- To create a virtual library of all reference material available in Official Statistics;
- To define and implement standards for the development of distance courses in the various fields of Official Statistics;
- To develop a controlled environment that will support the virtual library and give selective access to training courses;
- To develop a service for update and quality assurance of VL-CATS.

## PUBLICATIONS

The TES Institute is regularly publishing articles on its current activities in periodic Newsletter of several Statistical Institutes and some statistical journals as *Qüestió* (Quaderns d'Estadística i Investigació Operativa), edited by the Institut d'Estadística de Catalunya.

The TES Institute has started the production of TES Manuals on subjects covered by the vocational training programme:

- The first manual available at the TES institute presents *The role of statistics in a democracy*.
- The second manual to be published soon will cover the *Index numbers for spatial and temporary comparisons*.
- Further manuals will cover topics of *sampling techniques, seasonal adjustment methods*.

## TES NEWSLETTER

Beginning of February, the TES Institute has disseminated the first issue of its Newsletter named «Facts & Visions». This newsletter should be considered as an important information tool between the TES Institute and its partners from all over Europe focusing on matters related to training and staff development in the statistical world.

We hope that our partners will use «Facts & Visions» to report on their own activities. So, may we encourage you to use your keyboards for European information purposes.

Copies of «Facts & Visions» are available at the TES Institute as well as guidelines for the submission of articles.

## **GENERAL INFORMATION**

For further details on any of the above mentioned topics, please contact directly Ms. Valérie Vandewalle (co-ordinator for *Evaluation and Information*):

Phone: (352) 29.85.85.34

Fax: (352) 29.85.29

E-mail: [vvandewalle@tes-institute.lu](mailto:vvandewalle@tes-institute.lu)

<b>Code</b>	<b>Course Title</b>	<b>Days</b>	<b>Course Leader</b>	<b>Location</b>	<b>From</b>	<b>To</b>
DAT-202/2001	Use of Auxiliary Information in Sampling Surveys	4	Olivier Sautory	Paris	20-feb-01	23-feb-01
ASA-201/2001	Seasonal Adjustment Methods	5	Agustin Maravall	Luxembourg	26-feb-01	2-mar-01
PDS-001/2001	Basic Principles of Publication and Dissemination of Statistical Products	5	Ed Swires Hennessy	London	5-mar-01	9-mar-01
OSG-001/2001	The European Statistical System	3	Eurostat Experts	Luxembourg	19-mar-01	21-mar-01
SIS-203/2001	Geographical Information Systems	4	Marco Painho	Lisbon	26-mar-01	29-mar-01
MSI-101/2001	Adding Value through Strategic Management	5	John Scrivener	London	2-abr-01	6-abr-01
ECO-150/2001	Business Cycle Statistics	n 3	Klaus Reeh	Luxembourg	9-abr-01	11-abr-01
ECO-101/2001	Nomenclatures, Classifications and their Harmonisation	4	Niels Langkjaer	Vienne	23-abr-01	26-abr-01
SIS-201/2001	Statistical Disclosure Control	4	Peter Paul de Wolf	Voorburg	23-abr-01	26-abr-01
ECO-204/2001	The European System of Accounts (ESA95) - Goods and Services	3	Paul Konijn	Luxembourg	2-may-01	4-may-01
ECO-151/2001	Measuring the Economic Activity on the Internet	n 3	Pierre Dybman	Luxembourg	7-may-01	9-may-01
PDS-101/2001	Towards User-friendly Statistical Reporting	n 3	John Wright	London	9-may-01	11-may-01
ECO-203/2001	The European System of Accounts (ESA95) - Financial Accounts	3	Christine Coin	Luxembourg	21-may-01	23-may-01
PDS-105/2001	Marketing and Sales of Statistical Products and Services	3	Klaus Haagensen	Copenhagen	28-may-01	30-may-01
DAT-002/2001	Sampling Techniques and Practice	10	Prof. T.M.F. Smith	Southampton	18-jun-01	29-jun-01
ECO-001/2001	National Account Statistics in Practice (French)	10	François Lequiller	Paris	18-jun-01	29-jun-01
MSI-150/2001	Quality Management in Statistics	n 3	Lilli Japec Werner Grünwald	To be announced	25-jun-01	27-jun-01
DAT-105/2001	The Use of Administrative Sources for Statistical Purposes	n 3	Steve Vale	Helsinki	3-sep-01	5-sep-01
SOC-102/2001	Labour Cost and Labour Price Statistics	3	Steve Clarke	Luxembourg	12-sep-01	14-sep-01
DAT-207/2001	Advanced Sampling Techniques	n 3	Yves Tillé	Neuchatel	24-sep-01	27-sep-01
ECO-105/2001	Theory and Application of Enterprise Panel Surveys	5	Pierre Lavallée	Madrid	1-oct-01	5-oct-01
ECO-201/2001	Environmental Expenditure Statistics and Accounts (SERIEE)	n 4	Anton Steurer	Luxembourg	1-oct-01	4-oct-01
ASA-101/2001	Introduction to Applied Time Series Analysis	5	Agustin Maravall	Brussels	8-oct-01	12-oct-01
OSG-003/2001	Confidentiality and Protection of Privacy	3	Photis Nanopoulos	Luxembourg	15-oct-01	17-oct-01
ECO-103/2001	Enterprise Statistics	5	Johan Lock	Voorburg	5-nov-01	9-nov-01
SOC-001/2001	Systems of Social Statistics	5	Pieter Everaers	Heerlen	12-nov-01	16-nov-01

Xarxa temàtica  
**Enquestes i qualitat  
de la informació estadística**



**WORKSHOP  
DATA MINING**

Jornades internacionals sobre  
**Mineria de Dades**

Barcelona  
**14 i 15 de desembre de 2000**

**Col·laboren:**

- Institut d'Estadística de Catalunya
- Service pour la Science et la Technologie de l'Ambassade de France
- Universitat de Barcelona
- Universitat Politècnica de Catalunya

## **Jornades internacionals sobre Data Mining**

La preocupació per l'enorme volum de dades generades en certs contextos de recollida d'informació estadística, provoca la cerca d'instruments que facilitin la producció de resultats.

El data mining (o mineria de dades) estudia els procediments per extreure informació en aquests casos, on la quantitat de dades és molt gran. Permet tenir a l'abast els instruments de cerca, classificació, organització i categorització de la informació. També possibilita descobrir relacions, patrons o tendències difícils d'esbrinar amb els procediments convencionals.

Es pretén donar al Workshop un enfocament pràctic i afavorir la sinèrgia entre investigadors i professionals del sector públic i de l'empresa.

### **A qui s'adreça el Workshop?**

Aquestes jornades s'adrecen a estadístics i usuaris de l'estadística, investigadors o professionals del món institucional i empresarial preocupats per la gestió i l'anàlisi de la informació estadística.

### **Comitè de programa:**

Manuela Alcañiz (UB, Barcelona)  
Tomàs Aluja (UPC, Barcelona)  
Joan M. Batista (URL, Barcelona)  
Mònica Bécue (UPC, Barcelona)  
Germà Coenders (UG, Girona)  
Michael Greenacre (UPF, Barcelona)  
Montserrat Guillén (UB, Barcelona)  
Michel Meste (U. Perpignan, Perpignan)  
Pedro Puig (UAB, Bellaterra)  
Enric Ripoll (Idescat, Barcelona)  
Frederic Utzet (UAB, Bellaterra)  
Sylvie Viguier-Pla (U. Perpignan, Perpignan)

### **Programa provisional:**

Es realitzaran quatre sessions monogràfiques amb l'exposició de ponències en cadascuna d'elles. Per a una descripció detallada del programa i ponents consultar l'adreça: [www.eco.ub.es/~xteqie](http://www.eco.ub.es/~xteqie)

**Comunicacions lliures:**

Els participants que desitgin presentar una comunicació hauran de posar-se en contacte amb l'organització del Workshop abans del 10 de novembre de 2000.

**Preus d'inscripció:**

	<b>Abans 20.11.00</b>	<b>Després 20.11.00</b>
Participants	35.000 Pta	50.000 Pta
Universitaris i Adm. Pública	25.000 Pta	35.000 Pta
Membres de la Xarxa Temàtica	10.000 Pta	15.000 Pta

Els participants trametran un taló nominatiu amb la denominació de **Universitat Politècnica de Catalunya** adreçat a la instància indicada a l'apartat «Informació i inscripcions» i prèviament hauran d'omplir la butlleta d'inscripció electrònica que trobaran a l'adreça: [www.eco.upc.edu/~xteqie](http://www.eco.upc.edu/~xteqie)

**Informació i inscripcions:****XTEQIE-DM**

Mónica Bécue

Dept. Estadística i Investigació Operativa

Facultat de Matemàtiques i Estadística (UPC)

Pau Gargallo, 5

08028 Barcelona

Fax 34 93 401 58 55

E-mail: monica@eio.upc.es

**Lloc:**

Universitat de Barcelona

Facultat de Ciències Econòmiques i Empresarials

Aula Magna

Avgda. Diagonal, 690, Barcelona

08034 Barcelona

# **Informació per als autors i lectors**



## NORMES PER A LA PRESENTACIÓ D'ARTICLES A QÜESTIIÓ

La revista accepta, per a la seva publicació, articles originals no sotmesos a consideració en cap altra revista dins els àmbits de l'Estadística, la Investigació Operativa, l'Estadística Oficial i la Biometria. Els articles poden ser teòrics o aplicats, incloent aspectes computacionals i/o de caire docent, i poden presentar-se en anglès, francès, català o qualsevol altra llengua oficial a l'Estat espanyol.

Tots els originals destinats a les esmentades seccions temàtiques de *Qüestiió*, incloent-hi els articles per a la «Secció docent i problemes», seran sotmesos sistemàticament a unavaluació prèvia a càrrec d'especialistes independents i/o membres del Consell Editorial, llevat dels articles convidats per la revista i les reimpressions d'articles. El resultat de l'avaluació serà comunicat a l'autor principal als efectes d'eventuals correccions formals o dels seus continguts.

Per a totes les trameses d'originals, la revista emetrà un acusament de recepció la data del qual figurarà com a «data de rebuda» en la publicació de l'article. Per la seva banda, la «data d'acceptació» de l'article serà la data de recepció de la versió definitiva.

Per a la presentació d'articles, l'autor trametrà a la Secretaria de *Qüestiió* (Institut d'Estadística de Catalunya) dues còpies del treball mecanografiat en DIN A4, a una sola cara, a doble espai i amb marges amplis. Cada article ha d'incloure el títol, el nom de l'autor o autors, la seva afiliació i l'adreça completa, així com un resum de 75-100 paraules al principi de l'article, seguit de les principals paraules clau (en l'idioma original) i la seva adscripció a la classificació MSC2000 de l'American Mathematical Society. Abans de sotmetre els articles a la revista, s'aconsella els autors que revisin la correcció lingüística de textos d'acord amb l'idioma original i les eventuals traduccions a l'anglès.

Les referències bibliogràfiques es faran indicant el cognom de l'autor seguit de l'any de la publicació entre parèntesi [i.e.: Mahalanobis (1936), Rao (1982b)] i seran llistades alfabèticament al final de l'article; les referències múltiples d'un mateix autor s'ordenaran cronològicament. Les notes explicatives es numeraran correlativament i han d'aparèixer al peu de la pàgina corresponent. Les taules i figures també es numeraran correlativament en el text i seran reproduïdes directament dels originals tramesos en cas que no sigui possible la seva autoedició.

Una vegada evaluat satisfactoriament l'article cal que, a més de la versió impresa, l'autor el trameji en disquet de 3.5 polsades i en format MS-DOS, on han de constar de forma clara els noms dels autors i el títol de l'article. Aquesta versió final s'ha de tramestre preferiblement en el processador de textos *LATEX2<sub>e</sub>* [subsidiàriament, es poden tramestre els textos i les taules en Word Perfect —versió 6.0A o anterior— o ASCII]; en el cas de figures, diagrames o gràfics es recomanen els formats adients per als programes-editors PS, EPS o PCX. Els autors han de garantir la correspondència exacta entre la versió impresa i la còpia electrònica. D'altra banda, si l'article no està escrit en llengua anglesa s'haurà d'adjuntar la traducció del títol original, de l'abstract i de les paraules clau, així com un ampli resum en anglès (amb una extensió d'entre 2 i 5 pàgines i amb la mateixa estructura de l'article original).

La Secretaria de *Qüestiió* posa a disposició dels autors que ho sol·licitin plantilles en format *LATEX2<sub>e</sub>* per a la seva edició i les referències adients de la classificació de l'AMS.

**QÜESTIIÓ**  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona  
Tel: +34-93 412 15 36  
Fax: +34-93 412 31 45  
E-mail: questio@idescat.es

## **GUIDELINES FOR THE SUBMISSION OF ARTICLES FOR QÜESTIÓ**

The journal well comes submission of articles and contributions that are not being considered for publication in any other journal in the fields of Statistics, Operational Research, Official Statistics or Biometrics. Articles may be theoretical or applied, including teaching aspects and applications, and will be accepted in English, French, Catalan or any of the other official languages in Spain.

All originals assigned to the thematic sections of Qüestió, including articles for the «Teaching section and problems» will be systematically reviewed by independent referees and/or members of the Editorial Board, who will send a report to the main author of the article in order to correct, if necessary, any formal or content aspects. The articles invited by the journal and articles reprinted will be excluded from this evaluation process.

For all submissions, the journal will issue a receipt corresponding to the submission date, which will appear as «date received» in the final publication of the article. The «acceptance date» of the article, which will appear in its final publication, will be the date of sending the final version to the journal.

For the presentation of original articles, the author should send, to the Secretary of Qüestió (Institut d'Estadística de Catalunya), two copies of the paper typed on A4 sheets, one side of the paper only, double spaced and with wide margins. Each article should include the title, the name of the author or authors, their affiliation, full address and also an abstract of the paper (75-100 words) at the beginning of the article, followed by the main keywords (in the original language) and its assignation in the MSC2000 classification of American Mathematical Society. Before submitting their papers, authors are advised to seek assistance in the writing of their articles for the correct use of English and/or of original language.

Bibliographical references should state the author's name followed by the year of publication in brackets [e.g.: Mahalanobis (1936), Rao (1982b)] and they should be listed at the end of the article in alphabetical order; multiple references to the same author should be given in chronological order. Footnotes should be numbered in the article and appear at the foot of the corresponding page. Figures and tables are to be numbered in consecutive order in the text using Arabic numerals and will be directly reproduced from the originals submitted if it is not impossible to print them electronically.

Once the evaluation has been passed, the author is required to provide the article on a diskette (a 3.5-inch disk in MS-DOS format) together with its paper copy; it must be a new diskette and must bear very clearly the names of the authors and the title of the article. This final version should be processed by L<sup>A</sup>T<sub>E</sub>X 2<sub>E</sub>, preferably, or, failing that, by Word Perfect (6.0A or earlier) or ASCII for text and tables; for figures, diagrams or graphs, the appropriate formats of PS, EPS or PCX software tools are strongly recommended. Authors must ensure that the version of the electronic copy is exactly the same as the paper copy which accompanies it. Furthermore, if the article is not written in English, the translation of its original title, short abstract and keywords should be enclosed, as well as a full summary of the article in English (that is, 2-5 pages with the same structure as the original).

The Secretary of Qüestió can send, by request of the authors, the L<sup>A</sup>T<sub>E</sub>X 2<sub>E</sub> style of Qüestió for manuscript preparation and the appropriate classification of AMS references.

**QÜESTIÓ**  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona  
Tel: +34-93 412 15 36  
Fax: +34-93 412 31 45  
E-mail: questio@idescat.es

## **NORMES PER A LA PUBLICACIÓ D'ANUNCIS INSTITUCIONALS A QÜESTIIÓ**

Qüestiió convida les entitats patrocinadores, les institucions col·laboradores, els organismes públics i privats, i tota la comunitat científica vinculada a l'estadística o la investigació operativa, a la publicació d'anuncis institucionals sobre cursos, seminaris, congressos i activitats similars que, preferentment, tinguin lloc en el nostre país. Els textos poden presentar-se en anglès, francès, català o en qualsevol altra llengua oficial a l'Estat espanyol. Les iniciatives per a una possible publicació sempre són a instància de les entitats interessades, de manera que Qüestiió no fa una cerca sistemàtica d'esdeveniments d'aquesta naturalesa, ni té cap ànim d'exhaustivitat en les ressenyes d'activitats finalment publicades.

Una vegada aprovada la inclusió dels anuncis sol·licitats es procedirà a la seva publicació, i es reproduirà directament dels originals tramesos amb les mides adequades i la màxima qualitat tipogràfica possible; en aquest cas, Qüestiió no procedeix a cap mena de procés d'autoedició de la versió impresa que l'anunciant hagi trames. Si els originals es trameten en els mateixos termes electrònics exigits per als articles (vegeu «Normes per a la presentació d'articles a Qüestiió»), la revista procedirà a la seva autoedició. Si es desitja una qualitat superior a la reproducció simple o l'autoedició, o bé la seva publicació en color, els sol·licitants hauran de posar-se en contacte amb la Secretaria de Qüestiió per tal de trametre els fotolits dels textos originals corresponents.

La disposició dels textos i les figures adjuntes dels anuncis han de procurar la màxima intel·ligibilitat i claredat expositiva, sense atapeir la informació ni utilitzar formats o fonts de lletres excessivament petites. D'altra banda, la publicitat ha de ser fidedigna, exempta d'enganyos i respectuosa amb les persones i institucions. En qualsevol cas, la direcció de Qüestiió es reserva la decisió final pel que fa a la seva publicació.

L'anunciant es compromet a lluirar els textos/materials amb l'antelació que se li indiqui per a la inserció en els números/volums de Qüestiió que prèviament s'hagi establert. La revista no es fa responsable dels retards, per part de l'anunciant, que impedeixin la publicació de l'anunci en els termes previstos.

Mònica M. Jaime  
Secretaria de Qüestiió  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona  
Tel: +34-93 412 15 36  
Fax: +34-93 412 31 45  
E-mail: [questio@idescat.es](mailto:questio@idescat.es)

## **GUIDELINES FOR INSTITUTIONAL ADVERTISEMENTS IN QÜESTIÓ**

Qüestiió invites all sponsor entities, collaborating institutions, other public and private bodies and the entire scientific community related to Statistics or Operations Research to submit institutional advertisements on courses, seminars, congress and similar activities that will be held, preferably in our country. These will be accepted in English, French, Catalan or any of the other official languages in Spain. The initiative should always come from the entities interested in advertising them so that Qüestiió's aim is not to do a systematic search of these events and therefore does not publish a comprehensive list of such activities.

Once their insertion is approved the advertisements will be reproduced from the most accurate photocopy of the originals sent by the advertiser to Qüestiió in paper copy, with the appropriate size and at the best possible typographic quality. Therefore, in this case the journal does not elaborate any further editing process to the printed version that the advertiser has sent. If the original advertisements are sent in the same electronic format requested by the articles (please see «Guidelines for the submission of articles for Qüestiió») the journal will print it directly from the file. If a better quality than the simple reproduction or automatic printing or a colour version of the adverts is desired, the authors should contact the Secretary of Qüestiió in order to negotiate this.

The typesetting of texts and figures in the advertisement should have maximum intelligibility and clearness, neither compressing the information too much nor using formats or letter fonts that are too small. Furthermore, the information has to be reliable, without errors and respectful of the people and institutions. The management of Qüestiió has the right to a final decision concerning the insertion of the advertisement.

Advertisers commit themselves to give the text/materials on request in order to insert them in the issues of Qüestiió that have been previously agreed. The journal is not responsible for any delay from the announcer that could prevent the advertisement from being published on the agreed terms.

Mònica M. Jaime  
Secretaria de Qüestiió  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona  
Tel: +34-93 412 15 36  
Fax: +34-93 412 31 45  
E-mail: [questiio@idescat.es](mailto:questiio@idescat.es)

## NORMES PER A LA PUBLICACIÓ D'ANUNCIS PRIVATS O AMB FINALITAT COMERCIAL A QÜESTIÓ

Qüestió accepta la publicació d'anuncis privats o amb finalitat comercial sobre productes, serveis o altres eines promocionals a l'entorn de l'estadística o la investigació operativa. Els textos poden presentar-se en anglès, francès, català o en qualsevol altra llengua oficial a l'Estat espanyol. Les iniciatives per a una possible publicació sempre són a instància de les organitzacions que hi estiguin interessades, de manera que Qüestió no fa una cerca sistemàtica de novetats o productes d'aquesta naturalesa ni té cap ànim d'exhaustivitat en els anuncis finalment publicats.

Els anuncis en **blanc i negre** s'elaboren a partir de la fotòcòpia més acurada possible dels originals que tramen l'anunciant en versió impresa, amb les mides adequades i la màxima qualitat tipogràfica. Per tant, en aquest cas la revista no efectua cap procés d'edició ulterior respecte de la versió impresa que l'anunciant hagi trames. Alternativament, si els anuncis originals es tramenen en els mateixos termes formals exigits per als articles (vegeu «Normes per a la presentació d'articles a Qüestió»), la revista procedirà a la seva autoedició. Igualment, si es desitja una qualitat superior a la reproducció simple, els sol·licitants hauran de tramentar els fotolits dels originals corresponents o encarregar-los a Qüestió, que els facturarà separadament.

Els anuncis en **color** requereixen els fotolits dels textos originals, que poden ser subministrats directament per l'anunciant o bé encarregats per la revista a compte de l'anunciant; en el segon cas, l'anunciant ha de tramentar a la revista els originals impresos en color amb la màxima qualitat, per tal de filmar-los amb les millors garanties i condicions. El cost dels fotolits realitzats per Qüestió serà sempre a càrrec de l'anunciant, a qui se li repercutirà l'import i l'IVA d'aquests, juntament amb les tarifes que corresponen a la modalitat d'anunci per la qual hagi optat.

La disposició dels textos i figures adjuntes dels anuncis ha de procurar la màxima intel·ligibilitat i claredat expositiva, sense atapeir la informació ni utilitzar formats o fonts de lletres excessivament petites. D'altra banda, la publicitat ha de ser fidedigna, exempta d'engany i respectuosa amb les persones i institucions. En qualsevol cas, la direcció de Qüestió es reserva la decisió final de la seva inclusió.

L'anunciant es compromet a lluir els textos/materials amb l'antelació que se li indiqui per a la seva inserció en el(s) número(s)/volum(s) de Qüestió que prèviament s'hagi establert. La revista no es fa responsable dels retards per part de l'anunciant que impedeixin la publicació de l'anunci en els termes previstos.

### Imports:

1 pàgina en color (un número aïllat):	125.000 PTA + IVA
1 pàgina en color (tres números consecutius):	200.000 PTA + IVA
1 pàgina en blanc i negre (un número aïllat):	30.000 PTA + IVA
1 pàgina en blanc i negre (tres números consecutius):	50.000 PTA + IVA
1/2 pàgina en blanc i negre (un número aïllat):	20.000 PTA + IVA
1/2 pàgina en blanc i negre (tres números consecutius):	35.000 PTA + IVA

### Mides opcionals dels anuncis:

1 pàgina sencera (espai intern):	19.0 cm. x 12.3 cm.
1 pàgina sencera (espai extern):	23.8 cm. x 17.0 cm.
1/2 pàgina (espai intern):	9.5 cm. x 12.3 cm.
1/2 pàgina (espai extern):	11.9 cm. x 17.0 cm.

### Forma de Pagament:

- Transferència bancària al compte: 2013-0100-53-0200698577
- Xec bancari nominatiu a l'Institut d'Estadística de Catalunya
- Pagament amb targeta de crèdit

El pagament serà per l'import total de la factura corresponent, on hi figurarà el cost dels fotolits en el cas que l'edició de l'anunci hagi estat a càrrec de l'Institut. En el cas que l'anunciant necessiti una factura proforma, només cal que ho faci saber amb l'antelació suficient.

### Correspondència:

Mònica M. Jaime  
Secretaria de Qüestió  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona  
Tel: +34-93 412 15 36  
Fax: +34-93 412 31 45  
E-mail: questio@idescat.es

## **GUIDELINES FOR THE PRIVATE OR COMMERCIAL ADVERTISEMENTS IN QÜESTIÓ**

Qüestió accepts for their publication both private and commercial advertisements on products, services or other promotional tools related to statistics or operational research and will be accepted in English, French, Catalan or any of the official languages in Spain. The initiatives should always come from entities interested in advertising them so that Qüestió's aim is not to do a systematic search of news and therefore does not publish a comprehensive list of such private or profit activities.

The black and white advertisements are made out from the most accurate photocopy of the originals sent by the advertiser to Qüestió in paper copy with the appropriate size and at the best possible typographic quality. Therefore, in this case the journal does not elaborate any further editorial process to the printed version that the advertiser has sent. Alternatively, if the original advertisements are sent in the same formal terms required by the articles (please see «Guidelines for the submission of articles for Qüestió»), the journal will proceed to its autoedition. In the same way, if a better quality than the simple reproduction is wanted, the authors should send the photolits of the corresponding original texts or, on the other hand, order to Qüestió their fulfilment, which will be invoiced separately from the rates charged as advertisements.

The advertisements in colour need the photolits of the original texts, which can be provided directly by the advertiser or requested by Qüestió to the advertiser charge; in the second case, the advertiser must send to the journal the originals printed in colour with the best possible quality, so that they can be filmed at the best conditions and guarantees. The cost of the photolits made by Qüestió will always be charged to the advertiser together with the VAT derived from it, plus the prices corresponding to the type of the advertisement that has been chosen.

The set up of texts and figures of the advertisement should provide the maximum intelligibility and clearness, neither squeezing together the information nor using set ups or letter types that are too small. On the other hand the publicity has to be reliable, without fraud and respectful to the persons and institutions. The direction of Qüestió has the right of the last decision concerning the insertion of the advertisement.

The advertiser commits himself to give the texts/materials on request, in order to insert them in the issue(s) of Qüestió that had been previously agreed. The journal is not responsible for any delay from the announcer that could prevent the advertisement from being published in the agreed terms.

### **Rates:**

1 colour page (only one issue):	125.000 PTA + VAT
1 colour page (three consecutive issues):	200.000 PTA + VAT
1 black and white page (only one issue):	30.000 PTA + VAT
1 black and white page (three consecutive issues):	50.000 PTA + VAT
1/2 black and white page (only one issue):	20.000 PTA + VAT
1/2 black and white page (three consecutive issues):	35.000 PTA + VAT

### **Advertisement sizes (optional):**

1 full page (internal space):	19.0 cm. x 12.3 cm.
1 full page (external space):	23.8 cm. x 17.0 cm.
1/2 page (internal space):	9.5 cm. x 12.3 cm.
1/2 page (external space):	11.9 cm. x 17.0 cm.

### **Payment:**

- A bank transfer to account number: 2013-0100-53-0200698577
- A bank cheque to Institut d'Estadística de Catalunya
- Charge on a credit card

The payment should be for the amount shown at the invoice, where it will be shown the total cost of the photolits, in case that Qüestió would be in charge of the filming of the advertisement. If advertiser need a pro-forma invoice, he should let us know some time in advance so that Qüestió could send it to the proper address.

### **Mail address:**

Mònica M. Jaime  
Secretaria de Qüestió  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona  
Tel: +34-93 412 15 36  
Fax: +34-93 412 31 45  
E-mail: questio@idescat.es

## **Butlleta de subscripció a la revista *Qüestió***

Nom i cognoms \_\_\_\_\_

Empresa/Institució \_\_\_\_\_

Adreça \_\_\_\_\_

Codi postal \_\_\_\_\_ Ciutat \_\_\_\_\_

Tel. \_\_\_\_\_ Fax \_\_\_\_\_ NIF \_\_\_\_\_

Data \_\_\_\_\_

Signatura

Desitjo subscriure'm a **Quèstió** per a l'any 2001

Preu de subscripció vigent:

- Estat espanyol: 3.600 Pta/any (21,64€) (IVA inclòs)
  - Estranger: 4.000 Pta/any (24,04€) (IVA inclòs)

## Forma de pagament

- Transferència al compte 2013-0100-53-0200698577
  - Domiciliació bancària al compte número  

--	--	--	--	--	--	--	--	--	--	--	--
  - Xec nominatiu a l'Institut d'Estadística de Catalunya
  - Gir postal
  - En efectiu

Retorneu aquesta butlleta (o una fotocòpia) a:

**Qüestió**  
Institut d'Estadística de Catalunya  
Via Laietana, 58  
08003 Barcelona

Preu de números solts (actuals i endarrerits):

- Estat espanyol: 1.500 Pta/exemplar (9,02 €) (IVA inclòs)
  - Estranger: 1.700 Pta/exemplar (10,22 €) (IVA inclòs)



**Exemplar per a l'entitat bancària**

Autorització de domiciliació bancària per al pagamento de les subscripcions anuals de la revista ***Qüestió***

El sotassinat _____
autoritza el Banc/Caixa _____
Adreça _____
Codi postal _____ Ciutat _____
a abonar les subscripcions a la revista <b><i>Qüestió</i></b> amb càrrec al seu compte número <input type="text"/>
Data _____
Signatura

***Qüestió***  
**Institut d'Estadística de Catalunya**  
Via Laietana, 58  
08003 Barcelona

614

# Novetats editorials en matèria estadística de la Generalitat de Catalunya gener-desembre 2000

- **Anuari Estadístic de Catalunya, 1992-1999 CD-ROM**  
1999, 3.000 PTA (18,03 €),  
ISBN: 84-393-4914-9
- **Xifres de Catalunya 1999**  
versions en català, castellà, francès, anglès i alemany. Gratuit
- **Revista Qüestió Añ 2000 Vol. 24 núm. 1, 2, i 3**  
3.600 PTA (21,63 €),  
ISSN 0210-8054
- **Estadística, producció i comptes de la indústria 1998**  
gener 2000, 1.300 PTA (7,81 €), 371 pp.,  
ISBN: 84-393-5038-4
- **Projeccions de població de Catalunya 2010. Comarques i àmbits del Pla territorial**  
gener 2000, 1.600 PTA (9,62 €), 395 pp.,  
ISBN: 84-393-5009-0
- **Comptes de les administracions públiques de Catalunya 1996**  
maig 2000, 1.200 PTA (7,21 €), 138 pp.,  
ISBN: 84-393-5075-9
- **Estadística de població 1996 Vol. 8 Relació de la població amb l'activitat econòmica. Dades comarcals i municipals**  
maig 2000, 1.250 PTA (7,51 €), 192 pp.,  
ISBN: 84-393-5049-X
- **Mercat de treball 1999**  
març 2000, 1.500 PTA (9,01 €), 328 pp.,  
ISSN: 1137-0122, ISBN: 84-393-5092-9
- **Estadística comarcal i municipal 1999**  
abril 2000, 3.250 PTA (19,53 €), 712 pp.,  
ISSN: 1135-9331, ISBN: 84-393-5121-6
- **Anuari estadístic de Catalunya 2000**  
juliol 2000, 3.000 PTA (18,03 €), 832 pp.,  
ISSN: 1130-166-X, ISBN: 84-393-5184-4
- **Moviment natural de la població 1998**  
**Dades comarcals i municipals**  
juny 2000, 1.400 PTA (8,41 €), 106 pp.,  
ISBN: 84-393-5205-0
- **Comerç amb l'estrange 1997 Dades comarcals i municipals**  
setembre 2000, 1.400 PTA (8,41 €), 194 pp.,  
ISBN: 84-393-5202-6
- **Moviments migratoris 1997 Dades comarcals i municipals**  
setembre 2000, 1.600 PTA (9,62 €), 173 pp.,  
ISBN: 84-393-5204-2
- **Comerç amb l'estrange 1998 Dades comarcals i municipals**  
setembre 2000, 1.400 PTA (8,41 €), 194 pp.,  
ISBN: 84-393-5203-4
- **4. Indústria bàsica i productes metal·lífics**  
octubre 2000, 900 PTA (5,41 €), 79 pp.,  
ISBN: 84-393-5239-5
- **5. Maquinària i equips mecànics**  
octubre 2000, 900 PTA (5,41 €), 71 pp.,  
ISBN: 84-393-5240-9
- **Estadística de població 1996 Vol. 9 Situació professional de la població. Dades comarcals i municipals**  
maig 2000, 1.250 PTA (7,51 €), 118 pp.,  
ISBN: 84-393-5213-3
- **Estadística de població 1996 Vol. 7 Migracions: Any d'arribada a Catalunya. Dades comarcals i municipals**  
octubre 2000, 1.000 PTA (6,01 €), 108 pp.,  
ISBN: 84-393-5265-4
- **Estadística de població 1996 Vol. 10 Sectors d'activitat de la població. Dades comarcals i municipals**  
octubre 2000, 1.000 PTA (6,01 €), 110 pp.,  
ISBN: 84-393-5266-2
- **Els comptes de la Protecció social a Catalunya**  
setembre 2000, 1.400 PTA (8,41 €), 125 pp.,  
ISBN: 84-393-5206-9
- **Llars i famílies a Catalunya 1996**  
setembre 2000, 1.400 PTA (8,41 €), 207 pp.,  
ISBN: 84-393-5207-7
- **Estadístiques culturals a Catalunya**  
2.000 PTA (12,02 €), 190 pp.

## **LLIBRERIES DE LA GENERALITAT**

## Barcelona

Barcelona  
Rambla dels Estudis, 118 (tel. 93 302 64 62)  
llibrbcn@correu.cattel.com

Girona

Gran Via de Jaume I, 38 (tel. 972 22 72 67)  
llibrgi@ibernet.com

Lleida

**E-mail**  
Rambla d'Aragó, 43 (tel. 973 28 19 30)  
[llibrle@ibernet.com](mailto:llibrle@ibernet.com)

## Madrid

**blanquerna**  
Blanquerna. Llibreria catalana.  
Serrano, 1 (tel. 91 431 00 22)  
[blanquerna@nauta.es](mailto:blanquerna@nauta.es)

PUNT DE VENDA

Puigcerdà

Plaça del Rec, 5 (tel. 972 88 05 14)

## **VENDA PER CORREU**

Apartat 2800, 08080 Barcelona  
eadop@correu.gencat.es  
Plaça del Rec, 5 (tel. 972 88 05 1)