



Boletín Informativo Sociedad de Estadística e Investigación Operativa

**350 ANIVERSARIO DE LA MUERTE DE
JACOB BERNOULLI**

**20 AÑOS DEL PREMIO RAMIRO
MELENDRERAS**

**CREACIÓN DE UN CENTRO NACIONAL
DE MATEMÁTICAS**

ESTADÍSTICA Y SILLAS RESISTENTES

1 2 3 4 5 6 7



**Sociedad de Estadística e
Investigación Operativa**

BOLETÍN INFORMATIVO

Volumen 21, número 1

MARZO 2005

REDACCION

Editor: Jesús López Fidalgo
fidalgo@usal.es
Universidad de Salamanca

Editores Asociados:

Estadística:

Miguel Angel Gómez Villegas
ma_gv@mat.ucm.es
Universidad Complutense de Madrid

Investigación Operativa:

Justo Puerto Albandoz
puerto@us.es
Universidad de Sevilla

Aplicaciones:

Molina Fernández, Manuel
mmolina@unex.es
Universidad de Extremadura

Estadística pública:

Montserrat Herrador
herrador@ine.es
Instituto Nacional de Estadística

Edición técnica:

Facultad de Ciencias Matemáticas.-U.C.M. Despacho
502 Plaza de Ciencias, 3
28040 Madrid (Ciudad Universitaria)
Tel: 91 544 91 02
e-mail: oficina@seio.es - <http://www.seio.es>

Imprime SEROTEL - Pº de la Castellana, 87.
Dep. Legal: M-13647-1995
ISSN: 1699-4841
Copyright © 2005 SEIO
Para anuncios contacte: oficina@seio.es

Normas para los envíos de colaboraciones:

Los artículos se enviarán por correo electrónico al editor asociado correspondiente o al editor del Boletín. No deberán tener una longitud superior a 5 páginas.

Las cartas al editor se dirigirán por correo electrónico al editor. La limitación será de 200 palabras.

El resto de colaboraciones y noticias se dirigirán al corresponsal más cercano. Las referencias bibliográficas y de software se acompañarán de los datos necesarios para su localización y una reseña no superior a 120 palabras. Los resúmenes de tesis se limitarán a 200 palabras y contendrán: título, autor, directores, departamento, universidad y la fecha de lectura. Con relación a congresos y cursos bastará una breve reseña semejante a las publicadas en el Boletín.

El formato preferible para todas las colaboraciones es MS-Word.

INDICE

Editorial	3
El rincón del Presidente	4

1. Información académica y laboral

Noticias	6
Tesis	8
Oportunidades de trabajo	16
Publicaciones y software	17
Cursos y Seminarios	20
Congresos	22
Altas y bajas de socios	26

2. Estudios monográficos y opiniones sobre la profesión

*Historia del Premio Ramiro Melendreras, Domingo Morales González	26
* Informe de la acción especial BFM2002-12271-E, Manuel de León y Enrique Zuazua	30

3. Artículos de Estadística

* Jacob Bernoulli (1654-1705) a los 350 años de su muerte, Miguel A. Gómez Villegas	35
---	----

4. Artículos de Investigación Operativa

* El lema de Farkas: una herramienta para resolver nuevas aplicaciones, Juan José Salazar González ...	36
--	----

5. Artículos de Aplicación

* Análisis estadístico de la estabilidad en asientos de uso doméstico y público, Mónica Alacreul y Carmen Armero	40
--	----

6. Estadística Oficial

*La estimación en áreas pequeñas para la estadística oficial, Montserrat Herrador y Jorge Saralegui	47
---	----

EDITORIAL

Jesús López Fidalgo

En este primer Boletín de la Sociedad de Estadística e Investigación Operativa como editor me gustaría agradecer al Consejo Ejecutivo de la SEIO la confianza que ha puesto en mí para llevar a cabo este proyecto. Quiero también agradecer la disponibilidad de los corresponsales y de los editores asociados para colaborar en la edición del Boletín. Continuará editándose en papel, además de la edición electrónica, pero existirá la posibilidad de que los socios que lo prefieran soliciten recibirlo solamente en formato electrónico. Quizá en un futuro no muy lejano el Boletín pase a ser totalmente electrónico.

Para llevar a cabo este proyecto hemos querido formar un equipo de casi 40 corresponsales representando el mapa estadístico nacional. Las secciones del Boletín contarán con editores asociados encargados de fomentar y atender las colaboraciones que pudieran incluirse en cada edición. De modo general el Boletín contará con las secciones siguientes: Información académica y laboral, Artículos de divulgación en el ámbito de la Estadística, Artículos de divulgación en el ámbito de la Investigación Operativa, Artículos de aplicación, Estudios monográficos y opiniones sobre la profesión y Estadística Oficial. Se añade además una sección de cartas al editor (fidalgo@usal.es) y una sección que llevará por título “El rincón del presidente” en la que se pretende dar a conocer proyectos, logros y la información que el Presidente de la sociedad considere de interés para los socios. Se incluirá, cuando sea pertinente, un obituario en el que se recuerde a los que nos precedieron y nos van dejando en esta senda de la Estadística y la Investigación Operativa. Queremos que el editorial refleje una gran variedad de opiniones y puntos de vista. Por este motivo procuraremos recurrir a socios y no socios para elaborar el editorial de cada edición.

Quiero solicitar a los socios su colaboración para sacar adelante este proyecto. Para ello podéis acudir a vuestros corresponsales más cercanos en lo relativo a informaciones variadas como pueden ser: cursos ofrecidos por las Universidades u otros organismos, cursos internacionales; oportunidades de trabajo, becas, prácticas en empresas y en la administración pública; promoción y cambios de

destino de profesores universitarios, de estadísticos de la Administración Pública, de especialistas en Estadística e Investigación Operativa en empresas, u otro tipo de nombramientos; información sobre tesis leídas en Estadística e Investigación Operativa; ayudas a proyectos de investigación y desarrollo, descripción de grupos de investigación en Estadística e Investigación Operativa; publicaciones y software, ... La periodicidad del Boletín no permite insertar convocatorias inminentes. En este caso es preferible que enviéis la noticia directamente a la secretaria de la sociedad cuando surja: María Jesús Ríos Insua (maria_rios@mat.ucm.es).

En lo referente a colaboraciones relativas a Artículos de divulgación en el ámbito de la Estadística, Artículos de divulgación en el ámbito de la Investigación Operativa, Artículos de Aplicación, Estudios monográficos y opiniones sobre la profesión y Estadística Oficial, podéis enviar los trabajos directamente a los editores asociados o al editor del Boletín.

Estamos realizando gestiones para que el Boletín aparezca en la base de datos MathScinet, como ya ocurre con otros boletines similares.

No cabe duda de que nos encontramos en un momento trascendental para la universidad, en el que la Estadística y la Investigación Operativa han de asentarse en una posición adecuada. Esto es de gran importancia para el futuro y todos podemos estar vigilantes y activos para poner nuestro granito de arena. En la configuración del grado y el postgrado en el espacio europeo tenemos un reto apasionante.

Me gustaría llamar la atención de nuevo sobre algo que en mi opinión debería estar de alguna manera presente en la SEIO. Con el desciframiento del genoma humano la genética, o si se quiere la bioquímica, ha lanzado un S.O.S a la informática y a la estadística. Para esta nueva disciplina se ha ideado un nombre, “Bioinformática”, que desafortunadamente no nos incluye. Por supuesto se ha de acabar acudiendo a la estadística para el complejo análisis de estos datos. Pienso que somos ya muchos los

estadísticos que de una manera u otra estamos involucrados en grupos de investigación de esta materia. Por primera vez en el JCR de 2003 aparecen revistas como “Bioinformatics” o “Journal of computational biology” en el área de “Statistics and Probability”, lo que me parece una batalla importante ganada en esta guerra de la estadística aplicada. Pienso que deberíamos asegurarnos una plaza en el tren, que tarde o temprano lanzará la Bioinformática como área de conocimiento en la universidad.

Ciertamente, junto con esto podríamos abrir un interesante debate acerca del asesoramiento estadístico en otras áreas o en la empresa privada. En primer lugar, a la luz del seguimiento que se ha hecho de los diplomados y licenciados en estadística cabría plantearse cuál es el hueco profesional que viene a ocupar un “Estadístico”, incluyendo en este término la Investigación Operativa. Pienso que todos tenemos experiencia de la demanda de asesoramiento estadístico, quizá de especial magnitud en medicina. Habitualmente acuden a un amigo, o a un amigo de un amigo, para lograr este asesoramiento. Desde un punto de vista, digamos “de currículum”, a un investigador en estadística no le interesa aplicar los procedimientos tradicionales a problemas concretos, es decir, hacer asesoramiento. Aunque por otro lado, los mejores resultados de innovación en la investigación en estadística e investigación operativa suelen surgir de problemas reales.

El hecho es que no existe una estructura adecuada para cubrir esta demanda. ¿No sería razonable que

las universidades y centros de investigación dispusieran de un centro de asesoramiento estadístico del mismo modo que disponen de un centro de proceso de datos con informáticos? Ciertamente están surgiendo asesorías estadísticas (habitualmente “informáticas y estadísticas”) que podrían cubrir este hueco, pero parece que todavía queda un largo camino por andar. Aquí se podría incluso apuntar la idea compleja y peligrosa del intrusismo. Ahora que el número de nuestros alumnos disminuye de manera preocupante convendría potenciar de alguna forma su futuro profesional. Quizá de este modo, al ver que esta carrera proporciona un futuro halagüeño, crezca el interés por esta titulación.

Me gustaría hacer una reflexión sobre la investigación en nuestra área en los últimos años. Si de algo no cabe duda es de la subida del índice de impacto de la sociedad española. Cada vez es más frecuente ver nombres de investigadores españoles en las publicaciones más prestigiosas del área. Si bien es verdad que esto se ha logrado de una manera hasta cierto punto artificial, cuando se solicitan publicaciones de impacto para todo. Pero también es cierto que mediante este procedimiento, más o menos criticable, se ha logrado poner la investigación en Estadística, Probabilidad e Investigación Operativa en España en un lugar muy destacable en el marco internacional.

De nuevo quiero agradecerte de antemano tu colaboración y ponerme a tu disposición.

EL RINCÓN DEL PRESIDENTE

Domingo Morales González

Transcurridos cinco meses desde nuestro último Congreso Nacional de Estadística e Investigación Operativa, aprovecho la oportunidad que me brinda el editor del Boletín para informar de las actividades desarrolladas y de los acuerdos tomados por el Consejo Ejecutivo y los Consejos Académicos de la SEIO. Las líneas preferentes de actuación para mi periodo presidencial se expusieron durante la Asamblea General de Cádiz y con los continuos esfuerzos de María Jesús Ríos Insua, nuestra Secretaria General, y de todos los compañeros de los Consejos, se están llevando adelante.

Quiero comenzar recordando que nuestras revistas

TEST y TOP están cada vez más asentadas en el panorama científico internacional. En estos momentos estamos desarrollando un proyecto de renovación de las páginas web y de digitalización completa de las mismas. En el caso de TEST es obligado agradecer a Enrique Castillo Ron y a Juan Antonio Cuesta Albertos la gran labor realizada al frente de la revista durante el periodo 2002-2004. El índice de impacto de TEST en el año 2003 ha sido 0.581 y ocupa la posición 41 de entre 75 revistas referenciadas. Los nuevos editores de TEST son Leandro Pardo Llorente y María Ángeles Gil Álvarez. Estoy seguro que ellos harán todo lo posible para mejorar la relevancia científica de la revista. Los editores de

TOP siguen siendo Marco A. López Cerdá e Ignacio García Jurado. Me consta que llevan tiempo trabajando para conseguir que TOP entre también en la relación de revistas referenciadas en el ISI Journal Citation Report. A todos ellos, y a los editores técnicos María Asunción Lubiano Gómez y Antonio Alonso Ayuso, quiero expresarles mi agradecimiento por el trabajo que realizan para mantener el alto nivel de calidad de nuestras revistas.

Desde septiembre de 1993 hasta diciembre de 2004, Francisco Javier Quintana Martín ha estado al frente del boletín de la SEIO. Finalmente ha dejado la dirección con uno de sus personales y brillantes editoriales que serán difíciles de igualar. Muchas gracias, Paco, por tu valiosa colaboración durante estos años. Para esta nueva etapa, Jesús López Fidalgo (Universidad de Salamanca), con el apoyo de Miguel Ángel Gómez Villegas (Universidad Complutense de Madrid), Justo Puerto Albandoz (Universidad de Sevilla), Manuel Molina Fernández (Universidad de Extremadura) y Montserrat Herrador Cansado (Instituto Nacional de Estadística), ha aceptado el reto de dirigir el boletín e introducir un nuevo formato y contenidos. Asimismo se cuenta con una amplia lista de colaboradores y corresponsales de universidades que, sin lugar a dudas, contribuirán a completar la información publicada en el boletín.

Los organizadores del próximo congreso de la SEIO, que se celebrará en Tenerife, ya llevan tiempo trabajando en su preparación y pronto tendremos noticias concretas suyas.

En el panorama de las Matemáticas, la SEIO colabora y coordina actividades con el resto de las sociedades matemáticas españolas; en especial, con la Real Sociedad Matemática Española (RSME), la Sociedad de Matemática Aplicada (SEMA) y la Societat Catalana de Matematiques (SCM). Resultados visibles de esta cooperación han sido la creación del Comité Español de Matemáticas (CEMAT), las actividades encaminadas a solicitar un Centro Nacional de Matemáticas (CNMat), la celebración en Valencia del primer congreso conjunto de las cuatro sociedades (MAT.ES 2005), la participación en el Comité Organizador del Congreso Internacional de Matemáticas ICM2006 y los acuerdos firmados para facilitar la existencia de socios comunes a varias sociedades.

La conexión con la Estadística Pública es un

objetivo primordial de nuestra sociedad, siendo tarea de todos el potenciar la relación entre la Estadística académica y la oficial. Ese objetivo ha sido reconocido por los legisladores en el Real Decreto 53/2005, de 21 de enero, que contempla la representación de la SEIO en el Consejo Nacional de Estadística. Por su condición de perteneciente al Cuerpo Superior de Estadística, se ha propuesto a Pilar Ibarrola Muñoz (Universidad Complutense de Madrid) para desempeñar el cargo de consejera. Por otra parte, en los congresos de la SEIO se han venido organizando sesiones monográficas de Estadística Pública en las que se va poniendo de manifiesto la creciente cooperación de las universidades con las oficinas centrales y autonómicas de estadística. A partir de 2005 se incorpora al Comité Editorial del boletín la Jefa del Área de Metodología del INE Montserrat Herrador. Su labor será la de incluir artículos y noticias relacionadas con la Estadística Pública.

En el marco de las actividades de colaboración con la Administración, se ha propuesto a Juan García Laguna (Universidad de Valladolid) para la participación como miembro de la Comisión Técnica de apoyo al programa de Evaluación de las Matemáticas que se va a llevar a cabo en los centros de Educación Infantil, Primaria y Secundaria de Castilla y León.

También la SEIO participa en distintas organizaciones internacionales, tanto en el ámbito de la Estadística como en el de la Investigación Operativa. Así, por ejemplo, conviene citar algunas iniciativas de EURO de gran interés como son la edición de publicaciones divulgativas sobre la Investigación Operativa o la posible doble afiliación SEIO-EURO de nuestros socios. O bien, la publicación de un artículo de 1000 palabras sobre la SEIO en el boletín de la American Statistical Association.

La SEIO es consciente de la necesidad de ayudar y apoyar, en la medida de sus posibilidades, a los principales eventos relacionados con la Estadística y la Investigación Operativa que se celebren en nuestro país. El Congreso “Operational Research Peripatetic Post-graduate Programme (ORP3 2005)”, a celebrar en Valencia (Spain) del 6 al 10 de Septiembre del 2005, es un nuevo instrumento diseñado por la Asociación Europea de Sociedades de Investigación Operativa (EURO) para jóvenes investigadores y profesionales de esta disciplina. Por ello, y teniendo en cuenta la importancia del Congreso ORP3 2005 en la

difusión de la Investigación Operativa en aquellas personas que se inician en la investigación, la SEIO ha convocado seis becas de 200 euros cada una para apoyar la asistencia al congreso.

Además, la SEIO ha acordado participar como entidad colaboradora en el International Seminar on Nonparametric Inference (ISNI 2005), que tendrá lugar en La Coruña del 13 al 15 de julio, organizado por el área de Estadística e Investigación Operativa del Departamento de Matemáticas de la Universidade da Coruña.

La potenciación de los grupos de trabajo de la

SEIO y el apoyo de nuevas iniciativas, como la creación de un grupo de trabajo en didáctica de la Estadística y de la Investigación Operativa, serán también líneas prioritarias de actuación para 2005.

Por último, unas palabras de aliento y ánimo para todos. La SEIO necesita de vuestra colaboración y, en algunos casos, de vuestra implicación personal. Resulta gratificante constatar como crece el número de socios y el número de actividades. Aun queda mucho camino por recorrer pero lo importante es seguir avanzando. Un cordial saludo a todos.

1. INFORMACIÓN ACADÉMICA Y LABORAL

NOTICIAS

NOMBRAMIENTO

Vera Pawlowsky Glahn, Catedrática de la Universidad de Girona, ha dejado con fecha 1 de febrero de 2005 el cargo de Directora del Departamento de Informática y Matemática Aplicada, para pasar a ocupar el cargo de Directora de la Escuela de Posgrado.

PREMIO IAMG

El pasado mes de febrero, Raimon Tolosana, del Departamento de Informática y Matemática Aplicada de la Universidad de Girona, fue premiado con una ayuda de la IAMG para llevar a cabo el proyecto Krigeado en coordenadas: estimación de funciones de probabilidad. El premio consiste en una ayuda monetaria y se trata de una convocatoria internacional a la que pueden concurrir estudiantes de doctorado de todo el mundo. El proyecto premiado forma parte de la tesis de Raimon, dedicada al krigeado en coordenadas, cuyo objetivo general es resolver ciertos problemas relativos a la no-optimalidad de las técnicas de interpolación espacial (kriging) cuando los datos no siguen una distribución normal por tener un soporte acotado (p.ej., datos positivos, composicionales, multinomiales). Sobre la base de que toda técnica estadística ideada para datos reales puede aplicarse a las coordenadas reales de los objetos de un espacio real euclidiano, Raimon estudia las propiedades generales de un

estimador de krigeado en un espacio euclidiano genérico, y de varios casos particulares. En concreto, el proyecto explora la aplicación de esta técnica a observaciones de una distribución binomial o multinomial, consideradas como vectores en el simplex, el conjunto de vectores de componentes positivos de suma constante igual a uno. El objetivo es desarrollar una alternativa al krigeado de funciones indicatrices, o krigeado indicador (IK), que pretende estimar la probabilidad de un suceso en un punto no muestreado. IK es una de las técnicas más usadas en Geoestadística, dada su simplicidad, y sin embargo es también una de las más discutidas, puesto que a menudo ofrece estimas imposibles como p.ej. probabilidades negativas. Puede entenderse que esto se debe a que las funciones indicatrices (y las probabilidades) se manejan como si fueran números reales no acotados. Tomando en consideración la geometría logística del simplex como un espacio euclidiano, este problema se resuelve satisfactoriamente.

INSTITUTO CÁNTABRO DE ESTADÍSTICA

Se ha puesto en marcha el Instituto Cántabro de Estadística (ICANE), cuyo director es Juan M. Rodríguez Poo.

ENCUENTROS SANTANDER-VALLADOLID-TOULOUSE

Entre los días 2 y 4 de febrero de 2005 ha tenido lugar en el Chalet Supaéro de Bonascre un encuentro de estadística matemática al que han acudido representantes de las Universidades de Cantabria, Valladolid y Toulouse. Este encuentro sucede al primer encuentro de estadística matemática organizado en 2003 por la Universidad de Cantabria en Santander.

ESCUELA DE VERANO

La escuela de verano "Empirical Processes: Theory and Statistical Applications" tuvo lugar en Laredo del 30 de agosto al 3 de septiembre. Fue organizada por el prof. Juan A. Cuesta Albertos (Universidad de Cantabria) y el prof. Carlos Matrán (Universidad de Valladolid) e incluida en los cursos de verano de la Universidad de Cantabria. Fue organizado bajo los auspicios de la European Mathematical Society, EMS; y las Universidades de Cantabria y Valladolid, y patrocinado por el Banco Santander Central Hispano, UNESCO ROSTE y la ciudad de Laredo.

PROYECTOS EUROPEOS

El 1 de enero de 2005 ha finalizado el proyecto europeo (una red temática) llamado Pro-Enbis (Promotion of European Network for Business and Industrial Statistics: <http://www.enbis.org/pro-enbis/>). En este proyecto han participado del orden de 30 Universidades y empresas europeas. Pro-Enbis surgió de ENBIS (European Network for Business and Industrial Statistics: <http://www.enbis.org/>).

CONVOCATORIA DE CONFERENCIAS SATÉLITE DEL INTERNATIONAL CONGRESS OF MATHEMATICIANS (ICM 2006)

Satellite Conferences are one of the most important scientific activities surrounding the celebration of every ICM. The Executive Committee of ICM2006 encourages all members of our mathematical community to get involved in the organization of scientific meetings and workshops on the occasion of the International Congress of Mathematicians that will take place in

Madrid on August 22-30, 2006.

There are a small number of requisites in order for one of these meetings to be accepted as a "Satellite Conference", on top of which we should mention the scientific quality and the interest of the research topics proposed, as well as the previous experience of the organizers. Other criteria to be considered are the following;

- * The Conference may be organized by any scientific group in any of the research areas of current interest in mathematics

- * The Conference must have a strong international projection and, therefore, it should be well balanced with respect to the participation of local and international specialists.

- * For strategic reasons, we ask for certain proximity in time to the ICM in Madrid . To that end, every Satellite Conference should start neither before July 17th nor after September 18th, 2006 . In any case, it should not overlap with the celebration of the ICM.

We also encourage people to coordinate efforts and avoid potential conflicts and/or overlapping with similar initiatives from research groups in the same scientific area.

Formal aspects of the application :

- * The proposal to organize a Satellite Conference must be submitted to the Executive Committee of ICM2006 (see below) indicating a title, names of the local and scientific committees, dates and place of celebration and a short abstract with the topics that would cover. The proposal must necessarily include the name of the member of the organizing committee of the event that will act as coordinator and as contact person with the Executive Committee of ICM2006 .

- * The applications will be evaluated by a subcommittee of experts and approved by the Executive Committee of ICM2006 .

- * Once a proposal has been accepted, all the relevant information about the Conference will appear at the webpage of ICM2006. The organizers will create and maintain their own webpage that will be accessible from that of ICM.

- * The search for financial support of a Satellite Conference is the sole responsibility of its organizing committee. The Executive Committee of ICM2006 would be happy to provide letters of support for whatever application to public or private institutions the organizers may need.

- * The deadline to submit a proposal for the Satellite Conferences expires on October 31st , 2005 .

Applications should be sent by electronic mail to
Fernando Soria
Departamento de Matemáticas

Universidad Autónoma de Madrid
Tel: +34 91 497 4796
Fax: +34 91 497 4889
e-mail: fernando.soria@uam.es

TESIS

APORTACIONES A LA TEORÍA DE LOS PROCESOS DE RAMIFICACIÓN MULTITIPO

Autor: Rodrigo Martínez Quintana (U. Extremadura).
Director/es: Miguel González Velasco (U. Extremadura) y Manuel Mota Medina (U. Extremadura).
Fecha de lectura: 13 de Febrero de 2004.
Lugar: Universidad de Extremadura
Publicación: vol. 202. Servicio de publicaciones de la Universidad de Extremadura. 2004. ISBN 84-7723-646-1 (en CD-ROM).

Resumen: Esta tesis doctoral se encuadra dentro de la Teoría General sobre Procesos de Ramificación en tiempo discreto, centrándose en la familia de Procesos de Ramificación Multitipo. En ella se realiza la introducción de un nuevo modelo que permite estudiar de forma unificada a una importante clase de Procesos de Ramificación Multitipo Homogéneos. A dicho modelo se le ha denominado Proceso de Ramificación Multitipo Generalizado Dependiente del Tamaño de la Población. Tras proceder a su definición e interpretación intuitiva, se prueba que es una cadena de Markov multitipo homogénea, estableciéndose resultados relativos a la comunicación entre sus estados y a la clasificación de los mismos, obteniéndose además los principales momentos condicionados del proceso. Posteriormente se considera el estudio de la probabilidad de extinción, así como del comportamiento límite, en primer lugar bajo el contexto aún más general de una Cadena de Markov Homogénea, para posteriormente particularizar los resultados obtenidos al Proceso de Ramificación Multitipo Generalizado Dependiente del Tamaño de la Población. Más concretamente, se determinan condiciones bajo las cuales la cadena se extingue con probabilidad uno y condiciones que garantizan su no extinción con probabilidad positiva. Además se propone una clasificación de la misma donde se distingue entre los casos subcrítico, crítico y supercrítico, atendiendo fundamentalmente a las diferencias de comportamiento en cuanto a la extinción. En base

a esta clasificación se analiza el comportamiento límite de la cadena convenientemente normalizada y de funcionales lineales asociados a la misma, proporcionándose resultados relativos a la convergencia casi segura, en L_r para $1 \leq r \leq 2$, en probabilidad y en distribución hacia límites no degenerados en cero y finitos.

Finalmente, se aplican los resultados obtenidos a casos particulares de Procesos de Ramificación Multitipo Modificados. Más concretamente se consideran nuevos modelos denominados Procesos de Ramificación Multitipo Controlados con Control Aleatorio y Reproducción Dependiente del Tamaño de la Población. En dicha aplicación se pone de manifiesto la potencia de las técnicas desarrolladas a lo largo de la tesis.

ESTUDIO DE LA FIABILIDAD DE SISTEMAS CON REPARACIONES POSPUESTAS

Autor: Eva Teresa López Sanjuán (U. Extremadura).
Directores: Inmaculada Torres Castro (U. Extremadura) y Attahiru Sule Alfa (U. Manitoba, Canada)
Lugar: Universidad de Extremadura
Fecha de lectura: 19 de Noviembre de 2004.

Resumen: La presente tesis doctoral se encuadra dentro del marco de la teoría de la fiabilidad y de la teoría de procesos estocásticos. En este trabajo se presenta un nuevo proceso estocástico, denominado proceso potencial, que permite modelar tiempos de operatividad y/o de reparación en un sistema en deterioro. Para este proceso estocástico se estudian sus principales propiedades, sobre todo las que se refieren a convergencia, órdenes estocásticos y clases de envejecimiento. La segunda parte de esta tesis se dedica al estudio de una nueva política de mantenimiento de sistemas reparables que está basada en el concepto de reparación pospuesta. La idea principal de esta política es que la reparación de los fallos que sufre el sistema se pospone hasta un momento en que la reparación sea menos costosa. Esta política de mantenimiento se desarrolla bajo dos modelos. Para ambos modelos,

se analiza el beneficio esperado por período laborable con el fin de determinar cuál es la política de reparación óptima. Asimismo, se ilustran los resultados teóricos con ejemplos numéricos.

ESTIMACIÓN DE DENSIDADES: ALGUNOS RESULTADOS EXACTOS Y ASINTÓTICOS

Autor: José Enrique Chacón Durán (U. Valladolid).
Directores: Agustín García Nogales (U. Extremadura) y Jesús Montanero Fernández.(U. Extremadura).
Lugar: Universidad de Extremadura
Fecha de lectura: 10 de Diciembre de 2004

Resumen: La memoria estudia el problema de estimación de densidades en Estadística Matemática. En un marco muy general se expone la relación entre el concepto de suficiencia parcial y la estimación de densidades y se prueba también que la hipótesis de que la densidad esté en algún espacio L_p no es distinguible, en el sentido de Hoeffding y Wolfowitz. Para el caso concreto de la estimación de una densidad real univariante se estudia el estimador núcleo, con especial énfasis en el problema de elección de un ancho de banda óptimo. Bajo condiciones muy poco restrictivas se prueba la existencia de dicho ancho de banda óptimo, así como sus propiedades asintóticas y se describe la sucesión de anchos de banda óptimos hasta el orden del tercer término dominante cuando se utiliza un núcleo de cualquier orden k . Se describe la metodología bootstrap, presentando estructuras estadísticas apropiadas para el estudio de los estimadores bootstrap, tanto para tamaño muestral fijo como para el caso asintótico, y se aplica dicho método al problema de elección automática del ancho de banda. Así, se proporcionan dos nuevos selectores bootstrap de ancho de banda, utilizando un ancho de banda piloto de tipo local, y se estudian sus propiedades asintóticas, mediante teoremas de distribución límite del error relativo respecto de la sucesión de anchos de banda óptimos, y su comportamiento para tamaños muestrales finitos, mediante un estudio comparativo por simulación.

DISEÑO DE REDES DE TELECOMUNICACIONES Y DETERMINACIÓN DE TABLAS DE ENRUTAMIENTO CON ESTOCASTICIDAD EN EL FLUJO ORIGEN-DESTINO ENTRE NODOS

Autor : Juan Francisco Monge Ivars
Director : Laureano Escudero Bueno.
Lugar: Universidad Miguel Hernández de Elche.
Centro de Investigación Operativa.

Resumen: Se presenta un modelo estocástico tres etapas 0-1 mixto para formalizar matemáticamente el problema de localización de routers en una red de telecomunicaciones y el enrutamiento de paquetes a través de la red, bajo un ambiente de incertidumbre. La estocasticidad en el flujo de entrada a los nodos origen está representada por un conjunto de escenarios representativos.

Se expone la complejidad del problema tratado y la imposibilidad de obtener soluciones exactas en problemas de medianas dimensiones.

Se presenta un algoritmo heurístico, combinación de un procedimiento de generación de columnas y un algoritmo genético, para la resolución del problema.

Se concluye indicando el interés del problema tratado y su modelización como alternativa a los procedimientos habituales de resolución del problema mediante simulación.

MÉTODOS ESTADÍSTICOS PARA EL ANÁLISIS DE LIGAMIENTO GENÉTICO MEDIANTE POOLS DE ADN

Autor: Carlos Enrique Carleos Artime
Director: Norberto Corral Blanco
Lugar: Universidad de Oviedo. Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática

Resumen: Con objeto de mejorar sus cabañas, las asociaciones de ganaderos recurren a menudo a los conocimientos aportados por la Genética sobre la herencia de caracteres de interés económico. A tal fin, rutinariamente efectúan mediciones sobre los animales y recogidas de muestras de tejido. Por otro lado, el trabajo de laboratorio requerido por los métodos genéticos tradicionales (genotipado individual) resulta excesivamente caro para la mayoría de las asociaciones, lo que las obliga a reducir el número de muestras analizadas a unos pocos ejemplares. Se ha propuesto un nuevo procedimiento de análisis (genotipado por lotes) que, al genotipar simultáneamente muchos ejemplares, permite reducir notablemente los costos, a cambio de prescindir de una parte importante de la información obtenida mediante los métodos tradicionales. El propósito de esta tesis consiste en explorar diversas estrategias para aprovechar,

en la medida de lo posible, la información aportada por una muestra de genotipado por lotes. La memoria comienza con un capítulo de preliminares que expone las técnicas estadísticas utilizadas y los conceptos generales de Genética necesarios para la comprensión del trabajo. A continuación, se abordan los diferentes métodos de análisis del genotipado por lotes que hacen uso de los fenotipos en los lotes a través de momentos muestrales; se obtienen propiedades matemáticas de los métodos sugeridos por otros autores y se proponen nuevos métodos. En el capítulo siguiente se muestra la aplicación de técnicas de genotipado individual en los algoritmos diseñados para el genotipado por lotes a través de diversos enfoques basados en cadenas de Markov. Al final, se realiza un estudio comparativo mediante simulación.

TESTS NO PARAMÉTRICOS BASADOS EN UNA MEDIDA DE IGUALDAD ENTRE FUNCIONES DE DENSIDAD

Autor: Pablo Martínez Camblor
Director: Norberto Corral Blanco
Lugar: Universidad de Oviedo

Resumen: En esta tesis se define una medida de similitud (distancia) entre funciones de densidad que evalúa el área que tienen en común dos distribuciones de probabilidad absolutamente continuas. A partir de esta medida se construye un nuevo estadístico que permite cuantificar cuanto se parece una función de densidad y su estimación o, la similitud entre varias estimaciones independientes de una misma función de densidad. Posteriormente se realiza un estudio teórico acerca de las propiedades más importantes de este estadístico, calculándose su media y varianza y la distribución asintótica, así como comprobándose su convergencia casi segura, mediante técnicas de procesos estocásticos. Estos resultados teóricos se emplean en la construcción de una familia de test de hipótesis para estudiar el ajuste de una distribución a un modelo teórico (bondad de ajuste) y la identidad entre varias poblaciones. En cada uno de estos contrastes se analiza su potencia empleando métodos de simulación y se estudian las condiciones para que tengan un buen comportamiento. Finalmente, mediante técnicas similares a las anteriores, se demuestra la consistencia del método bootstrap suavizado y se evalúa su eficiencia práctica a través de simulaciones.

Este trabajo se desarrolla para variables unidimensionales, si bien, en su parte final se dan

algunas ideas para la generalización al caso multivariante. El método propuesto tiene la ventaja de poder comparar simultáneamente más de dos funciones de densidad, pudiendo considerarse una generalización de la medida $L1$ y, resultando ser, en ocasiones, más potente que los métodos usualmente utilizados en este tipo de contrastes.

DESCOMPOSICIÓN DE LA TRANSITIVIDAD EN ESTRUCTURAS DE PREFERENCIA ADITIVA

Autor: Susana Díaz Vázquez
Directores: Susana Montes Rodríguez, Bernard De Baets, Pedro Gil Álvarez
Lugar: Universidad de Oviedo. Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática

Resumen: En esta tesis se ha llevado a cabo un estudio de la propagación de la transitividad dentro de las estructuras de preferencia aditiva, por tratarse éste de un concepto fundamental en la teoría de la decisión multicriterio.

En el caso clásico se sabe que la transitividad de la relación de preferencia extendida asociada a una estructura de preferencia se puede caracterizar por las transitividades de la relación de preferencia estricta y la relación de indiferencia. Se ha introducido una definición de transitividad que generaliza la que habitualmente se emplea en la literatura y hemos buscado los tipos más fuertes de transitividad que se pueden asegurar para las relaciones de preferencia estricta e indiferencia suponiendo que la relación de preferencia extendida satisface distintos tipos de transitividad. Además se ha probado que nuestras implicaciones son las más fuertes posibles: hemos obtenido contraejemplos generales que ponen de manifiesto que no se puede obtener tipos más fuertes de transitividad bajo las mismas condiciones. Puesto que la influencia de la hipótesis de completitud está clara ya en el caso clásico, también hemos estudiado la influencia de esta propiedad en la descomposición de la transitividad en ambiente difuso. Hemos establecido las implicaciones más fuertes posibles tanto en ausencia de completitud como bajo completitud débil.

Por otra parte, hemos estudiado la implicación inversa. Hemos demostrado que la transitividad de la relación de preferencia extendida no se deduce ni siquiera de los tipos más fuertes de transitividad que se pueden suponer para las relaciones de preferencia estricta e indiferencia, de modo que la caracterización conocida en el caso clásico no

admite una traslación directa al caso borroso.

ANÁLISIS MULTIVARIANTE Y FUNCIONAL DEL PROCESO DE INERCIA EN MODELOS DE FUNCIÓN DE TRANSFERENCIA

Autor: Francisco M. Ocaña Peinado.

Director: Mariano J. Valderrama Bonnet.

Lugar: Dpto. Estadística e I.O. Universidad de Granada

Fecha de lectura: 1 de Febrero de 2005.

Resumen: La introducción en 1970 de los modelos ARIMA en la obra de los ingenieros G.E.P. Box y G.M. Jenkins, a partir de trabajos realizados en la década anterior, supuso un nuevo enfoque en el tratamiento de las series temporales, en cuanto que frente al concepto clásico en el que la serie se ajustaba a un cierto patrón o tendencia cuyas desviaciones constituían el núcleo del análisis, esta nueva metodología partía de que la herramienta básica para la modelización y el análisis la constituían las correlaciones existentes entre las variables aleatorias seriadas.

Pese a las limitaciones de estos modelos (estructura lineal, carácter temporal discreto, necesidad de disponer de series históricas largas, etc), los ARIMA se han aplicado en numerosas áreas científicas, tales como la Ingeniería, Economía, Ciencias del Comportamiento, etc., siendo una herramienta de predicción eficaz a corto plazo.

La formulación básica de un ARIMA puede concebirse como que la serie objeto de estudio se ajusta a un proceso estocástico de segundo orden (con momentos de primer y segundo orden finitos) que representa la respuesta u output de un sistema lineal a un impulso o input que sea ruido blanco, siendo la función de transferencia del filtro cociente de dos polinomios de grado finito, es decir, de tipo racional. En el propio texto de Box y Jenkins se plantea una generalización de estos modelos univariantes al caso en que la variable de respuesta temporal pueda explicarse por una o varias variables aleatorias que no sean de ruido blanco, es decir, se propone la extensión del modelo de regresión múltiple a caso dinámico. Esta nueva formulación da lugar a los denominados modelos de función de transferencia (MFT) o también, fundamentalmente en el ámbito econométrico, modelos de regresión dinámica.

El MFT consta de dos términos aditivos, el primero recoge la parte explicada por variables exógenas y el segundo, denominado proceso de inercia, representa la perturbación aleatoria. La

parte explicativa se plantea inicialmente con una estructura de retardos linealmente distribuidos a partir de la cual se obtienen formas más compactas en términos de modelos de retardos racionalmente distribuidos (Pankratz, 1991; Otero, 1993). En cuanto al proceso de inercia, la formulación básica parte de que se ajusta a un modelo ARIMA general.

En base a las limitaciones antes comentadas de los ARIMA, la presente tesis doctoral tiene por objeto desarrollar modelos más adecuados para representar la parte no explicada, permitiendo así optimizar su capacidad predictiva y formular modelos más generales. Concretamente, el tratamiento que se introduce está basado en la descomposición del el proceso de inercia en términos de sus componentes principales, tanto desde un punto de vista discreto como funcional y, como se demuestra en la memoria, este planteamiento proporciona una mayor precisión del modelo predictivo final.

La tesis doctoral está articulada en tres capítulos, estando el primero dedicado a introducir los conceptos básicos sobre MFT, ilustrándose con una aplicación a las Ciencias del Comportamiento, donde se elaboran sendos modelos dinámicos para explicar los síntomas del lupus eritematoso en función del estrés de un grupo de pacientes medido sobre una escala numérica protocolizada, y viceversa, lo que permite clasificarlos en tres grupos en función del orden del retardo en la parte explicada.

En el segundo capítulo se aborda la cuestión antes mencionada consistente en modelizar el proceso de inercia en términos de componentes principales a partir de su consideración como proceso estocástico de segundo orden. Esto permite lograr un grado de aproximación tan alto como se desee a medida que se incorporan nuevas variables a su representación numerable. La eficacia del método se muestra con dos aplicaciones, una de tipo financiero, donde se elabora un modelo para predecir la evolución del IBEX35 a partir del índice SP500 de la Bolsa de Nueva Cork, y otra en el campo medioambiental para predecir la concentración de polen de ciprés y de polen de olivo a partir de dos inputs alternativos aunque correlados entre sí: temperatura y horas de sol.

El capítulo tercero aborda el enfoque funcional del problema y, tras introducir las nociones básicas sobre modelos CARMA, dedicando especial atención al de orden uno, se presenta la versión del MFT en tiempo continuo, en términos de sistemas estocásticos lineales. A continuación se desarrolla el proceso de inercia en términos de componentes principales como serie de Karhunen-

Loève, obteniéndose aproximaciones en tiempo continuo, validándose la optimización de esta modelización mediante un ejemplo simulado.

ESTIMACIÓN DE MODELOS DE VOLATILIDAD ESTOCÁSTICA Y MODELOS DE COMPONENTES INOBSERVADOS CONDICIONALMENTE HETEROCEDÁSTICOS

Autora: Carmen Broto Pelegrín
Directora: Esther Ruíz Ortega
Lugar: Dpto. de Estadística, Universidad Carlos III
Fecha de lectura: 23 de septiembre de 2004

Resumen: En esta tesis se analizan algunos aspectos relevantes para el análisis de series temporales condicionalmente heterocedásticas. En primer lugar, se hace una revisión detallada de los métodos de estimación de los modelos de volatilidad estocástica (SV), cuyas propiedades, ventajas y desventajas comparativas no se habían analizado recientemente. Los distintos métodos se ilustran con datos simulados y con una aplicación empírica a rendimientos diarios del S&P 500. Seguidamente se propone un nuevo modelo de componentes inobservados y perturbaciones condicionalmente heterocedásticas de tipo asimétrico. En éste se considera que la serie es un modelo de paseo aleatorio más ruido, cuyas innovaciones siguen procesos GQARCH(1,1). Tras analizar las propiedades estadísticas del modelo, se obtienen la distribución asintótica y algunas propiedades para muestras finitas del estimador PMV. Finalmente, se propone la utilización de las correlaciones de los cuadrados de los residuos auxiliares para detectar la presencia de heterocedasticidad en cualquier componente y se presentan aplicaciones empíricas del modelo a diversas series financieras y de precios. Finalmente se generaliza el modelo mediante la incorporación de un componente estacional homocedástico, que puede ser determinista o estocástico. Analizadas las propiedades de los estimadores en muestras finitas, se ajusta el modelo a ocho series de inflación.

ANÁLISIS BAYESIANO DE SISTEMAS DE COLAS

Autora: María Concepción Ausín Olivera
Directores: Rosa Elvira Lillo Rodríguez y Michael Peter Wiper
Lugar: Dpto. de Estadística, Universidad Carlos

III
Fecha de lectura: 19 de Febrero de 2004

Resumen: En esta tesis, se proponen procedimientos Bayesianos para la inferencia, predicción y diseño de diferentes sistemas de colas que incluyen modelos con distribuciones generales para el tiempo entre las llegadas y/o el tiempo de servicio, con uno o varios servidores y con capacidad finita e infinita. Con las observaciones tomadas del proceso de llegadas y de servicio en el sistema, en primer lugar, se desarrollan métodos de estimación Bayesiana de densidades que permiten aproximar las distribuciones desconocidas asociadas a estos procesos. La estimación está basada en mixturas de distribuciones de tipo PH con un número desconocido de componentes que se aborda con diferentes métodos MCMC de dimensión paramétrica variable. Se analiza la congestión del tráfico en cada sistema observado basándose en la inferencia desarrollada sobre los parámetros del mismo. En caso de que el sistema sea estable, se describen procedimientos para la estimación Monte Carlo de las medidas de interés de cada modelo en equilibrio, tales como las distribuciones predictivas del número de clientes en el sistema, del tiempo de espera en cola o de la longitud de los periodos de ocupación. La predicción Bayesiana desarrollada se basa en diferentes propiedades de los sistemas en los que intervienen distribuciones PH y en algunos resultados conocidos de la Teoría de Colas clásica. En los modelos de colas con varios servidores, se abordan también problemas de diseño en los que el objetivo consiste en decidir el número óptimo de servidores que minimizan una función de coste que depende de las distribuciones estacionarias estimadas. La metodología desarrollada se ilustra con datos reales procedentes de un hospital geriátrico de Londres y un establecimiento bancario de Madrid. Por último, se analiza el comportamiento transitorio y el periodo de ocupación de modelos de colas completamente generales con un único servidor. Se obtiene un resultado que permite extraer numéricamente las raíces complejas de unas ecuaciones implicadas en las distribuciones transitorias de interés.

ATÍPICOS, CAMBIOS ESTRUCTURALES Y DISCRIMINACIÓN EN SERIES TEMPORALES MULTIVARIANTES

Autora: Pedro Galeano San Miguel
Director: Daniel Peña Sánchez de Rivera.

Lugar: Dpto. de Estadística, Universidad Carlos III

Fecha de lectura: 17 de Mayo de 2004

Resumen: Esta Tesis se centra en el análisis de algunas propiedades dinámicas de las series temporales univariantes y multivariantes. Concretamente, se analizan tres problemas habituales en el análisis de una serie temporal: (1) observaciones atípicas, que son aquellas observaciones afectadas por fenómenos dinámicos no relacionadas con la estructura dinámica de la serie temporal, (2) cambios en la varianza marginal de la serie, que corresponden a la observación de diferentes periodos en cuanto a la varianza de la serie y (3) criterios de selección de modelos, que consisten en la determinación de un modelo adecuado para una serie temporal. Las aportaciones de esta Tesis contribuyen a la resolución de estos tres problemas. Se propone un método de detección y estimación de datos atípicos en series multivariantes basado en la proyección del vector de series en direcciones adecuadas. El método propuesto tiene la ventaja de no requerir un modelo inicial para la serie. De esta manera se puede limpiar la serie de atípicos y posteriormente, obtener un modelo para estimar conjuntamente los parámetros, el tamaño y efectos de los atípicos. Para cambios de varianza, se proponen dos algoritmos para la detección y estimación de cambios en la varianza marginal de una serie multivariante. También se proponen contrastes para identificar cambios de correlaciones entre las componentes del vector de series. Finalmente, se propone un enfoque para seleccionar modelos, basado en análisis discriminante que unifica criterios clásicos y Bayesianos obtenidos mediante diferentes enfoques. Además, este planteamiento permite añadir un término de corrección en estos criterios que no modifican sus propiedades asintóticas pero que hacen mejorar su comportamiento en muestras finitas. El método se aplica para seleccionar modelos autorregresivos y de media móvil y un tipo especial de modelos no lineales, los modelos autorregresivos por umbrales.

CONTRASTES BASADOS EN RECORRIDOS PARA RAÍCES UNITARIAS Y COINTEGRACIÓN

Autora: Ana Elisabeth García Sipols.
Directores: Álvaro Escribano Sáez y Felipe M. Aparicio Acosta
Lugar: Dpto. de Estadística, Universidad Carlos III

Fecha de lectura: 12 de Febrero de 2004

Resumen: El tema central de la tesis es el desarrollo de contrastes robustos para raíces unitarias y cointegración uniecuacional. Para ello, se presentan los principales inconvenientes de los contrastes tradicionales ante la presencia de no linealidades, atípicos y cambios estructurales, efectos que producen distorsiones, bien en el tamaño, bien en la potencia de dichos contrastes. Se propone una metodología alternativa no paramétrica para contrastar la presencia de raíz unitaria antes las desviaciones de los supuestos clásicos mencionadas. Con esta metodología se intenta explotar una extensión del concepto de integración, en el que toman especial importancia los instantes de llegada de eventos significativos (cambios bruscos de nivel en un instante de la serie). Por ejemplo, cuando llega un nuevo máximo o mínimo en la serie de partida. Esta propiedad permite distinguir entre series estacionarias y no estacionarias, debido a la existencia de una importante llegada de nuevos máximos y mínimos en series con tendencias estocásticas, al contrario de lo que ocurre en las series con corta memoria. En base a esta propiedad, se plantean tres contrastes no paramétricos que son robustos frente a las anomalías descritas, hecho que se comprueba mediante experimentos Monte Carlo. Se analizan algunas propiedades asintóticas, y finalmente, se ilustra su buen funcionamiento mediante aplicaciones empíricas con series reales. Se amplía dicha metodología a contrastes de cointegración, proponiendo un contraste que hereda las propiedades de robustez de los contrastes anteriores. Se comprueban dichas buenas propiedades mediante nuevos experimentos Monte Carlo, y se analizan series reales univariantes así como series bivariantes.

MODELLING RISKS IN DISEASE MAPPING (MODELIZACIÓN DE RIESGOS EN LA REPRESENTACIÓN CARTOGRÁFICA DE ENFERMEDADES)

Autora: Berta Ibáñez Beróiz
Directoras: M. Dolores Ugarte y Ana F. Militino
Lugar: Universidad Pública de Navarra
Fecha de lectura: 22 de marzo de 2004

Resumen: La presentación geográfica de los riesgos de mortalidad es una herramienta básica en el análisis de datos de mortalidad. El creciente interés por la distribución de ciertas enfermedades entre los epidemiólogos y profesionales de la

salud pública y el desarrollo de programas informáticos han dado lugar a grandes avances en este campo.

En esta tesis se propone una estrategia de análisis de datos de mortalidad basada en la causa de la presencia de sobredispersión (en inglés, “overdispersion”), fenómeno que surge cuando el modelo homogéneo de Poisson no recoge toda la variabilidad presente en los datos. En esta memoria se consideran tres modelos: el modelo discreto de mixturas de Poisson, que incorpora la variabilidad extra en un modelo de Poisson mediante un efecto aleatorio de distribución discreta, los modelos markovianos, que permiten incorporar la dependencia espacial en los riesgos y el modelo de Poisson inflado en cero, que sirve para analizar causas de mortalidad muy poco comunes. Para el modelo discreto de mixturas de Poisson, se proponen distintos métodos para la construcción de intervalos de confianza para los riesgos que están basados en técnicas asintóticas y bootstrap. En particular, los intervalos de confianza BC-bootstrap alcanzan unas probabilidades de cobertura parecidas a las nominales. Para detectar la posible existencia de dependencia espacial, se propone el uso de un test score desarrollado para diversos modelos markovianos (modelos CAR) desde un enfoque empírico-Bayesiano, que se caracteriza por ser simple ya que sólo requiere ajustar el modelo bajo la hipótesis nula y además es potente. Finalmente, para contrastar la necesidad de utilizar el modelo de Poisson inflado en cero, se construye un contraste bootstrap-paramétrico que es potente y tiene mejores propiedades que los correspondientes test score y test de la razón de verosimilitudes derivados en este contexto. Los resultados obtenidos en esta tesis se ilustran con datos de mortalidad de la provincia de Navarra

ANÁLISIS DE INDICADORES SOCIALES: APLICACIÓN AL CASO ESPAÑOL

Autor: Luis Felipe Rivera Galicia
Directores: José Miguel Casas Sánchez y José Javier Núñez Velázquez
Lugar: Universidad Alcalá. Fac. Cc. Económicas y Empresariales. Departamento de Estadística, Estructura Económica y O.E.I.

Resumen: En el primer capítulo de la Tesis, se presentan los aspectos fundamentales en los que se centra el análisis de los Indicadores Sociales, que se constituyen como herramientas básicas para la medición del Bienestar. Se hace un recorrido histórico por las definiciones de

indicador social y se propone una nueva definición alternativa que engloba las anteriores. Además, se justifica la necesidad de la elaboración de un Sistema de Indicadores Sociales, introduciendo la problemática de su sintetización en una única medida.

En el segundo capítulo, se introduce el concepto de distancia como elemento determinante para la elaboración de un indicador sintético por comparación. A partir de este concepto, se selecciona la distancia DP2 como herramienta para la construcción de un indicador sintético, realizando un extenso análisis de la misma, en el que se presentan diversos resultados que permiten modificar su formulación original, para adaptarla a la metodología general que se propone en el capítulo siguiente. Se demuestra que la solución de la distancia DP2 no es única, y se establece un criterio de optimalidad que fija su resultado de manera unívoca.

En el tercer capítulo, se presenta una propuesta metodológica general para la construcción de indicadores sintéticos. Esta propuesta se divide en dos partes: la primera consiste en la homogeneización de las variables de partida, y la segunda se basa en la aplicación de una clase de operadores matemáticos que sirven para agregar la información contenida en las variables previamente homogeneizadas. De las posibles funciones de agregación, se seleccionan las lineales, conforme a lo que es habitual en los trabajos de este campo, y se analizan las propiedades que presentan. Se estudia cómo se pueden generar los diferentes métodos de elaboración de indicadores sintéticos dentro del marco metodológico propuesto, y se diseñan diversas alternativas que permiten dotar de capacidad de análisis dinámico al indicador sintético resultante.

En el capítulo cuarto, se lleva a cabo una aplicación de la metodología general propuesta y se estudia, a partir de ciertos indicadores de desigualdad, la evolución del nivel de vida en las Comunidades Autónomas en España, teniendo en cuenta el nivel medio de gasto que se observa en las mismas. Se clasifica a las Comunidades Autónomas en grupos, según la evolución de sus respectivos niveles de vida en relación con el nivel de vida del Total Nacional.

ANÁLISIS BAYESIANO DE LOS MODELOS DE REGRESIÓN POLITÓMICOS Y DE POISSON ROBUSTOS

Autora: Dña. Julia García Galisteo.
Directores: Dr. D. Fco. Javier González-Torre y

Dra. Dña. M. Lina Martínez García.
Lugar: Universidad de Málaga. Departamento de Estadística e Investigación Operativa
Fecha de lectura: 18 de febrero de 2005.

Resumen: En este trabajo se presenta un estudio, desde el punto de vista bayesiano, de los modelos de regresión dicotómicos y su generalización al caso politómico y de los modelos de regresión de Poisson.

En primer lugar, se generaliza el modelo usual de regresión dicotómico considerando funciones de nexo arbitrarias que verifiquen ciertas condiciones mínimas, al que denominamos modelo determinista. Se realiza, además, el estudio de posibles observaciones anómalas y se considera como criterio a la hora de elegir entre varias funciones de nexo el preferir aquella que sea más inmune a estas observaciones, lo cual conduce a la consideración de funciones de nexo que denominaremos robustas. Se propone para ello dos familias de distribuciones como posibles funciones de nexo, la familia t-Student y la familia de distribuciones e Box-Tiao. De ellas se elegirán aquellas que sean más robustas que las funciones de nexo clásicas Logística y probit.

Para el modelo de regresión de Poisson y el de regresión politómica, también se generalizan los modelos usuales, es decir, el modelo Loglineal y el Logístico Multivariante, respectivamente, introduciendo como funciones de nexo generales las pertenecientes a una cierta clase que se introduce en el modelo dicotómico.

Un problema importante que habitualmente se presenta cuando se analizan datos heterogéneos con estos modelos es el de la sobredispersión. Este problema se enfoca desde una perspectiva bayesiana considerando modelos bayesianos jerárquicos.

ESTUDIO DEL CUESTIONARIO DE EVALUACIÓN DEL PROFESORADO DE LA UPV MEDIANTE OPINIÓN DE LOS ESTUDIANTES. TRATAMIENTO ESTADÍSTICO

Autora: Mónica Martínez Gómez
Directores: José Miguel Carot Sierra y José Jabaloyes Vivas
Departamento: Estadística e Investigación Operativa Aplicadas y Calidad de la Universidad Politécnica de Valencia.
Fecha de lectura: 14 de Marzo de 2005

Resumen: La mejora de la calidad de las instituciones constituye un importante reto para

las universidades en los próximos años y el instrumento potencial para ello lo constituye la evaluación institucional en general y la evaluación de la actividad docente en especial.

El objetivo general del trabajo es desarrollar una metodología estadística adecuada para extraer, analizar e interpretar la información contenida en el Cuestionario de Evaluación Docente mediante Opinión de los Alumnos de la UPV, con la finalidad de optimizar su utilización práctica.

Mediante el encadenamiento y utilización conjunta de varias técnicas estadísticas multivariantes, se han propuesto recomendaciones orientadas a incrementar la eficacia en la utilización del cuestionario como herramienta de medida e indicador de calidad de la enseñanza en la universidad, para introducir actuaciones de mejora continua en los procesos educativos de la UPV.

APLICACIÓN DE MÉTODOS ESTADÍSTICOS MULTIVARIANTES PARA LA MODELACIÓN Y LA MONITORIZACIÓN DE UN REACTOR DISCONTINUO SECUENCIAL PARA EL TRATAMIENTO DE AGUAS RESIDUALES

Autor: Daniel Aguado García.
Directores: José Ferrer Polo (Dpto. Ingeniería Hidráulica y Medio Ambiente, Universidad Politécnica de Valencia), Aurora Seco Torrecillas (Dpto. Ingeniería Química, Universidad de Valencia) y Alberto J. Ferrer Riquelme (Dpto. Estadística e I.O. Aplicadas y Calidad, Universidad Politécnica de Valencia)
Fecha de lectura: 15 de Marzo de 2005

Resumen: Se ha desarrollado un sistema para la monitorización, diagnóstico en tiempo real y mejora de la calidad de un reactor discontinuo secuencial (SBR), operado para eliminar biológicamente el fósforo del agua residual, mediante métodos de control estadístico de procesos multivariantes basados en técnicas de proyección sobre estructuras latentes. Adicionalmente, se ha estudiado y comparado diversas técnicas para predecir la evolución de las variables de calidad (fósforo, potasio y magnesio), determinadas analíticamente en laboratorio, a lo largo de los lotes aprovechando la información contenida en las variables registradas por los sensores instalados en el reactor. Los resultados obtenidos en esta tesis han permitido comprobar que los modelos predictivos desarrollados reproducen satisfactoriamente la evolución de las variables de calidad medidas en laboratorio, y el

sistema de monitorización es capaz de detectar e identificar las variables responsables de las anomalías ocurridas durante la operación del SBR. Esta tesis doctoral supone una importante aportación en el campo de la modelación y monitorización de los procesos de tratamiento de

aguas residuales, permitiendo la obtención de un sistema integral de monitorización y predicción que permita una operación eficiente del proceso.

OPORTUNIDADES DE TRABAJO

UNIVERSIDAD REY JUAN CARLOS

El Grupo de Estadística y Ciencias de la Decisión (<http://bayes.escet.urjc.es>) de la Universidad Rey Juan Carlos ofrece plazas de profesores doctores para el curso 05/06. Se buscan especialistas en las materias de Estadística e Investigación Operativa. Los candidatos interesados deben estar en posesión del título de doctor (o equivalente) y deben remitir un Curriculum Vitae a Antonio Alonso-Ayuso (antonio.alonso@urjc.es) antes del 1 de abril de 2005.

UNIVERSITÄT GÖTTINGEN

At the Institute for Mathematical Stochastics of the Mathematical Faculty of the Georg-August-Universität Göttingen, a position is open for a W 3 Professorship (BesGr. W3 BBesO) in the field of Stochastics and its Applications, starting 1st of October 2005.

Applicants should have an outstanding research and publication record. The candidate is expected to take part in all teaching, research and administration tasks of the institute. Furthermore, he or she should represent an actual field in stochastics and its applications (possibly in life or computational sciences) which is strongly related to statistics. Active cooperation with the Center of Statistics is expected.

The placement prerequisites for a professorship appointment are given by § 25 NHG (Nds. GVB1. 2002 page 286). Details will be given on request.

The University of Göttingen is aiming to increase the share of female faculty members in research and teaching. Thus qualified women are especially encouraged to apply. Part time employment may also be possible. Disabled candidates will be given preference in the case of equal qualifications.

Applications with CV, list of publications, description of third party funded projects, teaching and research resume should be sent by 1th of May

2005 to the Dean of the Mathematics Faculty, Bunsenstr. 3-5, 37073 Göttingen.

UNIVERSIDADES DE GALICIA

Pre-announcement of three postdoctoral positions in Statistics in Spain

Three post-doctoral positions in Statistics will be offered, one at each of the three universities in Galicia (NW of Spain), as part of the programme "Isidro Parga Pondal" issued by the Government of Galicia.

Position at the University of Vigo

Flexible Statistical Inference Research Unit (FSI.RU). Scientist-in-charge: Jacobo de Uña-Álvarez, e-mail: jacobo@uvigo.es. Research area: Nonparametric (and semiparametric) statistical inference, including goodness-of-fit, curve estimation, multivariate survival analysis, spatial statistics, resampling methods, and time series.

Webpage:

<http://webs.uvigo.es/jacobo/jacobo.html>

Position at the University of Santiago de Compostela Statistics Group (STAT). Scientist-in-charge: Wenceslao González-Manteiga, e-mail: wenceslao@usc.es. Research area: Biostatistics and/or Spatial statistics.

Group webpage: <http://eio.usc.es/pub/stat>

Position at the University of A Coruña

Statistical Inference Research Group (SIRG). Scientist-in-charge: Ricardo Cao, e-mail: rcao@udc.es. Research area: Statistical Inference.

Group webpage:

http://www.udc.es/dep/mate/Dpto_Matematicas/Investigacion/ie.htm

The total maximum gross salary will be about 26075 euros per year, and the contract period will be of two years (with the possibility of getting prolonged). Deadline for applications is expected to be mid-May 2005.

The duties of the contracts are research activities, with the possibility of teaching (no more than 120 hours per year) if the selected candidate is willing

to.

The conditions for candidates are:

a) Being citizen of the European Union (EU) or having residence permission in Spain.

b) Having a Doctor degree. For Doctor degrees obtained in non EU countries the candidate needs homologation by the Spanish Ministry of Education.

c) A research stay period (after doctoral degree) of at least 6 months at institutions other than which the doctoral degree was obtained at.

Possible candidates are strongly encouraged to

contact the scientists-in-charge (Jacobo de Uña-Álvarez: jacobo@uvigo.es, Wenceslao González-Manteiga: wenceslao@usc.es and Ricardo Cao: rcao@udc.es) as soon as possible.

Further information on the application procedure will be available, in due time, at the Research and Development website of the Galician Government: <http://www.sxid.org>.

The three scientist-in-charge would be grateful if you could bring this position to the attention of all who might be interested in.

PUBLICACIONES Y SOFTWARE

GEOSTATISTICAL ANALYSIS OF COMPOSITIONAL DATA

Pawlowsky-Glahn, V. and Olea, R. A. (2004). Geostatistical Analysis of Compositional Data, Oxford University Press, Studies in Mathematical Geology Series 7, New York, NY (USA), ISBN 0195171667, 208 p.

El libro ofrece un marco general para el estudio de composiciones regionalizadas y, en particular, para el cokrigado de un vector completo. Introduce los conceptos básicos relacionados con composiciones regionalizadas, presenta la definición de estructura de covarianza espacial para las mismas, y analiza relaciones de correlación espacial cruzada nula. Finalmente, describe la técnica de estimación espacial conocida como cokrigado o cokriging, e ilustra los procedimientos mediante un ejemplo geológico.

MÉTODOS ESTADÍSTICOS EN INGENIERÍA

Rafael Romero Villafranca y Luisa Rosa Zúnica Ramajo. Métodos Estadísticas en Ingeniería. Universidad Politécnica de Valencia. Servicio de Publicaciones, 334 p. (02/2005).

El libro está planteado para cubrir las materias estadísticas más importantes para un ingeniero, partiendo de conocimientos estadísticos previos nulos y en una asignatura con pocos créditos. La obra incluye el tratamiento a nivel operativo de temas avanzados, como efectos de dispersión en anova, "orthogonal arrays", variables cualitativas en modelos de regresión, etcétera... La mayor parte del libro se centra en el análisis de

problemas reales, fruto de la experiencia profesional de los autores. Cada capítulo se completa con un anejo con ejemplos resueltos y las respuestas a numerosas cuestiones de autoevaluación, posibilitándose al lector el acceso una dirección en Internet, a los datos de los ejemplos reales manejados en el libro. (Consultar Prefacio e Índice detallado en <http://ttt.upv.es/~rromero/descargas/>)

INVESTIGACIÓN OPERATIVA. PROBLEMAS Y EJERCICIOS RESUELTOS

Martin Martin, Q. Santos Martín, M^a T. y de Paz Santana y. del R. Investigación Operativa. Problemas y ejercicios resueltos. PEARSON. Prentice Hall (Prentice Práctica). Madrid, 2005

Este libro de "Problemas de Investigación Operativa" recoge gran variedad de ejercicios, entre otros los ejercicios que aparecen en el libro de "Investigación Operativa" del profesor Quintín Martín Martín, publicado en esta misma editorial. Los alumnos a los que está orientado son: estudiantes de la Licenciatura en Matemáticas, Diplomatura de Estadística, Ingeniería Informática y de todas las áreas relacionadas con la Economía y Empresa. Los temas comienzan con una exposición teórica sencilla de los conceptos que van a ser utilizados en la resolución de los ejercicios. Los ejercicios resueltos se han elegido buscando interés y variedad en ellos.

Se incorpora al libro un CD que contiene una aplicación "Invesope" que nos resuelve problemas, mediante tablas, de Programación Lineal y Análisis de Sensibilidad.

PLANIFICACIÓN DE TURNOS EN UN AEROPUERTO: USO DE SIMULACIÓN Y METAHEURÍSTICOS

Silvia Casado Yusta. Planificación de Turnos en un Aeropuerto: Uso de Simulación y Metaheurísticos., 2005. 310 P. Servicio de Publicaciones de la Universidad de Burgos.

El problema tratado en este trabajo consiste en la planificación de los turnos del personal de un aeropuerto encargado de los controles de seguridad y de los mostradores de facturación. Más concretamente se trata de racionalizar los costes de este personal, a la vez que se garantiza un grado de fluidez mínima en cada uno de estos puntos. Se desarrolla un sistema para resolver el problema que consta de tres herramientas principales: un simulador, un método para optimizar simulaciones y un método para resolver el problema de programación de horarios. Este sistema proporciona diferentes soluciones u opciones al decisor y es susceptible de ser adaptado fácilmente a otros modelos dinámicos de colas: supermercados, bancos, centralitas de teléfonos, job-shop, etc.

MÉTODOS ESTADÍSTICOS PARA MEDIR, DESCRIBIR Y CONTROLAR LA VARIABILIDAD

Alberto Luceño (2004). Métodos Estadísticos para Medir, Describir y Controlar la Variabilidad. Colección de Textos Universitarios de la UC; número 1 del grupo de Ciencias Experimentales.

EXTREME VALUE AND RELATED MODELS WITH APPLICATIONS IN ENGINEERING AND SCIENCE

E. Castillo, A. S. Hadi, N. Balakrishnan and J. M. Sarabia. Extreme Value and Related Models with Applications in Engineering and Science. Wiley (<http://personales.unican.es/castie/extremes/>)

En <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-047167172X.html> hay disponible un paquete para trabajar con problemas de extremos en Mathematica.

FROM VECTORS TO TENSORS

Juan Ramón Ruiz-Tolosa and Enrique Castillo. Springer. Universitext. 2005 (<http://personales.unican.es/castie/Tensors/>).

En <http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-115-22-35351693-0,00.html> hay

disponible un paquete para trabajar con problemas de tensores en Mathematica. Trata sobre tensores, con alguna aplicación a la probabilidad. Hay una versión española de este libro publicada por la Academia de Ingeniería.

FUNCTIONAL EQUATIONS IN APPLIED SCIENCES

Functional Equations in Applied Sciences, by E. Castillo, A. Iglesias and R. Ruiz-Cobo.. Elsevier, 2005 (http://www.elsevier.com/wps/find/bookdescription.n.cws_home/704290/description#description)

El capítulo 13 lleva por título: Applications to Probability and Statistics.

REDES PROBABILÍSTICAS Y NEURONALES EN LAS CIENCIAS ATMOSFÉRICAS

J.M. Gutiérrez, R. Cano, A.S. Cofiño and C. Sordo. Monografías del Instituto Nacional de Meteorología (<http://grupos.unican.es/ai/meteo> o <http://grupos.unican.es/ai/meteo/Book.html>)

Este libro recopila temas relativos a la aplicación de técnicas estadísticas y de minería de datos / aprendizaje automático en distintos problemas meteorológicos, especialmente en la predicción local (downscaling).

El libro viene acompañado de un software para MatLab que ilustra los distintos algoritmos y métodos descriptos (<http://grupos.unican.es/ai/meteo/MeteoLab.html>) y que incorpora comandos para importar fácilmente observaciones y predicciones (GRIB), cálculo de EOF, CCA, regresiones, generadores de tiempo, aplicación de redes neuronales, visualización, etc. Este software se distribuye gratuitamente bajo licencia GNU.

Por otra parte, tomando como núcleo el software MeteoLab, se han desarrollado tecnologías específicas para la aplicación operativa de las técnicas más eficientes de downscaling, su ejecución en tiempo real, su almacenamiento en bases de datos, y el acceso interactivo vía Web. Se ha realizado una adaptación a medida para el Instituto Nacional de Meteorología (<http://meteo.macc.unican.es/prometeo>), que ofrece interactivamente predicciones locales diarias en más de 3000 localidades españolas de precipitación, temperaturas, viento, etc.

ESTADÍSTICA PRÁCTICA CON MINITAB

P. Grima, Ll. Marco, X. Tort-Martorell. Prentice Hall, 2004.

Dirigido a estudiantes y profesionales que deseen usar MINITAB como herramienta para la realización de análisis estadísticos. La información se presenta de forma muy visual y todos los procedimientos se explican con ejemplos. En los capítulos de casos prácticos, casi siempre basados en situaciones reales, el protagonismo lo tienen los problemas planteados, de forma que MINITAB es sólo el instrumento que ayuda a resolverlos

DISEÑO DE EXPERIMENTOS: SOLUCIONES CON SAS Y SPSS

L. Vicente, P. Girón, C. Nieto, T. Pérez (2005). "Diseño de Experimentos: Soluciones con SAS y SPSS". Ed. Pearson. Prentice-Hall.

Este libro se propone como referencia para quienes se plantean el Diseño de Experimentos desde un punto de vista práctico. Se trata de un material que permite, aclarar conceptos, identificar los modelos, aplicar los métodos estadísticos, manejar los paquetes SAS y SPSS e interpretar los resultados.

La novedad es:

La presentación de una serie de problemas, explicados detalladamente con el apoyo de los programas estadísticos, mostrando cómo se plantean, cómo se resuelven y cómo deben interpretarse los resultados. La experiencia docente de los autores permite estimar el tiempo de resolución y el grado de dificultad de cada problema.

Una colección de problemas propuestos.

Prácticas basadas en situaciones reales

Una guía de utilización de los programas estadísticos.

PUBLICACIONES EDITADAS POR EL INE EN ENERO Y FEBRERO DE 2005

Censo del Conde de Aranda. Tomo VII. Diócesis de Orense, Oviedo, Palencia y Pamplona

896 páginas. 50 € IVA incluido

Plan Estadístico Nacional 2005-2008

318 páginas. 17 € IVA incluido

Encuesta de Población Activa (EPA).

Principales resultados. Tercer trimestre 2004

122 páginas. 9,50 € IVA incluido

Encuesta del gasto de las empresas en protección ambiental 2002

334 páginas. 15,50 € IVA incluido

Estadísticas del Agua 2002

120 páginas. 8 € IVA incluido

Península Ibérica en Cifras 2004

Versión bilingüe en español y portugués. 36 páginas. 2,20 € IVA incluido

Encuesta Industrial de Empresas 2003

294 páginas. 15,50 € IVA incluido

Encuesta de Empleo del Tiempo 2002-2003. Tomo I. Metodología y Resultados Nacionales

378 páginas. 21,50 € IVA incluido

Tomo II. Resultados por Comunidades Autónomas

316 páginas. 24 € IVA incluido

España en la UE de los 25

Disponible versión en español y versión en inglés.
64 páginas. 2,20 € IVA incluido

Boletín Mensual de Estadística. Número 156. Diciembre 2004.

348 páginas, incluye CD-Rom.

Precio del ejemplar: 17 €. IVA incluido

Suscripción anual: 135,50 € IVA incluido

INEbase. Enero 2005

CD-Rom. Precio del ejemplar: 17,28 € IVA incluido

Suscripción anual: 151,14 € IVA incluido

Contenido:
Boletín Mensual de Estadística. Número 157 -
Enero 2005

Cifras de población referidas al 01/01/2004. Real Decreto 2348/2004, de 23 de diciembre
Estadística de Hipotecas 2003. Base antigua

INEbase. Diciembre 2004

CD-Rom. Precio del ejemplar: 17,28 € IVA incluido

Suscripción anual: 151,14 € IVA incluido

Contenido:
Boletín Mensual de Estadística. Número 156 -
Diciembre 2004

Encuesta continua de presupuestos familiares
2003 trimestre 1º

Encuesta continua de presupuestos familiares
2003 trimestre 2º

Encuesta de morbilidad hospitalaria 2002

Defunciones según la causa de muerte 2002

Cuenta satélite del turismo de España. Serie contable 1995-2003

Cuenta de renta de los hogares. Base 1995. Serie 1995-2002

ETCL. Serie 1º trimestre 2000- 3º trimestre 2004

Encuesta Industrial de Empresas. Serie 1993-2003

España en la UE de los 25

Obra divulgativa compuesta de una serie de gráficos, mapas y comentarios sobre temas tan variados como población, economía, mercado laboral, sociedad, y energía y medio ambiente. El contenido de esta publicación se basa en la selección de 25 indicadores básicos, publicados por la Oficina de Estadística de las Comunidades Europeas (Eurostat), que en su mayoría pertenecen al grupo de Indicadores Estructurales. Entre los resultados obtenidos en el estudio destaca que: España es la quinta economía de la UE por Producto Interior Bruto (PIB) y su nivel de precios está 14 puntos por debajo del nivel medio comunitario; además las españolas son las más longevas, la tasa bruta de natalidad es idéntica a la media de la UE-25 y el empleo femenino español registra el mayor crecimiento de todo el entorno europeo.

64 páginas. Precio: 2,20 € IVA incluido

Península Ibérica en Cifras 2004

En el marco de las Jornadas Ibéricas surgió la idea de elaborar esta publicación con el objetivo de presentar, de forma divulgativa, un conjunto de

indicadores que cubran los grandes temas que caracterizan la vida de nuestras sociedades. Esto permite, por una parte, establecer comparaciones entre Portugal y España; y por otra, mostrar la posición que cada uno de estos dos países ocupa en el contexto de la Unión Europea.

Edición bilingüe portugués-español.

36 páginas. 2,20 € IVA incluido

Encuesta de Empleo del Tiempo 2002-2003.

Se trata de la primera encuesta de ámbito nacional que se ha hecho de manera armonizada con las de otros países europeos siguiendo las recomendaciones de EUROSTAT. El estudio sobre la utilización del tiempo facilita información sobre el porcentaje de personas que realizan una determinada actividad en el transcurso del día y la duración media diaria dedicada a esa actividad por dichas personas. Esta información primaria permitirá, entre otras cosas, analizar la dimensión del trabajo no remunerado realizado por los hogares, la distribución de las responsabilidades familiares en el hogar, la participación de la población en actividades culturales y de ocio, el empleo del tiempo de grupos sociales especiales (jóvenes, desempleados, ancianos,...) y estimar la cuenta satélite de producción de los hogares.

Tomo I. Metodología y Resultados Nacionales.

378 páginas. 21,50 € IVA incluido

Tomo II. Resultados por Comunidades

Autónomas. 316 páginas. 24 € IVA incluido

CURSOS Y SEMINARIOS

III CURSO DE EXPERTO UNIVERSITARIO EN ESTADÍSTICA PÚBLICA.

Del 21/10/2004 al 15/06/2005 (320 horas - 32 Créditos) se imparte el III Curso de Experto Universitario en Estadística Pública, organizado por la Universidad de Sevilla, el Instituto de Estadística de Andalucía y el Centro Andaluz de Prospectiva. Más información en la dirección: <http://www.cica.es/~canp/iiicursoexpertoep.htm>

CURSO EXTRAORDINARIO (UNIVERSIDAD DE SALAMANCA): "TRATAMIENTO ESTADÍSTICO DE DATOS CON EL SPSS 12"

Más información: <http://www3.usal.es/~eioq/spss>
Quintín Martín Martín (e-mail: qmm@usal.es)

(Director del curso)

Objetivos:

- Poner al alcance de los usuarios de análisis estadísticos la potencialidad que ofrecen los programas de ordenador (Paquetes Estadísticos), en concreto del SPSS 12.
- Que el usuario de Paquetes Estadísticos sea capaz de discernir qué técnica estadística es la más apropiada para extraer la información subyacente en los datos, instruirle en la ejecución de esas técnicas estadísticas y, finalmente, que interprete los resultados proporcionados por ellas.

"ELEMENTOS DE ESTEREOLOGÍA Y GEOMETRÍA ESTOCÁSTICA" (UNIVERSIDAD DE CANTABRIA).

Luis Manuel Cruz Orive.

Curso de carácter metodológico-aplicado, para alumnos de Estadística o Matemática aplicadas, Biomedicina, Ciencias de Materiales, etc. Las técnicas están basadas en la probabilidad geométrica, la geometría integral, y la estadística, pero dichas disciplinas no son un prerrequisito para seguir el curso. Se impartirá en horario de 11:45 a 13:45 miércoles y viernes, a partir del 2 de marzo.

SEMINARIOS DEPARTAMENTO DE ESTADÍSTICA ABRIL-MAYO 2005 (UNIVERSIDAD CARLOS III DE MADRID)

Abril 2005

Viernes 1, Elena FERNANDEZ (Universitat Politècnica de Catalunya). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

Viernes 8, Alfonso NOVALES (Universidad Complutense de Madrid). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

Viernes 15, Lola UGARTE (Universidad Pública de Navarra). Hora: 13:15h, lugar: Sala 4.0.F16, Leganés.

Viernes 22, Agustín GARCIA NOGALES (Universidad de Extremadura). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

Viernes 29, Ramón DIAZ-URIARTE (Centro Nacional de Investigaciones Oncológicas). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

Mayo 2005

Viernes 6, Carlos ROMERO (Universidad Politécnica de Madrid). Hora: 13:15h, lugar: Sala 4.0.F16, Leganés.

Viernes 13, Davy PAINDAVEINE (Université Libre de Bruxelles). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

Viernes 20, Kenneth WALLIS (University of Warwick). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

Lunes 23, Nadine HILGERT (INRA, Montpellier). Hora: 13:15h, lugar: Sala Costas

Goutis (10.0.23), Getafe.

Viernes 27, Afonso GORDALIZA (Universidad de Valladolid). Hora: 13:15h, lugar: Sala Costas Goutis (10.0.23), Getafe.

CURSO DE ESPECIALIZACIÓN: BIOESTADÍSTICA, PRINCIPIOS PARA INTERPRETAR UN ESTUDIO CIENTÍFICO.

Dirigido a profesionales del campo de la salud que tengan que interpretar material científico.

Curso Semipresencial del Departamento de Estadística e Investigación Operativa (UPC) en colaboración con la Facultat de Matemàtiques i Estadística (FME, UPC) y la Facultat d'Informàtica de Barcelona (FIB, UPC).

Patrocinado por NOVARTIS (Instituto de Formación) y MEDICINA CLÍNICA

Inicio 5 de mayo de 2005. Duración 2 meses.

Lugar: Barcelona

Más información:

<http://www-fme.upc.es/bioestadistica>

SUMMER SCHOOL STATISTICAL MODELS IN FINANCIAL SERIES

8ª Summer School de la Sección Europea de la "Internacional Association for Statistical Computing" (IASC)

Organizada por el Departamento de Estadística e Investigación Operativa (UPC) la conjuntamente con el Departamento de Economía de la Empresa (UAB)

Objetivo:

Presentar temas avanzados de los modelos estadísticos aplicados a las series temporales financieras, insistiendo en los aspectos metodológicos y computacionales y en las aplicaciones.

La escuela está prevista para jóvenes investigadores, también está abierta a profesores recién incorporados e investigadores trabajando en el sector público o privado.

Profesores Invitados: Agustín Maravall (Banco de España), Daniel Peña (Universidad Carlos III, Madrid), Esther Ruiz (Universidad Carlos III, Madrid), Ruey Tsay (University of Chicago) y Cesar Villazón (Universitat Autònoma de Barcelona, Barcelona).

Fecha: del 17 al 22 de julio de 2005.

Lugar: Barcelona

Más información:

<http://www-eio.upc.es/smfs2005>

PLS'05 4TH INTERNATIONAL SYMPOSIUM ON PLS AND RELATED METHODS

Organizado por el Departamento de Estadística e Investigación Operativa (UPC) en colaboración con la Facultat d'Informàtica de Barcelona (FIB, UPC) y la escuela Hautes Etudes Commerciales (HEC, Paris)

Las técnicas PLS (Partial Least Squares), primero introducidas en 1966 por Herman Wold, han adquirido una sólida posición para solucionar los problemas aplicados en una variedad de campos como el marketing, aplicaciones industriales, ciencias de la información, etc. Siguiendo la tradición de los últimos simposios que se concentraban básicamente en un campo, el cuarto simposio internacional pondrá el énfasis en los problemas del marketing y sus aplicaciones.

Fecha: del 7 al 9 de setiembre de 2005

Lugar: Barcelona

Más información: <http://www.pls05.upc.edu>

CURSO SOBRE SEIS SIGMA EN BARCELONA

Organizado por el Departamento de Estadística e

Investigación Operativa (UPC)

La estructura y los contenidos son similares a los que realizan las grandes empresas para formar a sus "black belts". En ediciones anteriores (ya se han realizado 3 en Barcelona y otras 3 en San Sebastián) se han formado profesionales de más de 50 empresas entre las que se encuentran Visteon, Sony, Philips, Delphi Diesel System, Metaldyne, Ideal Standard, Alcoa, 3M, Banc Sabadell...

Fecha: Inicio 18 de marzo de 2005

Lugar: Barcelona

Más información: <http://www.upcq.net/index.php>

CURSO ESTADÍSTICA APLICADA A LA CALIDAD Y EL I+D

Organizado por el Departamento de Estadística e Investigación Operativa (UPC)

Nuevo programa estructurado en torno a 3 bloques (Conceptos clave, Control Estadístico y Diseño de Experimentos) de 30 horas cada uno y orientado a mostrar los conceptos y las técnicas más relevantes para la gestión de la calidad y el I+D.

Fecha: Inicio 1 de abril de 2005

Lugar: Barcelona

Más información: <http://www.upcq.net/index.php>

CONGRESOS

2005

ABRIL

- 4-5 IASE SATELLITE MEETING ON " STATISTICS EDUCATION AND THE COMMUNICATION OF STATISTICS" to be held in Sydney, Australia. Inf: bphillips@swin.edu.au; www.swin.edu.au/math/iase/commsat.html
- 5-12 INTERNATIONAL STATISTICAL INSTITUTE, 55th BIENNIAL SESSION; Sydney, Australia; Inf: ISI Permanent Office, Prinses Beatrixlaan 428, P.O. Box 950, 2270 AZ Voorburg, The Netherlands. Tel: +31-70-337-5737; Fax: +31-70-386-0025. E-mail: isi@cbs.nl
- 12-13 THE INTEGRATED SOCIETY FOR HUMAN RIGHTS ENVIROMENTS AND DEVELOPMENTS (INSHURED) ORGANIZES A CONFERENCE ON "POLITICAL RIGHTS VS. HUMAN RIGHTS" in Lumbini, Bhairahawa, Nepal. Inf: integratednepal@hotmail.com
- 13-14 IASS SATELLITE CONFERENCE AFTER THE ISI SESSION:"COMPLEX SAMPLING, RETROSPECTIVE SAMPLING AND MISSING DATA. A CONFERENCE IN HONOUR OF ALASTAIR J. SCOTT", to be held in Auckland, New Zealand. Inf:c.wild@auckland.ac.nz
- 13-16 FOURTH INTERNATIONAL CONFERENCE ON STATISTICS IN BUSSINESS AND INDUSTRY (ISBIS-4). It will be held in the vicinity of Cairns, on the Great Barrier Reef, an area of remarkable natural beauty and one of Australia's best vacation areas. Inf:nif@valuemetrics.com.au, [http:// www.action-m.com/isbis4](http://www.action-m.com/isbis4).
- 14-15 IAOS SATELLITE MEETING TO THE ISI SESSION ON " CHALLENGES IN MEASURING SMALL AND INDIGENOUS POPULATIONS", to be held at Te Papa Tongarewa, Wellington, New Zealand. Inf: <http://www.stats.govt.nz/ISIsatellitemeeting>, <http://www.stats.govt.nz/>, ISIsatellite@stats.govt.nz
- *21-23 FIFTH SIAM INTERNATIONAL CONFERENCE ON DATA MINING, to be held in Newport Beach, CA,USA. Inf:www.siam.org/meetings/sdm05, agoodman@uci.edu

MAYO

- *3-5 2005 SALFORD SYSTEMS: DATA MINING TRAINING AND WORKSHOP, Madrid. Inf: <http://www.cartdatamining.com/workshops.htm>
- 8-11 CANADIAN TRANSPORTATION RESEARCH FORUM 40TH ANNUAL CONFERENCE, Hamilton Ontario, Canada. Inf: <http://www.ctrf.ca/>
- 15-18 SIAM 2005 CONFERENCE ON OPTIMIZATION, City Conference Centre, Stockholm, Sweden. Inf: <http://www.siam.org/meetings/op05/>
- *19-21 UNITED STATES CONFERENCE ON TEACHING STATISTICS (USCOTS), to be held at the Ohio State University, Columbus, Ohio, USA. Inf: www.causeweb.org/uscotts/, rumsey@stat.ohio-state.edu
- *23-26 THE UNIVERSITY OF SIENA, ITALY, WILL HOST AN INTERNATIONAL CONFERENCE TO COMMEMORATE C. GINI AND M. O. LORENZ CENTENARY SCIENTIFIC RESEARCH. Inf: ginilorenz05@unisi.it
- 25-27 QUATRIÈME COLLOQUE FRANCOPHONE SUR LES SONDAGES, Québec (Canada), to be held at the Université Laval. Info: <http://www.crm.umontreal.ca/sondages2005/>, sondages2005@mat.ulaval.ca
- *25-27 X CONFERENCIA ESPAÑOLA DE BIOMETRÍA, Facultad de Psicología de la Universidad de Oviedo. Inf: <http://www.cop.es/biometria>

JUNIO

- 1-3 SPRING RESEARCH CONFERENCE (2005 SCR), Park City, Utah, USA. Inf: src2005.byu.edu, reese@stat.byu.edu
- *2-8 ISOLDE X (INTERNATIONAL SYMPOSIUM ON LOCATIONAL DECISIONS), Sevilla-Islantilla, SPAIN. Inf: <http://www.us.es/isolde>
- *5-9 ABS05 –2005 APPLIED BAYESIAN STATISTICS SCHOOL, BAYESIAN APPROACHES TO EVIDENCE SYNTHESIS AND DECISION MODELLING IN HEALTH CARE, Villa Monastero, Varenna (Lake Como), Italy. Inf: www.mi.imati.cnr.it/conferences/abs05.html
- *6-10 CIMMA 2005 will be held in Almería (south of Spain, 170 km or about 100 miles from Granada). Inf: <http://www.ual.es/Congresos/CIMMA2005/>
- 12-15 ANNUAL MEETING OF THE STATISTICAL SOCIETY OF CANADA, Saskatoon, Saskatchewan. Inf: bickis@math.usask.ca
- 12-16 2005 INTERNATIONAL SYMPOSIUM ON FORECASTING, San Antonio, Texas, USA. Inf: tyokum@angelo.edu
- 26-1 30TH CONFERENCE ON STOCHASTIC PROCESSES AND THEIR APPLICATIONS, on held on campus of the University of California at Santa Barbara, CA, USA. Inf: <http://www.pstat.ucsb.edu/projects/spa05>, feldman@pstat.ucsb.edu
- *27-30 "THE 2005 INTERNATIONAL MULTICONFERENCE IN COMPUTER SCIENCE AND COMPUTER ENGINEERING" (composed of 16 Joint Conferences), Las Vegas, USA. Inf: <http://www.world-academy-of-sciences.org>
- 28-29 THE THIRD EUROPEAN CONFERENCE ON INTELLIGENT MANAGEMENT SYSTEMS IN OPERATIONS, University of Salford, UK. Inf: <http://www.orsoc.org.uk/conf/imsio2005/main.htm>
- *29-1 17TH MINI EURO CONFERENCE ON "CONTINUOUS OPTIMIZATION IN INDUSTRY", Pecs, Hungary. Inf: <http://www.cs.elte.hu/opres/Pecs05EUROmini/>

JULIO

- *6-8 PROGIC2005: SECOND WORKSHOP ON COMBINING PROBABILITY AND LOGIC SPECIAL FOCUS: OBJECTIVE BAYESIANISM, Centre for Philosophy of Natural and Social Science, London School of Economics. Inf: <http://personal.lse.ac.uk/willia11/progic2005/>
- *10-15 20TH IWSM, INTERNATIONAL WORKSHOP ON STATISTICAL MODELLING, STATISTICAL SOLUTIONS TO MODERN PROBLEMS, Sydney, Australia. Inf: <http://www.uws.edu.au/iwsm2005/>
- 11-15 IFORS 2005: 17TH TRIENNIAL CONFERENCE OF THE INTERNATIONAL FEDERATION OF OPERATIONAL RESEARCH SOCIETIES, Honolulu, Hawaii. Inf: <http://www.informs.org/Conf/IFORS2005/>
- 11-15 20TH BRITISH COMBINATORIAL CONFERENCE, Durham, UK. Inf: <http://mcs.open.ac.uk/bcc2005/>
- *13-15 INTERNATIONAL SEMINAR ON NONPARAMETRIC INFERENCE (ISNI 2005) that will take place in A Coruña. Inf: <http://www.orzancongres.com/eventos/2005/isni/>

*17-22 THE 8TH ERS-IASC INTERNATIONAL SUMMER SCHOOL ON STATISTICAL MODELS IN FINANCIAL SERIES to be held in Barcelona (Spain). Inf: <http://www-eio.upc.es/smfs2005>

*20-23 ISIPTA '05, 4TH INTERNATIONAL SYMPOSIUM ON IMPRECISE PROBABILITIES AND THEIR APPLICATIONS SECOND CALL FOR PAPERS, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. Inf: <http://www.sipta.org/isipta05>

31-4 THE 18TH INTERNATIONAL CONFERENCE ON PRODUCTION RESEARCH, Fisciano (SA), Italy. Inf: <http://icpr18.unisa.it/default.htm>

AGOSTO

7-11 2005 JOINT STATISTICAL MEETINGS, Minneapolis, Minnesota, USA. To be held at the Minneapolis Convention Center. Sponsored by ASA, ENAR, WNAR, IMS and SSC. Inf: e-mail: meetings@amstat.org

*7-12 MAXENT 2005, 25TH INTERNATIONAL WORKSHOP ON BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING, San Jose State University, San Jose CA USA. Inf: <http://maxent2005.org>

*22-24 X SEMINARIO DE ESTADISTICA APLICADA DEL IASI, SOBRE " ESTADISTICA COMPUTACIONAL". Se realizará en la Pontificia Universidad Católica del Perú, en la ciudad de Lima, Perú. Inf: fabpan@cwpanama.net, lgasco@pucp.edu.pe

*22-25 14TH EUROPEAN YOUNG STATISTICIANS MEETING (14TH. EYSM), Debrecen (Hungria). Inf: luisj@uniovi.es.

SEPTIEMBRE

*6-10 OPERATIONAL RESEARCH PERIPATETIC POST-GRADUATE PROGRAMME (ORP3) 2005, Valencia, Spain. <http://www.orp3.com>

7-9 2005 ANNUAL INTERNATIONAL SCIENTIFIC CONFERENCE OF THE GERMAN OPERATIONS RESEARCH SOCIETY (GOR) , Bremen, Germany. Inf: <http://www.logistik.uni-bremen.de>

*8-9 IX CONGRESO DE INGENIERÍA DE ORGANIZACIÓN que se celebrará en Gijón. Inf: <http://cio2005.uniovi.es>

*12-16 CURSO ECAS "REGRESSION QUANTILES AND APPLICATIONS", this course will focus on Regression Quantiles and Applications, Belgium. Inf: <http://www.ulb.ac.be/soco/lmtd/ecas2005>

*14-16 IX CONGRESO DE METODOLOGÍA DE LAS CIENCIAS SOCIALES Y DE LA SALUD, Universidad de Granada. Inf: <http://www.ugr.es/~cmetodo/>

*16-17 EIGHTH WORKSHOP ON CASE STUDIES IN BAYESIAN STATISTICS, Carnegie Mellon University, Pittsburgh, PA. Inf: <http://www.stat.cmu.edu/bayesworkshop/2005/goodstudy.html>.

*18-22 2005 INTERNATIONAL WORKSHOP ON GLOBAL OPTIMIZATION GO05, Almeria, Spain. Inf: <http://dali.ace.ual.es/~go05/>

*21-23 SEMINARIO SOBRE "SISTEMAS ESTADÍSTICOS NACIONALES EN EL CONTEXTO REGIONAL DEL MERCOSUR". Organizado conjuntamente por el IASI y el Instituto Nacional de Estadística (INE) de Uruguay. Inf: fabpan@cwpanama.net, oandina@ine.gub.uy

OCTUBRE

*5-8 WORKSHOP ON STATISTICS IN GENOMICS AND PROTEOMICS, Hotel Estoril Eden, Monte Estoril, Portugal. Inf: <http://wsgp.deio.fc.ul.pt>

*19-21 THE 2ND COMPOSITIONAL DATA ANALYSIS WORKSHOP (CODAWORK'05). Universitat de Girona Inf: <http://ima.udg.es/Activitats/CoDaWork05/>

*28-31 3RD WORLD CONFERENCE ON COMPUTATIONAL STATISTICS & DATA ANALYSIS Amathus Beach Hotel, Limassol, Cyprus. Inf: <http://www.csdassn.org/europe/csda2005/>

NOVIEMBRE

13-16 INFORMS ANNUAL MEETING 2005, New Orleans, USA.

*23-25 18TH MINI EURO CONFERENCE ON VNS, Tenerife, Spain. Inf: <http://www.mecvns.com>

2006

MARZO

*20-24 CONFERENCE ON STOCHASTICS IN SCIENCE. In Honor of Ole E. Barndorff-Nielsen's 71st birthday. CIMAT, Guanajuato, Mexico. Inf: www.cimat.mx/Eventos/oebn-conference, pabreu@ciamat.mx

MAYO

*17-19 12TH IFAC SYMPOSIUM ON INFORMATION CONTROL PROBLEMS IN MANUFACTURING INCOM'2006, 2006 à Saint Etienne France. Inf: www.emse.fr/incom06/call_for_papers.html

28-31 SSC-2006: ANNUAL MEETING OF THE STATISTICAL SOCIETY OF CANADA, London, Ontario, Canada. Inf: bellhouse@stats.uwo.ca

JUNIO

*5-9 PROBABSTAT 2006 FIFTH INTERNATIONAL CONFERENCE ON PROBABILITY and STATISTICS, Smolenice Castle, Slovakia. Inf: <http://aiolos.um.savba.sk/~viktor/probabstat.html>

*8-11 XIX REUNIÓN ANUAL DE LA SOCIEDAD DE ECONOMÍA APLICADA ASEPELT-ESPAÑA. Badajoz. Inf: <http://asepelt2005.eweb.unex.es/>

JULIO

2-5 EURO XXI CONFERENCE 2006, Reykjavik, Iceland.

2-7 ICOTS-7: SEVENTH INTERNATIONAL CONFERENCE ON TEACHING STATISTICS, to be held in Salvador, Bahía, Brazil. Inf: <http://www.maths.aotago.ac.bz/icots7>, batanero@ugr.es

*16-21 XXIIIRD INTERNATIONAL BIOMETRIC CONFERENCE, to be held in Montreal, Quebec, Canada. Inf: <http://www.tibs.org>

AGOSTO

*6-10 2006 JOINT STATISTICAL MEETINGS, to be held in Seattle, Washington, USA. Inf: www.amstat.org/meetings/index.cfm?fuseaction=main, jsm@amstat.org

22-30 INTERNATIONAL CONGRESS OF MATHEMATICIANS MADRID 2006 (ICM-2006). Inf: <http://www.icm2006.org>

*28-1 17TH COMPSTAT SYMPOSIUM OF THE INTERNATIONAL ASSOCIATION FOR STATISTICAL COMPUTING (COMPSTAT 2006), Rome. Inf: <http://w3.uniroma1.it/compstat2006/>

2007

JULIO

*29-2 2007 JOINT STATISTICAL MEETINGS, to be held in Salk Lake City, Utah, USA. Inf: www.amstat.org/meetings/index.cfm?fuseaction=main, jsm@amstat.org

AGOSTO

22-29 INTERNATIONAL STATISTICAL INSTITUTE (ISI) to be held in Lisboa, PortugalS MADRID 2006 (ICM-2006). Inf: isi@cbs.nl <- ¿Es correcto todo esto?

2008

AGOSTO

*3-7 2008 JOINT STATISTICAL MEETINGS, to be held in Denver, Colorado, USA. Inf: www.amstat.org/meetings/index.cfm?fuseaction=main, jsm@amstat.org

2009

AGOSTO

*16-22 INTERNATIONAL STATISTICAL INSTITUTE (ISI) 57th BIENNIAL SESSION, to be held in Durban, South Africa. Inf: isi@cbs.nl

ALTAS Y BAJAS DE SOCIOS

ALTAS MES DE ENERO

Jorge Navarro Camacho
Saray Ugarte Bailbao
Alfonso Gordaliza Ramos
José Ramón Berrendero Díaz
Carlos Ramón García Alonso
Luis Ángel García Escudero
Fernando Reche Lorite
Maria Araceli Tuero Díaz

ALTAS MES DE FEBRERO

Vicente Fernández Rodríguez
Maria Del Pilar Moreno Navarro
Itziar Irigoien Garbizu
Maria Asunción Lubiano Gómez

Cesar Beltrán-Royo

ALTAS MES DE MARZO

Alberto Lekuona Amiano

BAJAS MES DE ENERO

Pedro Ruiz Bevia
Jorge García Robles
Yanira Del Rosario De Paz Santana
M^a Del Mar Soldevilla Moreno
Carmen Lucía Vidal Rodeiro

BAJAS MES DE FEBRERO

Yolanda García García
Nieves Atienza Martínez

2. ESTUDIOS MONOGRÁFICOS Y OPINIONES SOBRE LA PROFESIÓN

HISTORIA DEL PREMIO RAMIRO MELENDRERAS

Domingo Morales González

Ramiro Melendreras Gimeno fue catedrático de Estadística e Investigación Operativa de la Universidad de Murcia, en la cual, así como anteriormente en las Universidades Complutense de Madrid, de Santiago de Compostela y de Granada, ejerció su actividad docente. Su labor investigadora se centró principalmente en el campo de la programación matemática, aunque también trabajo en problemas de selección de la cartera, decisión multicriterio, teoría de la información o teoría de juegos.

Ramiro fallece el 17 de de julio de 1983 a la edad de 38 años. Un año después, en 1984, la Sociedad Española de Investigación Operativa, Estadística e Informática (denominada Sociedad de Estadística e Investigación Operativa (SEIO) a partir del 20 de diciembre de 1984) junto con la Fundación Ramiro Melendreras crean el premio “Ramiro Melendreras” al mejor trabajo presentado por un joven investigador en el congreso de la SEIO. Los aspirantes, menores de 30 años, deben presentar al concurso un trabajo inédito que es juzgado por el Comité de Programa del Congreso.

La andadura del premio comienza en el congreso de Granada de 1984, siendo

Wenceslao González Mantenga el primer premiado y Alberto Olivares González el último. Durante este periodo ha habido 15 convocatorias del premio, y únicamente ha quedado desierto en la edición del congreso de Calella (Barcelona) en 1994. Los primeros trabajos premiados se publicaron en las revistas de la SEIO, rompiéndose esa pauta con la transformación de Trabajos de Estadística y Trabajos de Investigación Operativa en Test y Top, respectivamente. Cabe destacar que los cinco primeros ganadores (ediciones de 1984, 1985, 1986, 1988 y 1989) son actualmente catedráticos de universidad y que los seis siguientes (ediciones de 1991, 1992, 1995, 1997, 1998 y 2000) son profesores titulares de universidad. En definitiva, el premio ha mantenido un alto nivel de calidad, como muestra la buena trayectoria profesional posterior de los investigadores galardonados y de sus artículos premiados.

Lamentablemente el premio no ha tenido una distribución muy equitativa por sexos dado que, de entre los 14 premios otorgados, solamente en las ediciones de 1997 y 2001 éste fue ganado por mujeres. Es curioso observar además que se han dado 7 premios a trabajos de Estadística y otros 7 a trabajos de Investigación Operativa,

produciéndose una alternancia sistemática entre los años 1988 y 2000. Profundizando un poco más en los temas de investigación objeto de premio, se puede establecer la siguiente clasificación: Inferencia no Paramétrica (1984, 1989, 1992), Teoría de la Información Estadística (1985, 1997, 2001), Optimización Multiobjetivo (1986), Teoría de Juegos (1988, 1991, 1998), Análisis de Supervivencia (2000), Programación Lineal Entera (1995), Programación Semi-Infinita (2003), Programación no Lineal (2004).

Las universidades de procedencia de los premiados han sido: Santiago de Compostela (1984, 1988, 1989, 1991), Complutense de Madrid (1985, 1986, 1997), A Coruña (1992), La Laguna (1995), Miguel Hernández de Elche (1998, 2001, 2003), Vigo (2000), Rey Juan Carlos (2004).

A continuación se presenta una breve descripción de los premios “Ramiro Melendreras” concedidos, hasta el momento, en los Congresos Nacionales de Estadística e Investigación Operativa.

Premio Ramiro Melendreras 1984

En el Congreso XIV, celebrado en Granada, fue premiado *Wenceslao González Manteiga* de la Universidad de Santiago de Compostela, por el trabajo titulado “Obtención del sesgo, varianza y error cuadrático medio de una familia axiomática de estimadores no paramétricos para funciones de densidad y regresión“. El trabajo se publicó en las actas del congreso, tomo I, páginas 163-179. Actualmente Wenceslao González Mantenga es Catedrático de Universidad (área de Estadística e Investigación Operativa) en la Universidad de Santiago de Compostela.

En el trabajo se construye una familia axiomática de estimadores no paramétricos para la función de densidad y de regresión, del tipo “delta secuencia”, que incluye a numerosos métodos de estimación (tipo núcleo, tipo histograma, etc.). Se obtiene una expresión general para el error cuadrático medio y se ilustran los resultados obtenidos en casos particulares y ejemplos numéricos.

Premio Ramiro Melendreras 1985

En el congreso XV, celebrado en Gijón (Asturias), fue premiado *Domingo Morales González* de la Universidad Complutense de Madrid, por el trabajo titulado “La medida de divergencia de Kagan en el muestreo secuencial

con procesos de Dirichlet“. El trabajo se publicó en Trabajos de Estadística, Vol. 1, N. 1, 88-96, 1986. Actualmente Domingo Morales González es Catedrático de Universidad (área de Estadística e Investigación Operativa) en la Universidad Miguel Hernández de Elche.

En el trabajo se extiende la medida de divergencia de Kagan para establecer una medida de la información que una muestra aleatoria da acerca de un proceso de Dirichlet en su globalidad. Después de analizar algunas de sus propiedades, se obtiene la expresión que toma dicha medida entre etapas n y $n+1$ del muestreo, y se estudian sus propiedades bayesianas. Finalmente, se introduce una regla de parada definida a partir de la información que se obtiene en el muestreo al pasar de una etapa a la siguiente.

Premio Ramiro Melendreras 1986

En el congreso XVI, celebrado en Torremolinos (Málaga), fue premiado *David Ríos Insua* de la Universidad Complutense de Madrid, por el trabajo titulado “Sobre soluciones óptimas en problemas de optimización multiobjetivo“. El trabajo se publicó en Trabajos de Investigación Operativa, Vol. 2, 49-67, 1987. Actualmente David Ríos Insua es Catedrático de Universidad (área de Estadística e Investigación Operativa) en la Universidad Rey Juan Carlos y Académico Numerario (electo) de la Real Academia de Ciencias.

En el trabajo se proporcionan caracterizaciones computacionales de distintos tipos de soluciones en diversos problemas de optimización multiobjetivo.

Premio Ramiro Melendreras 1988

En el congreso XVII, celebrado en Benidorm (Alicante), fue premiado *Ignacio García Jurado* de la Universidad de Santiago de Compostela, por el trabajo titulado “Un refinamiento del concepto de equilibrio propio de Myerson“. El trabajo se publicó en Trabajos de Investigación Operativa, Vol. 4, N. 1, 11-19, 1989. Actualmente Ignacio García Jurado es Catedrático de Universidad (área de Estadística e Investigación Operativa) en la Universidad de Santiago de Compostela.

En el trabajo se revisa el concepto de equilibrio propio de Myerson y se propone un refinamiento estricto de tal concepto. También se demuestra la existencia de una de las nuevas soluciones en la extensión mixta de cualquier juego n -personal finito en forma normal.

Premio Ramiro Melendreras 1989

En el congreso XVIII, celebrado en Santiago de Compostela (A Coruña), fue premiado *Ricardo Cao Abad* de la Universidad de Santiago de Compostela, por el trabajo titulado “Órdenes de convergencia para las aproximaciones normal y bootstrap en la estimación no paramétrica de la función de densidad”. El trabajo se publicó en *Trabajos de Estadística*, Vol. 5, N. 2, 23-32, 1990. Actualmente Ricardo Cao Abad es Catedrático de Universidad (área de Estadística e Investigación Operativa) en la Universidad de A Coruña.

El trabajo se ocupa de las distribuciones usadas para construir intervalos de confianza para la función de densidad en una situación no paramétrica. Se comparan los órdenes de convergencia para el límite normal, su aproximación “plug-in” y el método bootstrap. Se deduce que el bootstrap se comporta mejor que las otras dos aproximaciones, tanto en su forma clásica, como con la aproximación bootstrap normal.

Premio Ramiro Melendreras 1991

En el congreso XIX, celebrado en Segovia, fue premiado *Luciano Méndez Naya* de la Universidad de Santiago de Compostela, por el trabajo titulado “Un nuevo algoritmo para la resolución de juegos bimatriciales”. El trabajo se publicó en *Trabajos de Investigación Operativa*, Vol. 7, N. 1, 87-93, 1992. Actualmente Luciano Méndez Naya es Profesor Titular de Universidad (área de Economía Aplicada) en la Universidad de Santiago de Compostela.

En el trabajo se da un nuevo algoritmo para la resolución de juegos bimatriciales basado en encontrar las respuestas óptimas a las estrategias de cada jugador. El desarrollo del algoritmo se basa en un teorema de convexidad que se demuestra en el artículo.

Premio Ramiro Melendreras 1992

En el congreso XX, celebrado en Cáceres, fue premiado *José Antonio Vilar Fernández* de la Universidad de A Coruña, por el trabajo titulado “Estimación núcleo de la función de regresión con instantes muestrales aleatorios”. El trabajo se publicó en *Test*, Vol. 4, N. 1, 137-178, 1995. Actualmente José Antonio Vilar Fernández es Profesor Titular de Universidad (área de Estadística e Investigación Operativa) en la Universidad de A Coruña.

En el trabajo se investiga el comportamiento de un estimador no paramétrico de la función de regresión $m(x) = E(Y | X=x)$, de un proceso estocástico estacionario $Z(t) = (X(t), Y(t))$, en base a una muestra aleatoria $Z(t(1)), \dots, Z(t(n))$, suponiendo que los instantes de muestreo $t(i)$ están irregularmente espaciados en el tiempo. Bajo condiciones de dependencia mixing sobre $Z(t)$ y para dos estructuras estocásticas muy generales del proceso puntual de tiempos de muestreo $T = \{t(1), t(2), \dots\}$, se obtienen la consistencia en media cuadrática y la normalidad asintótica del estimador núcleo de Nadaraya-Watson y de una versión recursiva de este último. Un análisis comparativo con los resultados clásicos con tiempos regularmente espaciados pone de relieve la notable influencia del proceso T , tanto en tasas de convergencia más lentas como en cotas asintóticas mayores para el error cuadrático integrado medio (MISE).

Premio Ramiro Melendreras 1994

En el congreso XXI, celebrado en Calella (Barcelona) el premio fue declarado desierto.

Premio Ramiro Melendreras 1995

En el congreso XXII, celebrado en Sevilla, fue premiado *Juan José Salazar González* de la Universidad de La Laguna, por el trabajo titulado “Técnicas Poliédricas para Garantizar la Privacidad en Tablas Estadísticas Públicas”. El trabajo se publicó en *Mathematical Programming*, Vol. 84, 283-312, 1999. Actualmente Juan José Salazar González es Profesor Titular de Universidad (área de Estadística e Investigación Operativa) en la Universidad de La Laguna.

El trabajo propone un modelo de Programación Lineal Entera para el problema de determinar las celdas a ocultar en una tabla estadística bidimensional cuando se desea proteger algunas celdas sensibles mediante el llamado “método de supresión de celdas”. Este método es de uso frecuente en institutos de estadística como el INE, pero tradicionalmente se aplicaba de forma manual. Adicionalmente se realiza un análisis poliédrico de la envolvente convexa de las soluciones enteras de dicho problema, generándose así nuevas desigualdades válidas para fortalecer la relajación lineal del modelo. Junto con el Prof. Matteo Fischetti (Universidad de Padua, Italia) se implementa luego un algoritmo de ramificación y corte que dio lugar a la publicación citada en la referencia. Este trabajo base abrió una línea de investigación que

ha generado otras publicaciones del autor con otros investigadores nacionales y extranjeros, desarrolladas fundamentalmente a través de dos proyectos europeos: uno en el IV programa marco (Esprit Project 20462-SDC) y en el V programa marco (IST-2000-25069-CASC).

Premio Ramiro Melendreras 1997

En el congreso XXIII, celebrado en Valencia, fue premiada *María del Carmen Pardo Llorente* de la Universidad de Complutense de Madrid, por el trabajo titulado “Inferencia en poblaciones multinomiales basada en la Rphi-divergencia”. El trabajo se publicó en *Journal of Multivariate Analysis*, Vol. 69, 65-87, 1999. Actualmente María del Carmen Pardo Llorente es Profesora Titular de Universidad (área de Estadística e Investigación Operativa) en la Universidad de Complutense de Madrid.

En el trabajo se introduce y estudia una nueva familia de test estadísticos basada en la medida de Rphi-divergencia introducida por Burbea y Rao para contrastar bondad de ajuste. La distribución asintótica de estos estadísticos es una ji-cuadrado tanto bajo hipótesis nula simple como compuesta. Para hipótesis compuesta, los parámetros desconocidos se estiman tanto por máxima verosimilitud como por mínima Rphi-divergencia.

Premio Ramiro Melendreras 1998

En el congreso XXIV, celebrado en Aguadulce (Almería), fue premiado *Joaquín Sánchez Soriano* de la Universidad Miguel Hernández de Elche, por el trabajo titulado “La solución arco igualitaria: una solución ad hoc para la clase de los juegos de transporte”. El trabajo se publicó en *European Journal of Operational Research*, Vol. 150, 220-231, 2003. Actualmente Joaquín Sánchez Soriano es Profesor Titular de Universidad (área de Estadística e Investigación Operativa) en la Universidad Miguel Hernández de Elche.

El objetivo principal del trabajo consistió en introducir un concepto de solución particular para los juegos de transporte. Esta solución está relacionada con la estructura del problema de transporte subyacente, en particular con las soluciones óptimas para él. La solución es llamada solución *pairwise* igualitaria. En el trabajo se compara la nueva solución con el núcleo y el valor de Shapley del juego. Finalmente, se estudian las propiedades que satisface la nueva solución y se proporcionan dos caracterizaciones axiomáticas de la misma.

Premio Ramiro Melendreras 2000

En el congreso XXV, celebrado en Vigo (Pontevedra), fue premiado *Jacobo de Uña Álvarez* de la Universidad de Vigo, por el trabajo titulado “Eficiencia asintótica de la estimación límite-producto bajo censura proporcional”. El trabajo aun no está publicado. Actualmente Jacobo de Uña Álvarez es Profesor Titular de Universidad (área de Estadística e Investigación Operativa) en la Universidad de Vigo.

En el trabajo se estudia la eficiencia de la estimación límite-producto de Kaplan-Meier (la técnica estándar para el análisis de datos censurados) bajo el modelo de censura proporcional de Koziol-Green. En particular, se plantea y analiza la cuestión de cuán deficiente puede ser la estimación límite-producto en comparación con la basada en el estimador ACL (el estimador no paramétrico de máxima verosimilitud bajo censura proporcional). El estudio se realiza para el problema de la estimación de un parámetro genérico, que es una función arbitraria de k funcionales. Uno de los resultados principales del trabajo es que la eficiencia de la estimación límite-producto relativa a la ACL es mayor o igual que la proporción teórica de datos no censurados, independientemente del parámetro que se estime, y del modelo de distribución subyacente; además, se demuestra que esta cota no puede ser mejorada.

Premio Ramiro Melendreras 2001

En el congreso XXVI, celebrado en Úbeda (Jaén), fue premiado *Isabel Molina Peralta* de la Universidad Miguel Hernández de Elche, por el trabajo titulado “Los estadísticos de Rényi en el contraste de hipótesis para varias muestras, con una aplicación a datos familiares”. El trabajo se publicó en *Communications in Statistics - Theory and Methods*, Vol. 32, 415-434, 2002. Actualmente Isabel Molina Peralta es Profesora Visitante (área de Estadística e Investigación Operativa) en la Universidad Carlos III de Madrid.

El Test de la Razón de Verosimilitudes da, mediante un cociente, una medida de "distancia" entre el modelo restringido y el no restringido a la hora de resolver un contraste de hipótesis. Esta idea es extrapolable a otras pseudo-distancias, tales como las pertenecientes a la familia de divergencias de Rényi. En el trabajo se obtiene la distribución asintótica de los estadísticos de contraste basados en divergencias de Rényi, así como del Test de la

Razón de Verosimilitudes, en problemas basados en varias muestras. Esto se aplica a datos familiares normales multivariantes extraídos de distintas poblaciones, donde se plantea un contraste de igualdad de coeficientes de correlación lineal.

Premio Ramiro Melendreras 2003

En el congreso XXVII, celebrado en Lleida, fue premiado *Francisco Javier Toledo Melero* de la Universidad Miguel Hernández de Elche, por el trabajo titulado “Distancia al mal-planteamiento en programación lineal ordinaria y semi-infinita”. El trabajo se publicará en *Mathematical Programming, Series A* (Publicado online en diciembre de 2004). Actualmente Francisco Javier Toledo Melero es Profesor Ayudante Doctor (área de Matemática Aplicada) en la Universidad Miguel Hernández de Elche.

Este trabajo se desarrolla en el contexto de la Programación Lineal Semi-Infinita, que engloba el de la Programación Lineal Ordinaria. Se dice que un problema está mal-planteado con respecto a una propiedad cuando perturbaciones arbitrariamente pequeñas de los datos pueden dar lugar tanto a problemas que tienen dicha propiedad como a problemas que no la tienen. El trabajo está dedicado al estudio del mal-planteamiento con respecto a las propiedades de consistencia y de resolubilidad, proporcionando en ambos casos caracterizaciones de los problemas mal-planteados en términos de los

datos del problema, así como fórmulas explícitas de las correspondientes distancias al mal-planteamiento, las cuales, convierten el problema de determinar distancias en espacios infinito-dimensionales en el de calcular distancias en espacios finito-dimensionales.

Premio Ramiro Melendreras 2004

En el congreso XXVIII, celebrado en Cádiz, fue premiado *Alberto Olivares González* de la Universidad de Rey Juan Carlos, por el trabajo titulado “Optimización no convexa utilizando una búsqueda lineal adaptada”. El trabajo figura en las actas del congreso, pero aun no está publicado en revista científica. Actualmente Alberto Olivares González es Ayudante LOU (área de Estadística e Investigación Operativa) en la Universidad de Rey Juan Carlos.

En el trabajo se propone un nuevo algoritmo para la resolución de problemas de optimización sin restricciones no convexos. El algoritmo garantiza la convergencia a puntos que cumplen las condiciones de optimalidad de segundo orden. A partir de dos direcciones de descenso (una dirección de tipo Newton y otra de curvatura negativa) el algoritmo incluye un procedimiento de decisión que selecciona en cada iteración el modelo de búsqueda lineal más adecuado. El trabajo concluye con algunos experimentos numéricos que corroboran la eficacia del procedimiento.

INFORME DE LA ACCIÓN ESPECIAL BFM2002-12271-E MINISTERIO DE EDUCACIÓN Y CIENCIA CENTRO NACIONAL DE MATEMÁTICAS (CNMat)

Manuel de León¹. Coordinador de Matemáticas de la ANEP
Enrique Zuazua². Gestor del Programa Nacional de Matemáticas

1. Introducción.

En los últimos 25 años la investigación matemática española ha experimentado un incremento que podríamos calificar de espectacular, pasando de una producción en artículos en ISI del 0,3% en 1980 al casi 5% actual. En la Tabla I se aprecia ese crecimiento en los últimos 11 años por quinquenios acumulados: en la primera fila se incluye el porcentaje de artículos en revistas incluidas en el Journal Citation Reports (JCR) en los cuales figura al menos un autor español, y en la segunda fila el factor medio de impacto respecto a la media mundial.

1993-97	1994-98	1995-99	1996-00	1997-01	1998-02	1999-03
3.46	3.66	3.88	4.18	4.42	4.53	4.65
-17	-14	-15	-16	-13	-13	-6

Tabla I: Producción española en Matemáticas y Factor de impacto medio. Fuente: Web of Knowledge, ISI-THOMSON.

Este crecimiento se ha producido sin ningún tipo de orientación, sujeto por tanto a asimetrías temáticas y territoriales:

- Por una parte, se han desarrollado de una manera quizás excesiva algunas áreas de investigación.

- Por otra, se observan grandes carencias, sobre todo en las áreas más próximas a las aplicaciones, tanto a las Ciencias como en conexión con los desarrollos tecnológicos.

Otro elemento que ha contribuido a este cierto desajuste es la proliferación de universidades y en consecuencia de Departamentos y Licenciaturas de Matemáticas, que han llevado a un aumento significativo del número de equipos de investigación. Se ha producido así una tendencia a la atomización, contrariamente a los intentos de aunar esfuerzos y equipos de los diferentes Planes Nacionales en las convocatorias anuales de proyectos.

Estas circunstancias quedan reflejadas en la todavía escasa incidencia de las Matemáticas españolas en el contexto internacional. Nadie puede negar la existencia de grupos de gran calidad e impacto con unas fuertes conexiones internacionales, pero los datos bibliométricos muestran que España no posee instituciones con influencia en el campo. En la tabla II, se muestran las cinco instituciones españolas que aparecen en Web of Knowledge, entre 165, con sus posiciones relativas en citas por artículo:

Puesto	Institución	Artículos	Citas totales	Citas por artículo
# 109	UAM	468	1475	3,15
# 124	UCM	783	2356	3,01
# 138	UAB	447	1206	2,70
# 147	UB	459	1195	2,60
# 150	UGR	840	2090	2,49

Tabla II: Datos de las 5 instituciones españolas incluidas en Essential Science Indicators, Enero 1, 1994 - Octubre 31, 2004.

Los índices de citas por artículo pueden contrastarse con el índice medio en el mundo, de 2,51 citas por artículo en ese período.

Este análisis, que se podría extender a otras áreas científicas además de las Matemáticas, pone de manifiesto la necesidad de construir nuevas estructuras para que las Matemáticas españolas afronten los desafíos científicos y tecnológicos del siglo XXI en condiciones competitivas con los países desarrollados. El reciente Programa Nacional de Matemáticas ha supuesto un primer paso en esa dirección, y ya en él se ponía de manifiesto esta necesidad y se proponía la creación de un Centro Nacional de Matemáticas que, de una manera similar a lo que ocurre en otros países, contribuyera a vertebrar y orientar la investigación en el área. Esta nueva estructura debería contemplar con un

énfasis especial la labor de formación de jóvenes investigadores, en particular en las líneas frontera de la investigación interdisciplinar y en los campos emergentes.

El CNMat, debido a la amplitud de su ámbito de influencia, podría:

- Acometer tareas y plantearse objetivos que están fuera de las posibilidades de las distintas Universidades (y de los Institutos de Matemáticas actuales o futuros).
- Permitiría perfeccionar la orientación de la investigación matemática, fomentar las direcciones más útiles a la sociedad y conectar a los investigadores matemáticos con ésta.
- Facilitaría que las regiones menos competitivas elevaran el nivel de calidad de su investigación matemática, con la consiguiente influencia en el entorno científico-tecnológico.

El desarrollo de la propuesta contenida en el Programa Nacional de Matemáticas fue objeto de una Acción Especial del MEC (BFM2002-12271-E), dirigida por Manuel de León (Coordinador de Matemáticas de la ANEP) y Enrique Zuazua (Gestor de Matemáticas en el Plan Nacional), en el marco de la cual se organizaron dos reuniones con expertos nacionales e internacionales (véase el Anexo). La propuesta que sigue es consecuencia de las reflexiones surgidas tras las mismas.

2. Objetivos

El Centro Nacional de Matemáticas (CNMat) persigue los siguientes objetivos:

- Coordinar, dinamizar y orientar la investigación matemática española.
- Desarrollar la interacción entre las Matemáticas y otras Ciencias.
- Desarrollar la interacción de las Matemáticas con los sectores tecnológicos y financieros.
- Desarrollar la interacción de las Matemáticas con la Industria, fomentando la transferencia del conocimiento matemático a los sectores productivos.
- Desarrollar, en colaboración con las Universidades, Programas de Doctorado y Escuelas de formación en investigación con fuerte impacto internacional.
- Identificar y desarrollar programas de investigación estratégicos y, en particular, favorecer y mejorar el nivel de calidad de la actividad investigadora en Matemáticas en las zonas del Estado donde se detecte esta necesidad.
- Proporcionar un marco adecuado para la incorporación de investigadores de forma

temporal, en coordinación con los diferentes programas a desarrollar.

- Fomentar la cooperación internacional en Matemáticas contribuyendo especialmente a la construcción del Espacio Europeo de Investigación y facilitando la participación de grupos españoles en redes europeas.
- Favorecer e impulsar la cooperación internacional con potencias emergentes en Matemáticas, y especialmente con los países latinoamericanos, con los cuales ya existe una colaboración significativa.
- Colaborar al óptimo funcionamiento y coordinación de actividades de los diferentes Institutos Universitarios y Centros de Investigación en marcha. A la vez, impulsar la creación de otros nuevos en las diferentes Comunidades Autónomas, con la cofinanciación de éstas, en especial en aquellos casos de áreas de especialización con mayor presencia: Institutos de Matemática Industrial, de Computación, de Finanzas, etc.
- Servir de foro a la comunidad matemática española en los ámbitos de la enseñanza en los diferentes niveles educativos.
- Favorecer la divulgación de las Matemáticas en la sociedad española.
- Proporcionar a la comunidad matemática española acceso a infraestructuras como:

- o Biblioteca de referencia con acceso electrónico
- o Centro de Documentación científica
- o Centro de Computación
- o Soporte y ayuda a las sociedades matemáticas españolas
- o Soporte a las revistas matemáticas españolas de prestigio internacional, así como a otras publicaciones de prestigio dirigidas por investigadores españoles.
- o Soporte a las redes temáticas.
- o Soporte a las redes europeas.

3. Estructura del CNMat

3.1 Personalidad jurídica

El CNMat podría configurarse como un Organismo Autónomo con personalidad jurídica y patrimonio propios, que dependiera orgánicamente del Ministerio de Educación y Ciencia.

Una alternativa sería la constitución de una Fundación.

En cualquier caso, el CNMat establecería convenios de colaboración con el CSIC, Comunidades Autónomas, Universidades y otras instituciones públicas y privadas, con el

fin de desarrollar adecuadamente sus programas. De este modo, las Comunidades Autónomas, las distintas Universidades, el CSIC, las Instituciones privadas, etc. se convertirían en los distintos agentes promotores del CNMat, con derecho a participar en su gestión

En el marco de esos acuerdos, el CNMat coordinaría sus esfuerzos con los de los centros de investigación ya existentes, siempre intentando concentrar esfuerzos y recursos y buscando sinergias.

3.2 Órganos de gobierno

El CNMAT contaría con los siguientes órganos de gobierno:

- Director, Subdirector y Gerente.
- Junta de Gobierno
- Comisión Científica asesora

El Director, Subdirector y Gerente serán nombrados por el MEC a propuesta de la Junta de Gobierno. Los mandatos serán por un periodo de cuatro años, renovables por una sola vez. Ellos serán los únicos miembros permanentes mientras duren sus mandatos. El Director y el Subdirector serán matemáticos que desarrollan su actividad en España, y con experiencia probada en investigación y en gestión de la investigación. La dedicación a sus cargos será a tiempo completo. El Gerente será contratado directamente a cargo del presupuesto anual del CNMat a fin de conseguir una gestión más ágil.

La Junta de Gobierno incluirá representantes del MEC y de todos los agentes que participen en el CNMat, y tendrá como misión seleccionar entre las propuestas presentadas y que previamente hayan sido evaluadas favorablemente por el Comité Científico. La Junta de Gobierno tendrá un número de miembros adecuado para garantizar su operatividad, y se reunirá al menos una vez al año. El Director, Subdirector y Gerente serán miembros de oficio de la Junta de Gobierno. La Junta de Gobierno elaborará y aprobará tan pronto como sea posible un reglamento de funcionamiento del CNMat, donde quedarán recogidas con precisión (entre otras cosas) sus atribuciones, así como la de los otros órganos de gobierno.

La Comisión Científica asesora será nombrada por el MEC, estará formada por investigadores del más alto nivel internacional, y tendrá como misión aconsejar a la Junta de Gobierno sobre las propuestas presentadas al CNMat. La

Comisión Científica podrá incorporar expertos de los ámbitos empresarial y tecnológico así como científicos no pertenecientes al área de las Matemáticas, a fin de favorecer las aplicaciones y la interdisciplinaridad.

3.3 Sede e Instalaciones

El CNMat contaría con una sede física, en un edificio propio, que debería estar situada en un campus universitario que proporcionase el entorno más favorable para el desarrollo de sus actividades. Por tanto, es muy conveniente que ese entorno reúna las características más apropiadas para que el CNMat pueda convertirse en el plazo más breve posible en un centro de referencia nacional e internacional. Se enumeran a continuación algunas de ellas, que son las que reúnen habitualmente los centros similares con un mayor impacto internacional:

- Compromiso en firme de cofinanciación por parte de las universidades y Comunidades Autónomas implicadas.
- Existencia de una Licenciatura de Matemáticas en ese campus.
- Existencia de al menos un Programa de Doctorado de prestigio.
- Existencia en ese entorno de un número apreciable de grupos de investigación de calidad en Matemáticas, que mantengan relaciones internacionales del más alto nivel, con pertenencia a redes europeas y fuertes contactos con los países de vanguardia.
- Tradición en series de seminarios y coloquios.
- Existencia de un Centro de Computación capaz de cubrir al nivel adecuado las necesidades de los grupos de investigación en Matemática Aplicada y Computacional.
- Entorno científico y tecnológico que permita una interacción fácil de los matemáticos con otros científicos y tecnólogos.
- Ubicación geográfica idónea y máximas facilidades de acceso.

En resumen, el CNMat debe servirse del mejor entorno posible para poder dar el mejor servicio a la totalidad de la comunidad matemática española.

El edificio albergaría:

- Locales para uso del Director, Subdirector y Gerente
- Locales para uso del personal de administración adscrito
- Biblioteca
- Despachos para visitantes
- Aulas para seminarios
- Salas de ordenadores
- Salas de reuniones

- Locales para uso de las sociedades y publicaciones matemáticas

En caso de ser necesaria la construcción de un edificio, los terrenos para ello serían cedidos por la universidad correspondiente. Los costes de construcción o acondicionamiento deberían ser asumidos por todos los agentes involucrados en el proyecto: MEC, Comunidades Autónomas, Universidades, CSIC, Instituciones privadas, en cuantías que se establecerían en su momento.

Complementariamente, el CNMat debería contar con un Centro de Encuentros, que podría gestionarse en colaboración con alguna Comunidad Autónoma. Dada la diversidad de la oferta y la tradición existente en España en centros de reuniones es previsible que esta cuestión pueda resolverse con facilidad, con las máximas posibilidades de éxito y con un coste muy bajo.

El CNMat dispondrá de una Unidad Administrativa propia. El presupuesto del CNMat contemplará la contratación del personal de administración necesario para cumplir sus objetivos. Este personal será contratado por la Dirección del CNMat bajo la supervisión de la Junta de Gobierno.

4. Programas a desarrollar en el CNMat

El CNMat tiene como objeto estimular y contribuir a encauzar las iniciativas más relevantes en el ámbito de las Matemáticas. Para ello abrirá anualmente una convocatoria en la que se admitirán todo tipo de proyectos que serán seleccionados por la Junta de Gobierno tras escuchar a la Comisión Científica, teniendo en cuenta criterios de calidad y oportunidad, así como la cofinanciación ofrecida por los promotores. En este punto conviene subrayar que el CNMat tiene también vocación de optimizar recursos y convertirse en un receptor de las inversiones privadas de empresas y otras instituciones sin ánimo de lucro en el ámbito de las Matemáticas. Entre los programas y actividades que desarrollará el CNMat están los siguientes:

- Trimestres/semestres temáticos
- Coloquios y seminarios
- Escuelas de especialización
- Programas de Doctorado
- Acogida de investigadores nacionales y extranjeros
- Jornadas de Política Científica y Tecnológica
- Convocatorias de becas asociadas a los trimestres/semestres temáticos
- Convocatorias de becas asociadas a los programas de investigación estratégicos

- Divulgación científica de las Matemáticas, aspectos educativos e impacto social

Los programas serán desarrollados principalmente en la sede del CNMat, aunque éste estimulará también actividades matemáticas de diversa índole en otros centros del estado.

5. Financiación

La financiación basal del CNMat (mantenimiento de la sede y programas) estarán a cargo de:

- Dirección General de Investigación
- CSIC
- Comunidades Autónomas
- Universidades
- Financiación privada de Fundaciones y Empresas
- Fondos de Programas Europeos

La financiación del CNMat no debería estar sujeta a fluctuaciones, sino que debería ser consolidada, a fin de poder diseñar adecuadamente sus programas a medio y largo plazo. En caso de constituirse en un organismo autónomo, este presupuesto consolidado debería aparecer anualmente en los Presupuestos Generales del Estado. Naturalmente, el CNMat concurriría adicionalmente a las convocatorias de la administración central, administraciones autonómicas y fondos europeos.

Puesta en marcha inicial del CNMat

El CNMat tendría una Comisión Gestora provisional a lo largo del año 2005 hasta que el Centro se configurara definitivamente. Esta Comisión Gestora sería nombrada formalmente por el MEC y trabajaría en coordinación directa con la DGI hasta ese momento.

La decisión sobre la ubicación de la sede del CNMat sería tomada por el MEC, y tendría en cuenta las recomendaciones de la Comisión Gestora.

Las actividades del CNMat deberían comenzar el 1 de enero de 2006, aprovechando la celebración en Madrid del International Congress of Mathematicians en ese año. Aún cuando sería difícil que el CNMat dispusiera en esa fecha de un edificio, sin embargo podría ya comenzar a desarrollar determinados programas como son el impulsar un Programa de Doctorado de Calidad de ámbito internacional y

trimestres temáticos. En cualquier caso el nombramiento y actividades de la Comisión Gestora deberían mantenerse hasta la puesta en marcha efectiva del CNMat.

ANEXO

El principal objetivo de la Acción Especial del MEC (BFM2002-12271-E) fue llevar adelante una reflexión articulada sobre el formato del futuro Centro Nacional de Matemáticas.

Para llevar adelante tal reflexión, los responsables de la Acción, Manuel de León y Enrique Zuazua, diseñaron dos reuniones:

- 17 y 18 de junio de 2004, en el CSIC, reuniendo responsables de la Política Científica española así como a destacados investigadores españoles con puestos de responsabilidad en la gestión de algunas iniciativas ya existentes (o en trámite) en nuestro país: CRM, Institutos Universitarios de Matemáticas, Institutos en campos afines.
- 18 y 19 de noviembre, en la Universidad de Santiago de Compostela, con el apoyo local del Director del Instituto de Matemáticas de Santiago, Juan José Nieto Roig, y el Decano de la Facultad de Matemáticas, Juan Manuel Viaño Rey. En esta reunión se trazó un amplio panorama por investigadores internacionales de la organización de la investigación matemática en Europa, Estados Unidos y Latinoamérica. Debemos destacar la presencia de los máximos responsables de la organización de la investigación matemática en Estados Unidos (William Rundell, Director de la División de Matemáticas de la NSF) y Francia (Aline Bonami).

Las conclusiones reflejadas en el informe precedente son fruto de las reflexiones de los organizadores de estos dos eventos (MdL & EZ) tras haber escuchado y debatido ampliamente el tema con los participantes en estas dos reuniones.

¹ Profesor de Investigación del CSIC, Instituto de Matemáticas y Física Fundamental, Consejo Superior de Investigaciones Científicas, Serrano 123, 28006 Madrid. Email: mdeleon@imaff.cfmac.csic.es

² Catedrático de Matemática Aplicada, Departamento de Matemáticas, Universidad Autónoma de Madrid, Cantoblanco 28049 Madrid. Email: enrique.zuazua@uam.es

3. ARTÍCULOS DE ESTADÍSTICA

JACOB BERNOULLI (1654-1705) A LOS 350 AÑOS DE SU MUERTE

Miguel A. Gómez Villegas



El 27 de Diciembre de 1654 nació en Basilea (Suiza) Jacobo James Bernoulli, por lo tanto en el año 2004 se cumplen los 350 años del nacimiento del autor del Arte de la Conjetura (Ars Conjectandi), uno de los creadores de los fundamentos del Cálculo de Probabilidades.

Jacob Bernoulli procedía de una familia protestante que tuvo que escapar de Holanda huyendo de las guerras de religión entre católicos y protestantes.

Inició sus estudios de arte que acabó en 1671, para licenciarse seguidamente en teología en 1676. Al mismo tiempo, y en contra de la opinión de su padre, estudió matemáticas y astronomía. Después de su graduación trabajó en Ginebra como tutor, para viajar seguidamente por Francia, Holanda, Inglaterra y Alemania. Volvió a Basilea en 1683 y obtuvo la plaza de profesor de mecánica en la universidad de Basilea. Rechazó trabajar como eclesiástico para dedicarse al estudio de las matemáticas, consiguiendo la cátedra de matemáticas de la universidad de Basilea en 1687, manteniéndose en el mismo puesto hasta su muerte en 1705.

En el campo de la matemática, Bernoulli realizó importantes contribuciones en álgebra, cálculo infinitesimal, cálculo de variaciones y series.

Su hermano Johan Bernoulli (1667-1748) también estudió matemáticas y aunque de joven recibió clases de Jacob, los dos acabaron teniendo problemas por la paternidad de distintos trabajos. Parece ser que éste fue el principal motivo por el que El Arte de la Conjetura no vio la luz hasta que el sobrino de Jacob, Daniel Bernoulli, se decidió a publicarlo, cuando ya la polémica se había enfriado. Alguno de los biógrafos de Jacob describe a éste como *”terco, obstinado, agresivo, vengativo, con un cierto sentimiento de inferioridad, pero sin embargo íntimamente convencido de sus propias habilidades”*.

Como se ve, en todas las familias, aunque sean de matemáticos, existen problemas.

El trabajo por el que es más conocido es su libro El Arte de la Conjetura. Representa la transición desde las aportaciones realizadas por Christian Huygens en su memoria El Razonamiento en los Juegos de Azar (De Ratiociniis in Ludo Aleae) a una nueva Teoría, centrada ya en el concepto de probabilidad.

El Arte de la Conjetura se divide en cuatro partes. La primera es una reedición del trabajo de Huygens; la segunda recoge la teoría de las combinaciones y de las permutaciones, en la tercera resuelve 24 problemas relativos al concepto de esperanza de varios juegos, concepto que había sido introducido por Huygens. La cuarta parte es, desde el punto de vista de la evolución y el desarrollo de la estadística y el cálculo de probabilidades, la más interesante. En ella se incluyen aplicaciones a temas civiles, morales y económicos. Tres aspectos interesa destacar con respecto a esta parte.

El primero es que en ella se prueba por primera vez la convergencia en probabilidad de la media muestral a la media poblacional para variables aleatorias de Bernoulli. Con notación actual, Bernoulli prueba que si $X_i \sim \text{Bernoulli}(\theta)$ entonces

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \theta.$$

Su demostración es, aunque larga, perfectamente rigurosa y más natural que otras demostraciones que se

hicieron después. Llega a la necesidad de probar este resultado al tratar de determinar, hoy decimos estimar, el valor de θ , la probabilidad de éxito en una sola repetición.

El segundo aspecto que interesa destacar de esta parte cuarta, es que en ella introduce lo que llama grado de certeza moral, que es ya una aproximación al concepto de probabilidad subjetiva. Así dice

"Hay certeza moral cuando la probabilidad es casi igual a la certeza total... si una cosa es considerada moralmente cierta cuando tiene 999/1000 de certeza, otra ser moralmente imposible cuando tenga 1/1000 de certeza".

Un estudio bastante detallado de la saga de los Bernoulli puede verse en el libro de Sánchez y Valdés (2001). Una traducción al castellano de la parte cuarta del Arte de la Conjetura puede verse en Rivadulla (1993).

Con la ley débil de los grandes números para la distribución de Bernoulli, el autor ofrece por primera vez una concreción matemática de la medida de la incertidumbre. Bernoulli no sólo prueba cualitativamente que a mayor número de observaciones hay menos incertidumbre en el resultado de la probabilidad de éxito para la distribución que lleva su nombre, él cuantifica además la afirmación, garantizando el número de observaciones que han de realizarse, para que la frecuencia se aproxime a la probabilidad.

Esto nos lleva al tercer aspecto que nos interesa destacar de la última parte del Arte de la Conjetura y es que el trabajo queda incompleto, porque al tratar de determinar el número de observaciones a realizar para obtener la certeza moral de la probabilidad de éxito, con las aproximaciones que utiliza, el número de observaciones al que llega es demasiado grande como para que éste sea de utilidad práctica.

Como dice Stigler (1986) Jacob Bernoulli dió el primer paso de lo que sería un largo y fecundo viaje. Que este aniversario sirva para que cada vez más investigadores se conviertan en compañeros de tan interesante viaje.

Referencias

-Rivadulla, A. (1993) Teoría de probabilidades (Ars Conjectandi, Parte Cuarta, Basilea 1713), Boletín de la Sociedad Española de Historia de las Ciencias, 16, 389-418.

-Sánchez Fernández, C. Y Valdés Castro, C (2001) Los Bernoulli Geómetras y Viajeros, TresCantos (Madrid):Nívola.

-Stigler, S. M. (1986) The History of Statistics. The measurement of uncertainty before 1900, Cambridge: Harvard University Press.

4. ARTÍCULOS DE INVESTIGACIÓN OPERATIVA

EL LEMA DE FARKAS: UNA HERRAMIENTA PARA RESOLVER NUEVAS APLICACIONES

Juan José Salazar González

DEIOC – Universidad de La Laguna (jjsalaza@ull.es)

Resumen

Farkas (1986) dio (entre otros resultados) una condición matemática necesaria y suficiente para determinar cuándo un poliedro es o no vacío, y que posteriormente se ha denominado *Lema de Farkas*. Tradicionalmente este resultado se considera como un “resultado teórico”, o al menos así lo clasifican nuestros alumnos. El objetivo de estas notas es mostrar que el Lema de Farkas también tiene gran valor práctico dentro de la Optimización Matemática porque su oportuno uso dentro de algoritmos de “ramificación y corte” permite resolver problemas de tamaño mayor. Para alcanzar este objetivo se presentan cuatro aplicaciones reales, y en cada una se ilustra cómo utilizar el Lema de Farkas.

Introducción

Sea A una matriz con m filas y n columnas, denotadas éstas por a^1, \dots, a^n , sea y un vector de \mathfrak{R}^n y sea b un vector de \mathfrak{R}^m . Tradicionalmente el Lema de Farkas se presenta en la siguiente versión:

El sistema $Ay=b, y \geq 0$ tiene solución y en \mathfrak{R}^n si, y sólo si, el sistema $A^T u \geq 0, b^T u < 0$ no tiene solución u en \mathfrak{R}^m .
--

Intuitivamente es un resultado trivial: o bien b está en el cono convexo que generan las columnas de A (es decir, existe escalares no negativos y_1, \dots, y_n tales que $b = y_1 a^1 + \dots + y_n a^n$), o bien existe un hiperplano que deja a un lado el vector b y a otro lado las columnas de A (es decir, existe un gradiente u tal que $u^T a^i \geq 0$ para todo $i=1, \dots, n$, pero $u^T b < 0$). Ahora bien, a la vez es también un resultado fundamental. Así lo hacemos saber a nuestros alumnos, principalmente argumentando que sobre él se sustenta, por ejemplo, la Teoría de la Dualidad en Programación Lineal.

El objetivo de estas notas es acentuar esta importancia, haciendo especial énfasis en sus inmediatas implicaciones prácticas. La clave es observar que el Lema de Farkas da las condiciones que tiene que cumplir A y b para que el poliedro definido por

$$P = \{y \in \mathbb{R}^n : Ay=b, y \geq 0\}$$

sea no vacío. Estas condiciones son:

$$u^T b \geq 0 \text{ para toda dirección } u \text{ del cono poliédrico } C = \{u \in \mathbb{R}^m : A^T u \geq 0\}.$$

El Teorema de Caratheodory garantiza que este cono está generado por un número finito de vectores u^1, \dots, u^d , llamados *direcciones extremas*. Teniendo presente este resultado y la convexidad de C , para asegurar que $P \neq \emptyset$ basta controlar que se cumplan un número finito de desigualdades: $b^T u^j \geq 0$ para todo $j=1, \dots, d$. En sí mismo éste me parece un resultado destacable porque muchos de nuestros alumnos conocen cómo verificar si un sistema de igualdades lineales es o no factible, pero no tantos responden a cómo se puede determinar si un sistema de desigualdades lineales admite o no solución.

Es cierto que en la práctica el número d de direcciones extremas de un cono poliédrico puede ser muy grande, y por tanto las condiciones anteriores pueden parecer empíricamente inútiles. Sin embargo, es fundamental observar que para verificar si se cumplen *todas* las desigualdades anteriores *no* es necesario controlarlas una a una. Basta resolver el problema de Programación Lineal:

Minimizar $b^T u$ Sujeto a $u \in C$.

Gracias al Lema de Farkas, $P = \emptyset$ si y sólo si ese programa lineal es no-acotado, y en este caso una dirección de no-acotación u^* (caracterizada por cumplir $A^T u^* \geq 0$ y $b^T u^* < 0$) da además el certificado que lo demuestra. Lo importante de este resultado no es que la pregunta “¿ $P = \emptyset$?” equivale a resolver un problema lineal, ya que esto es incluso más evidente resolviendo “*minimizar* $0^T y$ *sujeto a* $y \in P$ ”. Lo importante es que para garantizar que P contiene algún punto y sería necesario imponerle a b la condición $b^T u^* \geq 0$ (llamada *corte*).

Este simple razonamiento es la idea base que permite resolver varias aplicaciones reales, como veremos en las próximas secciones. La característica común a estas (y otras) aplicaciones es que en ellas se debe buscar una decisión x de manera que exista un vector y cumpliendo una familia de desigualdades lineales cuyo término independiente depende de x . En ocasiones estas variables y , junto con las x , pueden suponer demasiadas variables para constituir un modelo matemático resoluble en la práctica. La introducción dinámica de cortes mediante la resolución de problemas lineales da una alternativa que reduce las variables del modelo a sólo las x , haciéndolo más tratable en la práctica. Adicionalmente los cortes que se generan pueden en ocasiones fortalecerse aumentando o disminuyendo (en inglés “*lifting*”) algunos de sus coeficientes o su término independiente, dando así modelos matemáticos con relajaciones lineales más ajustadas.

El esquema global corresponde a una técnica iterativa de hiperplanos de corte. En cada iteración se resuelve un *problema maestro* que es una relajación del problema original y que sólo tiene variables de tipo x . A partir de la solución óptima x^* se resuelve el programa lineal anterior (llamado *subproblema*), y si resulta no acotado entonces se tiene una dirección de no-acotación u^* . Con esta dirección se construye un corte para fortalecer el problema maestro de la siguiente iteración, evitando así que vuelva a aparecer la solución x^* . Dado que el número de direcciones d es finito, este proceso iterativo es un algoritmo que termina con una solución óptima de la aplicación original. Cuando x es un vector de números enteros, puede verse como la técnica de introducida en Benders (1962). Sin embargo, como característica especial, aquí aparece el hecho de que en varias aplicaciones donde x es un vector de enteros, no es necesario mantener que el problema maestro sea un problema de Programación Entera. En efecto, es posible utilizar este esquema iterativo resolviendo sólo problemas maestros y subproblemas lineales si se aplica en cada nodo de un árbol de ramificación sobre el vector de variables x . Es decir, el procedimiento puede verse como una técnica de *ramificación y corte*.

La primera aplicación que veremos nace en los Institutos de Estadística cuando desean proteger celdas con valores confidenciales en tablas estadísticas. El problema es decidir qué ocultar pero minimizando la pérdida de información en la que se incurre. Otra aplicación nace en el campo del transporte, cuando se necesita diseñar la ruta de un vehículo con capacidad limitada para que recoja y entregue demandas de un producto a través de unos clientes. El objetivo es minimizar los costes de la ruta mientras se mantienen los límites de carga del vehículo y las demandas de los clientes. La tercera aplicación aparece cuando se necesitan depurar encuestas estadísticas que no cumplen un conjunto de reglas de coherencia. En estos procesos el primer objetivo es localizar el menor número de errores potenciales pero garantizando que existirá un valor imputable durante la corrección de dichos errores (segundo objetivo). Por último se presenta una aplicación en el campo de las telecomunicaciones, donde se desea decidir concentradores a instalar de manera que puedan realizarse conexiones entre orígenes y destinos con los menores costes.

Protección de Datos Confidenciales

Los institutos de estadística nacionales o autonómicos recogen información de individuos o empresas bajo el compromiso de garantizar la privacidad. Luego agregan estos datos en tablas donde hay totales (por filas, columnas, etc.). Aplicando ciertas reglas determinan las llamadas *celdas sensibles*, y cuyos valores no pueden hacerse públicos porque revelarían información confidencial. Como consecuencia estos valores se reemplazan por asteriscos, dando lugar a las llamadas *supresiones primarias*. Ahora bien, debido a que las tablas muestran totales marginales, la sola supresión de las celdas sensibles no garantiza que un intruso no calculará su valor original sin más que usar las relaciones lineales que dan los valores totales a partir de los valores internos. Por tanto se deben suprimir también (quizás) otras celdas no sensibles, y que se llaman *supresiones secundarias*. Ahora bien, ¿qué celdas elegir como secundarias? Cada supresión conlleva una pérdida de información, y el objetivo es minimizar la pérdida global como consecuencia de todas las supresiones, pero asegurando en todo momento que un intruso no podrá calcular (ni exacta ni aproximadamente) los valores originales de las celdas sensibles.

La decisión de suprimir o no una celda i puede modelarse como una variable x_i que toma el valor 1 en caso de optar por sustituir su valor por un asterisco, y 0 en otro caso. Una solución x para toda la tabla será válida (aunque quizás no óptima) si garantiza que un intruso no detecta el valor exacto de ninguna celda sensible. Lo deseable, por tanto, es que las celdas i con $x_i=1$ generen suficiente “incertidumbre” para cada celda sensible, entendiendo por incertidumbre el hecho de que existan valores potenciales para los asteriscos que garanticen un nivel de protección mínimo y un nivel de protección máximo para cada celda sensible. Estas garantías se pueden modelizar mediante vectores de variables y , cada uno representando una posible tabla para cada celda y cada para cada nivel. Es decir, se hablaría de dos vectores y^{j+} e y^{j-} para cada celda sensible i en una tabla. Si pensamos en una tabla bidimensional con 100 filas y 100 columnas, y con un 10% de celdas sensibles, esto significaría un modelo con 10,000 variables 0-1 y con 2 millones de variables continuas. En su conjunto esto determina un problema de Programación Matemática demasiado grande para ser manejable en la práctica.

El Lema de Farkas da otra alternativa de resolución. Efectivamente para que una solución x sea factible se requiere que existan los vectores y^{j+} e y^{j-} , es decir, que un conjunto dependiente de x sea no vacío. Ahora bien, es posible reemplazar esta enorme cantidad de variables por unos cortes que se incorporan dinámicamente en un problema maestro a base de resolver una secuencia de subproblemas lineales. El lector puede encontrar detalles técnicos de este procedimiento en Salazar (2004), donde se aplica este esquema iterativo en diferentes metodologías dentro del Control de la Privacidad en Tablas Estadísticas.

Transporte de una mercancía a través de varios clientes

Supongamos que una compañía debe diseñar la ruta de un vehículo que debe mover un producto entre varios clientes. El producto puede ser, por ejemplo, dinero, y los clientes pueden ser sucursales de una misma entidad bancaria. Algunos clientes dan unidades de este producto, mientras otros requieren demandas de ese mismo producto. Se conocen el coste de ir desde un cliente hasta otro y la capacidad máxima del vehículo. ¿Cómo debe servir el vehículo la demanda de los clientes para minimizar el coste global de la ruta?

Claramente hay que diseñar una ruta x , y en este sentido el problema se parece al clásico problema del Viajante de Comercio. El vector x tiene una variable decisoria 0-1 para cada arco de la red. La gran diferencia es que ahora hay demandas (positivas y negativas) en los clientes y capacidad en el vehículo, lo que hace que no todo circuito Hamiltoniano sea válido. Para garantizar la validez de una ruta x es

necesario que exista una carga y_a asociada a cada arco a con $x_a=1$. Y es aquí donde vuelve a aparecer la utilidad del Lema de Farkas. Por un lado se puede construir un modelo con las x y con las y , pero las características del problema hacen mucho más conveniente usar el Lema de Farkas para reemplazar las variables y por cortes en las variables x . Es más, el subproblema que aparece en esta aplicación tiene una estructura combinatoria especial, y es posible caracterizar todas las direcciones extremas. En consecuencia la familia de cortes, aunque siempre de gran tamaño, admite ser fortalecida, originando así problemas maestros con mejores características. Estas consideraciones han permitido resolver problemas con dimensiones antes no resueltas. Los detalles de estas investigaciones pueden consultarse en el artículo Hernández y Salazar (2004).

Depuración de Encuestas

Muchos organismos públicos recogen información de individuos o empresas como base para análisis estadísticos sobre los que tomar de decisiones. Haciendo uso de las llamadas *reglas de coherencia*, estos organismos detectan que algunas encuestas contienen errores y deben ser depuradas antes de entrar con las otras en el proceso de análisis. Por ejemplo, en base a una de estas reglas se puede considerar que hay un error en la encuesta de un individuo cuya edad dice “12 años” y cuyo estado civil dice “divorciado”. En base a discusiones previas, tales como la original en Fellegi y Holt (1979), está ampliamente aceptado que la forma de corregir una encuesta que no satisfaga todas las reglas es alterando el menor número de sus valores. Por tanto, el primer objetivo a afrontar durante la depuración de una encuesta es el problema de optimización que busca el menor número de campos a modificar de manera que existan luego valores imputables. Sea x_i la variable decisoria que toma el valor 1 si el campo i debe modificarse, y 0 en otro caso, y sea y_i el valor imputable al campo i cuando $x_i=1$. Una vez más sucede que se buscan soluciones x asociadas a un conjunto que debe contener alguna solución y . Cuando las respuestas son números reales y las reglas son relaciones lineales, estos conjuntos son poliedros, y el Lema de Farkas vuelve a dar una alternativa eficaz para resolver problemas de grandes dimensiones.

En esta aplicación la vinculación entre las variables x_i y las variables y_i es a través de los extremos del intervalo de valores para la respuesta i . Es decir, si el campo i puede tomar valores entre f_i y g_i , y si el valor original en la encuesta es h_i , entonces esta vinculación viene dada por las desigualdades $h_i + (f_i-h_i)x_i \leq y_i \leq h_i + (g_i-h_i)x_i$. Cuando f_i y g_i son valores muy alejados, el modelo matemático con sólo las variables x , más los cortes que vienen de aplicar el Lema de Farkas, ofrece diversas mejoras frente a un modelo matemático sin estos cortes y con las variables y . El lector puede encontrar los detalles técnicos en Riera y Salazar (2004).

Diseño de una red de comunicaciones

En el diseño de una red de comunicaciones se deben localizar concentradores y establecer las conexiones con el menor costo posible de manera que se garantice la existencia de caminos para enviar paquetes de información desde unos terminales a otros. La existencia de límites en las capacidades de los concentradores y de las conexiones obliga a fraccionar los paquetes. Por tanto las comunicaciones pueden ser flujos que circulan por una red con capacidades. Lo fundamental es que hay un costo asociado a todos los concentradores y conexiones por los que pase alguna cantidad de flujo. La decisión de si se abre o no un concentrador en un nodo i puede modelizarse mediante una variable x_i , y la decisión de si se establece una conexión o no a través de un arco a de la red se puede modelizar mediante una variable x_a . Para que una decisión x sea válida debe garantizarse que exista un flujo y que permita todas las comunicaciones deseadas, y es aquí donde vuelve a intervenir el Lema de Farkas. En efecto, sin el Lema de Farkas sería necesario trabajar con un modelo con variables (x,y) mientras que gracias al Lema de Farkas es posible eliminar las variables y a cambio de insertar dinámicamente algunos cortes. En esta aplicación concreta el subproblema que genera tales cortes tiene una especial estructura que permite una segunda descomposición reduciendo su resolución a una serie de problemas de caminos mínimos en un grafo. Se pueden consultar los detalles en Rodríguez y Salazar (2004).

Agradecimientos

El trabajo de investigación sobre recogidas y entregas de una mercancía a través de clientes ha sido realizado conjuntamente con Hipólito Hernández Pérez (hhperez@ull.es), el trabajo sobre depuración de encuestas estadísticas con Jorge Riera Ledesma (jriera@ull.es), y el trabajo sobre la red de comunicaciones con Inmaculada Rodríguez Martín (irguez@ull.es). Todos somos miembros del *Departamento de Estadística, Investigación Operativa y Computación* de la Universidad de La Laguna,

financiados en parte por los proyectos nacionales TIC2000-1750-C06-02 y TIC2003-05982-C05-02. Estas notas nacieron de una charla impartida en las *I Jornadas SEIO-RSME sobre Programación Matemática*, Elche, 6-7 de mayo de 2004.

Bibliografía

J.F. Benders, "Partitioning procedures for solving mixed-variables programming problems", *Numerische Mathematik* 4 (1962) 238-252.

G. Farkas, "A Fourier-féle mechanikai elv alkalmazásainak algebrai alapjáról", *Mathematikai és Fizikai Lapok* 5 (1986) 49-54. [en húngaro]

I.P. Fellegi, D. Holt, "A systematic approach to automatic edit and imputation", *Journal of the American Statistical Association* 71 (1976) 17-35.

H. Hernández-Pérez, J.J. Salazar-González, "A Branch-and-Cut Algorithm for the Traveling Salesman Problem with Pickups and Deliveries", *Discrete Applied Mathematics* 145 (2004) 126-139.

J. Riera-Ledesma, J.J. Salazar-González, "Algorithms for automatic data editing", *Statistical Journal of the United Nations ECE* 20 (2003) 255-264.

I. Rodríguez-Martín, J.J. Salazar-González, "Decomposition Approaches for a Capacitated Hub Problem", *Lecture Notes in Artificial Intelligence* 3315 (2004) 154-163.

J.J. Salazar-González, "A Unified Mathematical Programming Framework for different Statistical Disclosure Limitation Methods", en prensa en *Operations Research*, 2005.

4. ARTÍCULOS DE APLICACIÓN

ANÁLISIS ESTADÍSTICO DE LA ESTABILIDAD EN ASIENTOS DE USO DOMÉSTICO Y PÚBLICO

Mónica Alacreu¹ y Carmen Armero

Àrea d'Epidemiologia. Direcció General de Salut Pública.

Departament d'Estadística i Investigació Operativa. Universitat de València.

1. Introducción

La certificación de calidad de un producto tiene como objetivo garantizar al consumidor la resistencia y seguridad del artículo que va a adquirir. Además, desde el punto de vista comercial, le proporciona un signo de distinción que permite diferenciarlo de la competencia de otros en el mercado. En el sector del mueble, esta garantía se obtiene sometiendo al mobiliario a unos ensayos de acuerdo a su naturaleza y función según las especificaciones que determina la norma prevista para su uso. Estas pruebas se realizan en laboratorios especializados, que disponen de la acreditación y de la maquinaria adecuada para simular, en poco tiempo, las acciones que el mueble tendría que soportar durante varios años en su régimen normal de uso.

La Asociación de Investigación y Desarrollo en la Industria del Mueble y Afines (AIDIMA) es una entidad privada de ámbito nacional y sin ánimo de lucro, cuyo objetivo básico es lograr el progreso y desarrollo tecnológico del sector español del mueble e industrias afines, el incremento en la calidad de producción y el fortalecimiento de las exportaciones. Entre los diferentes servicios que oferta se encuentra el Laboratorio para el Mueble Acabado (LMA), que es un centro técnico donde se realizan los ensayos y certificaciones de las diversas familias de muebles. Una de las más comunes y en la que centramos este trabajo, es la sillería de uso doméstico y público, que abarca tanto las sillas (asientos sin reposabrazos) como los sillones (asientos con reposabrazos).

El análisis estadístico que presentamos en este trabajo es parte del realizado durante el disfrute de una beca de formación profesional práctica que la primera autora realizó en AIDIMA sobre la recopilación y tratamiento estadístico de resultados del LMA. Durante ese periodo analizamos los resultados obtenidos con la norma UNE EN 1022: 1998, que es una publicación que regula la metodología y las especificaciones a seguir en todos los ensayos que evalúan la estabilidad de los asientos de uso doméstico y público en España.

Este trabajo está estructurado en cuatro apartados. Tras esta introducción, explicaremos en el siguiente apartado las dificultades que tuvimos en el diseño y recopilación del banco de datos. Después, presentaremos una descripción básica de las variables que caracterizan la estabilidad de los asientos y finalizaremos aplicando técnicas básicas de supervivencia para analizar el comportamiento de la fuerza

aplicada para producir el “vuelco” de cada uno de los asientos considerados.

2. ¿De dónde salieron los datos?

Para garantizar un banco de datos de buena calidad, nuestra primera tarea fue estudiar qué familia de muebles se ensayaba más frecuentemente en el laboratorio. Una vez seleccionada la sillería de uso doméstico y público, tuvimos que elegir el periodo de tiempo en el que centraríamos el trabajo. Teniendo en cuenta que la documentación que estábamos consultando se guardaba en el archivo general y que éste había sufrido cambios en el almacenamiento, llegamos a la conclusión de que lo más sensato era trabajar con los datos de los años 2000, 2001 y 2002. Una vez seleccionado el mobiliario y el periodo de análisis, elegimos, de entre todos los ensayos que se realizan a los asientos, aquellos que evalúan su estabilidad. Posiblemente, la tarea más tediosa de este proceso fue la consulta del archivo, ya que éste carecía de versión informatizada y todo el trabajo de recopilación lo tuvimos que hacer a mano.

Finalmente, el banco de datos quedó formado por las observaciones tomadas a 121 asientos que registramos en 6 variables:

- Año: Es el año, 2000, 2001 ó 2002, en que fue ensayado el asiento.
- Uso: Se refiere al uso que, previsiblemente, va a recibir el asiento. Existen tres categorías: uso doméstico normal, doméstico severo o público normal, y público severo.
- Reposabrazos: Es una variable dicotómica que toma el valor 1 si el asiento lleva reposabrazos (es un sillón) y 0 si no lo lleva (porque es una silla).
- Vdelantero: El vuelco delantero es la fuerza, expresada en newtons (N), que resiste el asiento antes de volcar hacia delante. Para que un asiento sea considerado estable en esta prueba debe resistir al menos 20 N.
- Vtrasero: El vuelco trasero es la fuerza, en newtons, que resiste un asiento antes de que vuelque hacia atrás. En esta prueba un asiento se considera estable si resiste, como mínimo, una fuerza superior a $0.2857(1000-h)$ N, siendo h la altura desde la base del asiento hasta el suelo expresada en milímetros (mm).
- Vlateral: El vuelco lateral es la fuerza, en newtons, que debe resistir el asiento antes de volcar de lado. Un asiento es estable en esta prueba si resiste al menos 20 N.

Las tres últimas variables nos permiten evaluar la estabilidad de los asientos. Las tres primeras, Año, Uso y Reposabrazos, son factores que utilizaremos para intentar explicar y analizar la estabilidad de los distintos asientos. Debemos comentar que las cargas aplicadas para los vuelcos laterales son diferentes en las sillas y sillones, y por eso, tanto en la descripción estadística de los datos como en el análisis de supervivencia diferenciaremos entre ambos.

3. Descripción de los datos.

En este apartado se muestra una descripción básica de las tres variables que miden la estabilidad de los asientos en forma de tablas con los estadísticos habituales, histogramas para visualizar la forma en que se distribuyen los datos y diagramas de cajas para describir su comportamiento según los diferentes factores considerados. Nótese que, en todo este apartado, todas las gráficas mantienen la misma escala para que la comparación visual sea homogénea.

En la tabla 1 se presenta el tamaño muestral, mínimo, máximo, media y desviación típica de las variables que describen la estabilidad de los distintos asientos.

	n	Mínimo	Máximo	Media	Desv. Típica
Vdelantero	121	0	278	89.51	52.59
Vtrasero	120	107	493	222.29	60.01
Vlateral para sillas	66	22	183	75.61	35.84
Vlateral para sillones	55	10	250	85.56	60.67

Tabla 1: Estadísticos básicos de las variables que miden el tiempo registrado hasta que se produce un vuelco delantero, trasero y lateral.

Puede observarse que el mínimo de la variable Vdelantero es 0 N. Esto significa que algún asiento ha volcado hacia delante durante la preparación previa al ensayo, que consiste en cargar el asiento con 600 N simulando que hay una persona sentada. El número de observaciones de la variable Vtrasero es 120, ya

que hubo un asiento al que no se le practicó esta prueba. Es también interesante observar que la dispersión de los datos del vuelco lateral en sillones es casi el doble de la registrada en las sillas.

En la figura 1 se presenta un histograma de los datos de las variables que registran los vuelcos estudiados. En los histogramas correspondientes al Vdelantero y Vlateral, tanto para sillas como sillones, la primera barra corresponde al número de asientos que no han superado los 20 N, y que por lo tanto, no han sido considerados como estables. En este ensayo se han encontrado 4 asientos inestables respecto a la prueba del Vdelantero, todas las sillas han sido diagnosticadas como estables y 2 sillones como inestables respecto al Vlateral.

La distribución de los datos de la variable Vtrasero está más desplazada hacia la derecha que la de las otras variables y presenta una cierta simetría respecto de los valores alrededor de 200 N. Recordamos también que la estabilidad de un asiento en esta prueba depende del diseño del mueble, de forma que cuanto mayor es su altura, menos fuerza, paradójicamente, debe superar para ser estable. Por ejemplo, los asientos con una altura superior a 720 mm deben resistir una fuerza mínima de 80 N. En la figura 2 representamos la fuerza que ha resistido cada asiento antes de volcar hacia atrás y la fuerza que debía resistir, según su altura, para ser considerado estable. En este caso, son 4 los asientos que han sido considerados inestables. En la figura 3 representamos las variables Vdelantero, Vtrasero y Vlateral con respecto a los factores considerados. Podemos destacar que existe una mayor variabilidad de las observaciones en el grupo de los sillones que en el de las sillas, posiblemente porque en el grupo de los

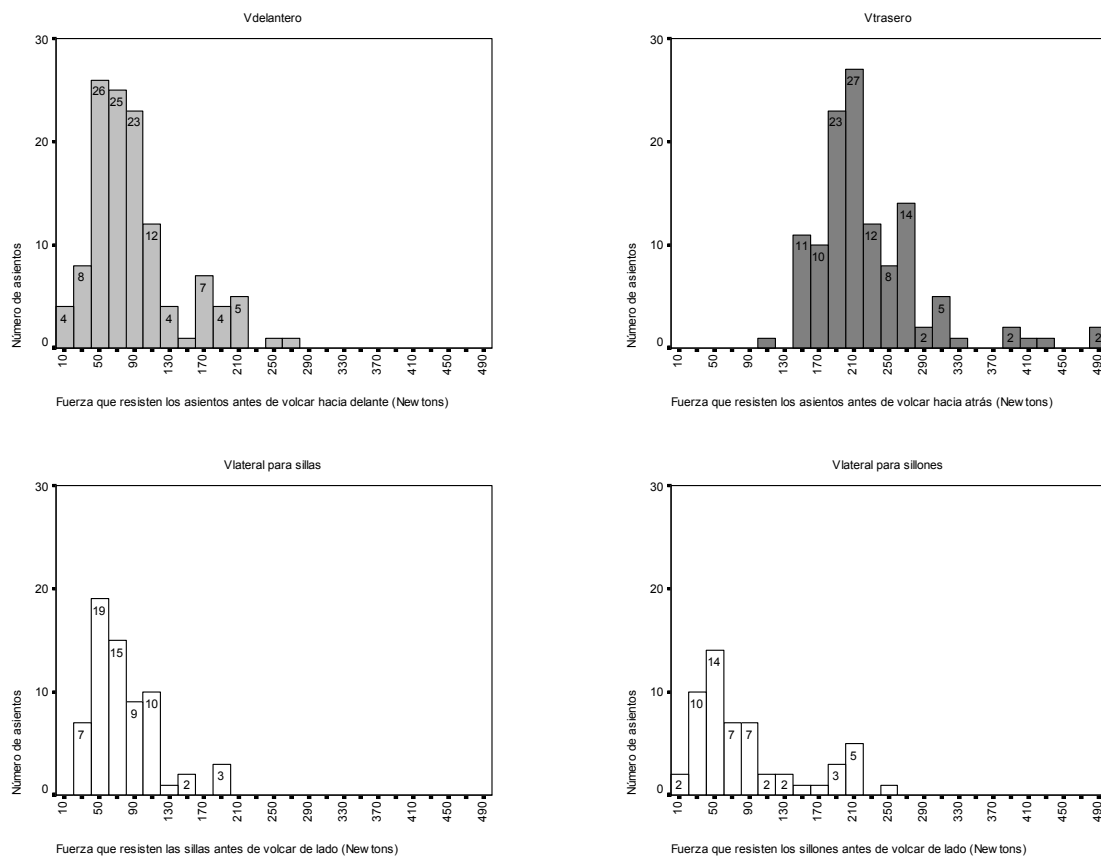


Figura 1: Histogramas de las variables que registran el vuelco delantero, trasero y lateral, éste último para sillones y sillas por separado.

sillones se consideran tanto sillones de comedor con un cierto tamaño, como sillones que se parecen a una silla, sólo que en su estructura están dotados de reposabrazos. Intuitivamente, al aplicar las mismas cargas a dos muebles tan dispares se van a obtener resultados muy diferentes. Por otra parte, las cajas que representan el vuelco delantero siguen una tendencia decreciente cuanto más severo sea el uso de asientos. Es decir, los asientos destinados a uso público severo tienen menos resistencia al vuelco.

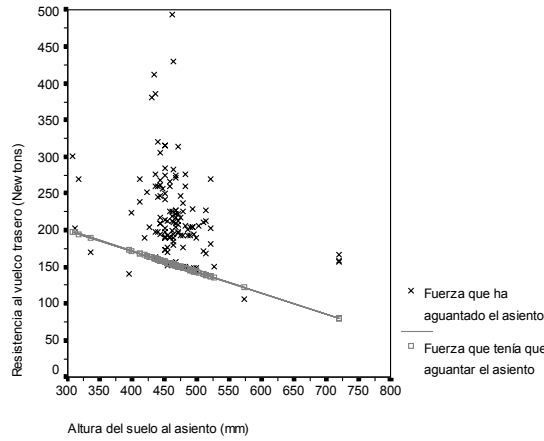


Figura 2: Resistencia al Vtrasero y fuerza necesaria para el cumplimiento de la estabilidad, respecto a la altura del asiento.

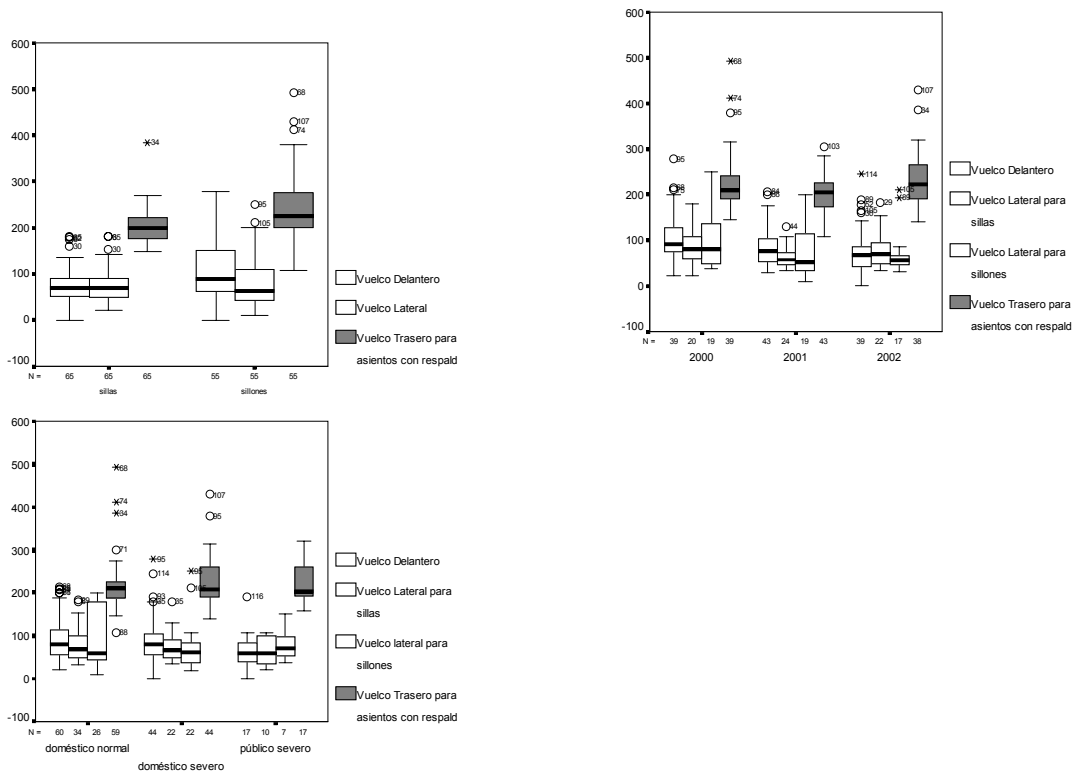


Figura 3: Diagramas de cajas que resumen las variables Vdelantero, Vtrasero y Vlateral, respecto los factores Reposabrazos, Uso y Año.

delantero que los asientos de uso doméstico normal. Sin embargo, las cajas que representan el vuelco trasero tienen una tendencia creciente con la severidad de su uso. Esto significa que los asientos de uso público severo son más resistentes al vuelco trasero que los asientos de uso doméstico normal. Por último, parece que con el paso de los años, los asientos se fabrican más resistentes al vuelco trasero y cada vez menos resistentes al vuelco delantero. Esta sería una línea de investigación a seguir que precisaría de datos de periodos de tiempo mucho más amplios.

4. Análisis de supervivencia.

El análisis de supervivencia es una técnica estadística poco habitual en temas de mobiliario.

Se utiliza para describir datos correspondientes al tiempo que transcurre desde un punto que marca el inicio del experimento hasta que se produce el evento de interés. Este es el caso, por ejemplo, del tiempo que pasa desde que se diagnostica una enfermedad terminal hasta que el paciente fallece. Sin embargo, estas técnicas son también útiles para describir datos que, aun preservando esta estructura, tienen un contexto de aplicación y un significado muy diferente. Por la naturaleza de las pruebas que evalúan la estabilidad de los asientos, podemos interpretar las fuerzas que se aplican a los asientos como magnitudes continuas que se administran hasta que se produce un evento. En este caso las fuerzas jugarían el papel del tiempo mencionado inicialmente y el evento sería el vuelco del asiento.

En la tabla 2, mostramos de forma básica y para cada una de las variables por separado, la estimación semi-paramétrica de la supervivencia con el método de Kaplan–Meier. Observamos que para volcar un asiento hacia delante son necesarios, en promedio, aproximadamente 89 N de fuerza, mientras que la fuerza media estimada, para volcar un asiento hacia atrás, es más de 222 N. En general, se observa que para volcar una silla de lado se necesita una media de 76 N, a su vez para volcar un sillón son necesarios alrededor de 86 N. Si miramos los percentiles, aplicando una fuerza de 48 N sobre el conjunto de las sillas volcamos de lado el 25% (sobreviven el 75%), sin embargo con la misma fuerza logramos volcar más del 25% de los sillones. Por otra parte, aplicando 94 N de fuerza fallaría el 75% de las sillas en el Vlateral y menos del 75% de los sillones. Esto significa que hay más variabilidad en la resistencia al vuelco lateral de los sillones que en la resistencia al vuelco lateral de las sillas. Hay sillones que son menos estables que “la peor” silla y los hay más estables que “la mejor” silla.

	Media	Error estándar	Intervalo de confianza al 95%	Percentiles		
				25	50	75
Vdelantero	89.51	4.78	(80.14, 98.88)	107	79	54
Vtrasero	222.29	5.48	(211.55, 233.03)	250	209	190
Vlateral para sillas	75.61	4.41	(66.96, 84.25)	94	69	48
Vlateral para sillones	85.56	8.18	(69.53, 101.60)	110	64	42

Tabla 2: Resumen de supervivencia para las variables Vdelantero, Vtrasero, Vlateral para sillas y Vlateral para sillones.

En la tabla 3 presentamos un resumen de la supervivencia de las variables que describen los diferentes vuelcos, considerando, como factor de agrupación, el uso previsible que deberían recibir los asientos.

	Uso	Media	Error Estándar	Intervalo de Confianza 95%	Percentiles		
					25	50	75
Vdelantero	doméstico normal	94.98	6.44	(82.36, 107.60)	112	81	55
	doméstico severo	92.27	8.60	(75.41, 109.14)	104	79	54
	público normal	63.06	10.68	(42.13, 83.99)	83	59	40
Vtrasero	doméstico normal	218	8.22	(202.00, 234.20)	227	210	189
	doméstico severo	226	9.15	(208.31, 244.19)	260	209	190
	público normal	226.59	11.68	(203.70, 249.48)	260	204	194
Vlateral para sillas	doméstico normal	79.50	6.60	(66.57, 92.43)	101	69	49
	doméstico severo	74.95	7.10	(61.03, 88.88)	90	65	49
	público normal	63.80	10.24	(43.73, 83.87)	101	48	34
Vlateral para sillones	doméstico normal	96.38	13.46	(70.01, 122.76)	180	53	44
	doméstico severo	74.59	11.91	(51.25, 97.93)	84	57	38
	público normal	79.86	15.12	(50.22, 109.50)	110	70	46

Tabla 3: Resumen de supervivencia para las variables Vdelantero, Vtrasero, Vlateral para sillas y Vlateral para sillones distinguiendo los diferentes niveles de uso.

El promedio de fuerza necesario para volcar hacia delante un asiento de uso doméstico normal es de casi 95 N, mientras que la fuerza necesaria en asientos de uso público severo es de 63 N. Los percentiles también muestran diferencias en cuanto al comportamiento de los tres usos. Aplicando 81 N de fuerza sobre los asientos de uso doméstico normal, conseguimos que vuelquen la mitad. Lo mismo pasa si aplicamos 79 N sobre los asientos de uso doméstico severo. Por último, con tan sólo 59 N de fuerza logramos volcar el 50% de los asientos de uso público severo.

Si nos fijamos en la variable Vtrasero, el comportamiento de la supervivencia para los distintos niveles de uso son bastante similares. El percentil 25 es el que más diferencias marca, ya que con 227 N de fuerza volcamos el 75% de los asientos de uso doméstico normal mientras que son necesarios 260 N para volcar el mismo porcentaje de asientos en los otros dos niveles.

Observando los resultados obtenidos para la variable Vlateral, vemos que para que vuelque lateralmente una silla de uso doméstico normal es necesaria una media de 79.5 N. El promedio para sillas de uso público severo es mucho menor, alrededor de 64 N. Si miramos los percentiles, podemos ver que con una fuerza de 49 N conseguimos volcar el 25% de las sillas de uso doméstico normal, el 25% de uso doméstico severo y más del 50% de las sillas de uso público severo. En el caso de los sillones, la fuerza media necesaria para volcar un sillón lateralmente ha aumentado considerablemente con respecto a la media en las sillas. Es decir, para volcar lateralmente un sillón se necesita en general más fuerza que para volcar una silla. Con 53 N de fuerza vuelca el 50% de los sillones de uso doméstico normal, el 46% de los sillones de uso doméstico severo y, aproximadamente, el 30% de los sillones de uso público severo.

La tabla 4 recoge los resultados de tres pruebas (test Log-Rank, Test de Breslow y test de Tarone-Ware) que valoran las posibles diferencias en la supervivencia de los asientos en relación a sus distintos usos. Cuando, como consecuencia de la estratificación de un factor, comparamos varias curvas de supervivencia que se cruzan, es conveniente evitar la prueba Log Rank cuando las mayores discrepancias se encuentran al inicio de las curvas.

	Test Log Rank		Test Breslow		Test Tarone-Ware	
	estadístico	p-valor	estadístico	p-valor	estadístico	p-valor
Vdelantero	5.65	0.0594	7.43	0.0244	6.86	0.0323
Vtrasero	0.64	0.7256	0.82	0.6651	0.92	0.6328
Vlateral	1.63	0.4428	1.00	0.6064	1.27	0.5304

Tabla 4: Test de comparación de las distribuciones de supervivencia de los distintos usos, en cada una de las variables vuelco.

En el caso de la variable Vdelantero, la prueba más adecuada es la de Tarone - Ware. Como las gráficas de las tres curvas de supervivencia se cruzan (figura 4), el test Log Rank no es capaz de detectar diferencias entre los distintos grupos, y sin embargo tanto la prueba de Breslow como la de Tarone - Ware sí que lo son. Podemos afirmar que, al menos, dos de los usos presentan diferencias significativas, en la variable Vdelantero. Para las variables Vtrasero y Vlateral, las tres pruebas anteriores no son capaces de detectar diferencias significativas en la supervivencia entre los asientos de uso doméstico normal, doméstico severo y público severo.

Para cada uno de los vuelcos estudiados, la figura 4 muestra la función de supervivencia obtenida con el método de Kaplan-Meier y, a su derecha, la estratificación de cada gráfica según su previsible uso. Como vemos en la primera gráfica, la variable Vdelantero tiene al principio la pendiente muy acentuada y después se atenúa, es decir al principio, aumentando poco la fuerza aplicada a los asientos, vuelcan hacia delante más rápidamente. Así, aplicando 100 N vemos que han caído, aproximadamente, el 70 % de los asientos. Sin embargo, si aplicamos 100 N para provocar el vuelco trasero, no falla ningún asiento. Por otra parte, vemos que la gráfica del Vlateral para sillones es la que tiene la pendiente más pronunciada. Aplicando entre 20 N y 100 N de fuerza, vuelcan de lado el 80 % de las sillas. Por último, en la última gráfica estratificada, vemos diferencias entre los distintos usos, para los sillones que vuelcan con más de 80 N.

No queda bien esta página, habría que ajustarlo mejor. Quizá poniendo este último párrafo detrás de la figura

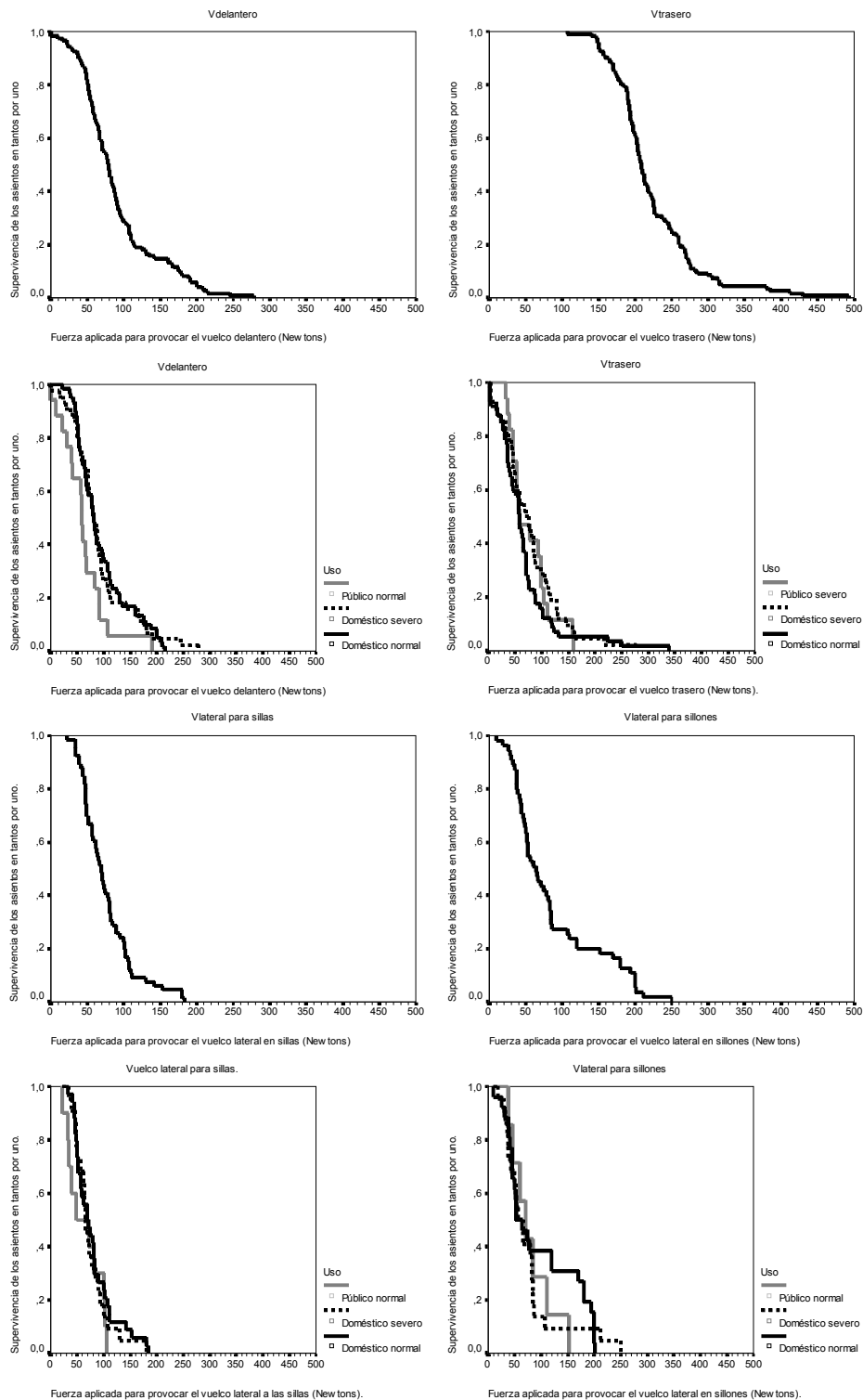


Figura 4: Funciones de supervivencia de las variables Vdelantero, Vtrasero y Vlateral para sillas y sillones. Funciones de supervivencia distinguiendo entre los diferentes usos.

5. Referencias.

Pérez-Hoyos S. Análisis de Supervivencia. Quaderns de salut pública i administració de serveis de Salut, 11. Valencia: Institut Valencià d'Estudis en Salut Pública, 1997.
 Bermúdez José D. Técnicas Estadísticas en el Análisis de Supervivencia. Departamento de Estadística e Investigación Operativa. Universitat de València, 1997. Disponible en <http://www.uv.es/~bermudez/>.

6. ESTADÍSTICA OFICIAL

LA ESTIMACIÓN EN ÁREAS PEQUEÑAS PARA LA ESTADÍSTICA OFICIAL

Montserrat Herrador, Jorge Saralegui
Instituto Nacional de Estadística

Resumen

La necesidad de disponer sistemáticamente de información estadística para dominios pequeños, se ha venido consolidando en los últimos años entre los objetivos de los sistemas de estadísticas oficiales. El artículo aborda los problemas ligados a la utilización de estimadores asistidos por modelos como forma de superar la limitación de la escasez de muestra en operaciones estadísticas cuya información tiene interés en pequeños dominios, con especial atención a los aspectos de calidad estadística, como la relevancia, transparencia y fiabilidad de los datos oficiales a difundir.

1. Introducción

La demanda de estadísticas oficiales con un gran detalle en la desagregación, tanto en el campo de la estadística económica como en el de la estadística social y laboral, no deja de crecer. En consecuencia, la necesidad de disponer sistemáticamente de datos publicados para dominios pequeños, se ha venido consolidando en los últimos años entre los objetivos de los sistemas de estadística oficiales, al mismo tiempo que se han venido desarrollando diversas líneas de investigación sobre la utilización de estimadores asistidos o basados en modelos, como forma de superar la limitación de la escasez de muestra en las operaciones estadísticas cuya información tiene interés en pequeños dominios.

Los problemas de la estimación en pequeños dominios surgen por el aumento en los costes y en la complejidad de los diseños muestrales que aspiren a alcanzar cotas aceptables de calidad de las estimaciones en todas las áreas o dominios de interés para los usuarios, lo cual, a su vez, puede tener consecuencias negativas en la calidad de las estimaciones para los dominios superiores previstos inicialmente en el proyecto de encuesta. Los límites, por razones de coste, para la ampliación sin restricciones de los tamaños muestrales en todos los dominios de interés, deben ser interpretados en un sentido amplio: bajo el punto de vista de la recogida y producción de datos, claro está, pero también bajo el punto de vista de la carga de respuesta a las unidades a contactar en la encuesta. El aumento de tamaños muestrales para mejorar la eficiencia en dominios pequeños debe tener en cuenta ambos costes, sin olvidar otras pérdidas de calidad debidas a los plazos de obtención de resultados y al impacto de determinados errores ajenos al muestreo (falta de respuesta, errores de medida, efectos entrevistador, etc.) de consecuencias más negativas cuanto mayores son los tamaños muestrales.

Tras hacer una exposición en las dos primeras secciones del problema general de la estimación en áreas o dominios pequeños, la ponencia continúa en la sección 3 con la situación en particular en nuestro país. La sección siguiente analiza la casuística ligada a la utilización de modelos en la estadística oficial, para pasar en las secciones 5 y 6 a una presentación de las principales clases de estimadores de áreas pequeñas actualmente en estudio para su aplicación a las encuestas del Instituto Nacional de Estadística (INE), y a los problemas relacionados con la calidad esperada de estas nuevas técnicas de producción de información. El artículo finaliza en la sección 7 con un análisis del papel de una oficina central de estadística en este campo.

2. Los dominios de estimación

Los dominios de estimación en una encuesta por muestreo son aquellos subconjuntos del ámbito poblacional para los que es necesario producir, bajo determinados requisitos de calidad, estimaciones de todas o algunas de las variables objetivo. Son por lo tanto unidades de análisis de los resultados de la encuesta, pero pueden o no ser también unidades de observación.

El estimador directo de diseño para el total de una característica en un dominio d suele tomar la forma general del estimador de Horwitz –Thomson; es decir,

$$\hat{Y}_{d\pi} = \sum_{S_d} \frac{y_k}{\pi_k} \text{ si } N_d \text{ es desconocido; } \hat{Y}_{d\pi} = \sum_{S_d} \frac{N_d}{\hat{N}_d} \frac{y_k}{\pi_k} \text{ si } N_d \text{ es conocido;}$$

donde S_d es la muestra en el dominio d , $\hat{N}_d = \sum_{S_d} \frac{1}{\pi_k}$ el estimador del total poblacional y π_k es la probabilidad de que la unidad k pertenezca a la muestra.

En este tipo de estimadores, los valores muestrales de la característica y_k aparecen multiplicados por un ‘peso inicial’ o ‘peso de diseño’ que en fases posteriores es corregido, intentando mantener la forma lineal del estimador, por los efectos de cambio de estrato, incidencias de marco, falta de respuesta, equilibrado a poblaciones externas, u otros, de acuerdo a modelos implícitos más o menos complejos. El estimador directo de la media en el dominio queda así de la forma más general

$$\hat{Y}_d^{direct} = \sum_{S_d} \frac{y_k w_k}{\sum_{S_d} w_k}, \text{ donde } w_k \text{ incorpora los diversos ajustes.}$$

No obstante, este tipo de ‘modelos’, por razones históricas difíciles de asimilar, no suelen ser considerados como tales en la literatura, y paradójicamente todavía es frecuente hoy en día encontrar polémicas académicas sobre la ventaja o no de utilizar modelos en la estimación del muestreo de poblaciones finitas, cuando, como queda dicho, en la mayoría de los proyectos, los modelos están presentes en diversas fases de la elaboración de resultados.

El pequeño dominio puede tener o no una dimensión territorial, pero en cualquier caso el problema se presenta cuando S_d es muy pequeño o nulo. Independientemente del procedimiento de estimación directa previsto, el pequeño dominio puede aparecer tanto en encuestas económicas como en encuestas de hogares, aunque en este caso los avances de la teoría hayan sido muy superiores. Baste citar como ejemplos del problema la gran presión de las peticiones dirigidas al INE para suministrar datos estructurales (o incluso coyunturales) de una encuesta económica a nivel de 4 ó 5 dígitos de la Clasificación de Actividades Económicas (CNAE), desagregados según la tipología del tamaño de las empresas, con lo que el estimador directo debe ser calculado con tamaños muestrales muy pequeños (pudiendo ser nulos si se introduce, además, como es frecuente, la dimensión territorial). En encuestas de hogares es más común el caso del pequeño dominio definido por la dimensión territorial exclusivamente, siendo cada vez más intensa la demanda de información (tasa de paro, índice de pobreza relativa, gasto per cápita) por pequeños territorios, como la comarca, la isla, el distrito, etc.

En España, prácticamente todos los diseños de encuesta contemplan la Comunidad Autónoma -NUT II, unidad estadística territorial de nivel II en terminología del sistema estadístico europeo- como dominio de observación y análisis, por lo que el problema de estimación en áreas pequeñas aparece a partir del nivel NUT III (provincia), aunque no en todos los casos. Por ejemplo, la Encuesta de Población Activa (EPA) publica datos por estimación directa por provincias. Sin embargo, la Encuesta de Presupuestos Familiares (EPF), sólo publica algunos indicadores de gasto y equipamiento por comunidades autónomas. En este último caso, para la publicación de datos provinciales se recomienda la acumulación de muestras de varios años/trimestres, procedimiento válido en ausencia de otros modelos que demuestren mayor robustez. Llama la atención el escaso recurso, por parte de los analistas y usuarios en España, a la utilización de muestras acumuladas de varios periodos para obtener estimaciones (en este caso provinciales), tanto en encuestas de hogares como en las estructurales de empresas.

Pero no hay que olvidar que las necesidades de información en áreas pequeñas territoriales en España, van más allá de la provincia. El INE está recogiendo en la actualidad, a través de las diversas autoridades con competencia en la materia, la situación administrativa del nivel NUT IV (“comarca” en España) con fines estadísticos Y, en este contexto, adquiere mayor urgencia la disponibilidad de métodos de estimación “poco exigentes” en tamaños muestrales para el nivel NUT IV.

El problema de las áreas pequeñas se puede presentar también ante demandas de información sobre subdominios ad hoc, como por ejemplo una cuenca (o parte de ella) hidrográfica, un cinturón industrial, una zona fronteriza ..., por lo que los modelos disponibles deben contemplar también el hecho de la demanda estadística para dominios que no han sido tenidos en cuenta a priori en el diseño.

3. Usuarios y aplicaciones

Por principios de calidad, todo esfuerzo de innovación metodológica debe tener en cuenta la tipología y expectativas de los usuarios efectivos y potenciales. Destacan, en primer lugar, como demandantes de las estadísticas en dominios pequeños (con al menos una dimensión territorial), los centros de decisión de la propia administración local competente en el “pequeño territorio”, así como los responsables de políticas más generales (CCAA, Estado) pero de aplicación en el ámbito local. Estas políticas cubren un amplio espectro, como las relativas a educación y formación, vivienda, salud y dependencia, lucha contra la exclusión social, medio

ambiente, desarrollo sostenible, detección y catalogación de “ecozonas”, cubriendo diversos campos de acción política y de preocupación social.

Dependiendo de cada país y circunstancia, son unas u otras necesidades de información las que mayor peso aportan a la demanda inicial de estadísticas en áreas pequeñas. En el caso español, probablemente sea la iniciativa de las autoridades regionales (comunidades autónomas) la que mayor impulso está aportando, en paralelo al desarrollo de las competencias autonómicas generales.

Una parte de la actividad estadística orientada a atender esta demanda, sin embargo, no entra dentro de la especificidad de los modelos y estimadores en áreas pequeñas, sino en el campo más amplio de las “estadísticas para áreas pequeñas”, de gran desarrollo también en nuestros días. Se trata de bancos de datos, asociados o no a sistemas de información geográfica (SIG), en los que se acumula transversal y longitudinalmente una gran masa de información y metainformación de fuentes diversas, incluidos indicadores simples y sintéticos, relativos a cada pequeña área, con un esquema de mantenimiento que puede responder a un modelo teórico más o menos desarrollado, como se puede observar en la gran diversidad de sistemas de indicadores disponibles. Por ejemplo, el sistema de información de este tipo que mantiene el INE, del que se extraen periódicamente informes publicados (Indicadores Sociales de España), integra también una dimensión territorial, que llega hasta el nivel municipal en algunos indicadores básicos.

Precisamente, los métodos de estimación en áreas pequeñas pueden servir para abastecer una pequeña parte de las necesidades de información de estos sistemas o, a su vez, ser receptores de una parte de información de estos bancos de datos de ámbito local, para ser utilizada como abastecedora de variables explicativas en los modelos subyacentes a los estimadores. Afortunadamente, la componente académica y de investigación universitaria, al menos a nivel internacional, aunque de manera muy incipiente en España, ha tenido –o está teniendo- un papel muy importante en el desarrollo y uso de estos modelos, lo que se percibe en la bibliografía que existe sobre el tema y que no deja de crecer día a día. Y no sólo en el campo de la estadística, sino también en otras disciplinas como las ciencias de la salud, de la educación, la sociología o la geografía del desarrollo.

4. Algunas limitaciones a la aplicación de modelos en la Estadística Oficial

Antes de aplicar sistemáticamente, aunque sea de forma parcial, los métodos de estimación asistida o basada en modelos es necesario ser muy conscientes de los riesgos y limitaciones que afectan a estas técnicas, lo cual explica en parte su relativamente lenta (y escasa) implantación en la estadística oficial, y no solo en la española.

En primer lugar, nos encontramos con el problema de la calidad y de la disponibilidad de fuentes externas que proporcionen las variables auxiliares que explican los modelos. Una fuente externa, de carácter exhaustivo, como un censo o un registro administrativo, puede muy bien alcanzar cotas de calidad aceptables para sus propios fines, pero ser inservible en la aplicación de un modelo, aunque en éste intervengan variables aparentemente análogas en la encuesta y en la fuente externa. En el caso de los censos, normalmente el principal problema es su larga periodicidad, sin olvidar algunos problemas de conceptualización en variables de uso posible en los modelos (ej.: situación de actividad económica de la persona). En los registros administrativos, como el registro de demandantes del INEM, por seguir con el ejemplo, de gran potencialidad a priori para explicar la variable estadística relativa a la tasa de desempleo, a la hora de su implantación pueden presentar inconvenientes - de cobertura e inestabilidad de la definición administrativa, principalmente- de cierto impacto en el rendimiento del modelo.

En el caso de los registros de población, entre los que podemos considerar al Padrón Continuo español, algunos fenómenos de importancia creciente, como la doble residencia o la formación de hogares complejos, disminuyen la propiedad explicativa de variables tan esenciales como la distribución de población por sexo y edad en un área pequeña. En el ámbito económico, una fuente exhaustiva como el Directorio Central de Empresas del INE, puede presentar limitaciones metodológicas graves a la hora de producir agregados para pequeños territorios, entre otras cosas por lo problemático de distribuir territorialmente la actividad de las empresas multilocalizadas o con vínculos complejos de unidades legales de actividad económica.

La utilización de modelos en estadística oficial debe hacer frente también a los problemas de falta de transparencia para determinados grupos de usuarios, como los medios de comunicación o algunos centros de decisión política, más acostumbrados a la estadística de origen administrativo u obtenida por estimación directa. Y ello en el contexto paradójico de la presencia habitual de modelos subyacentes en los estimadores finales de casi todas las encuestas, como se ha comentado anteriormente. Algunos modelos presentan cierta complejidad para la estimación de parámetros y sus varianzas, con software avanzado, apto solamente para personas expertas,

por lo que su uso e interpretación tiene asociados ciertos costes no presentes en la difusión de otras estadísticas.

Las grandes series de datos oficiales, con abundancia de interrelaciones multivariantes, que deben ser consistentes con los valores marginales, no se adaptan bien a la aplicación de estimadores basados en modelos para pequeños dominios, cuya consistencia a niveles más agregados, o con dimensión temporal, no siempre está garantizada, por lo que estos estimadores deben ser producidos conjuntamente con una metainformación y detalle metodológico difícil de asimilar por algunos usuarios.

En otro orden de cosas, especialmente en las estadísticas de empresas, el escaso tamaño muestral de algunas subpoblaciones impone un límite inferior al tamaño del dominio de estimación, por razones de confidencialidad. Algunos usuarios, por otra parte, buenos conocedores de un área pequeña, pueden no ser muy comprensivos con una estimación anómala en un dominio por ellos tan bien conocido, por muy “correcto” que sea el método de estimación empleado, lo cual contribuye a la preocupación por el riesgo de aparición de sesgos incontrolados, que afecta potencialmente a los estimadores basados en modelos.

5. Variables, modelos, estimadores

En los modelos de áreas pequeñas, las variables explicativas responden a dos tipologías muy distintas que condicionan el modelo: las que llamaremos aquí variables X , disponibles tanto en la observación muestral directa como en la fuente externa, a nivel de área pequeña, o las variables que llamaremos Z , o variables de área, disponibles solamente en los agregados procedentes de la fuente externa, pero que no son recogidas en la encuesta sobre la que se estimarán los parámetros del modelo. Un ejemplo clásico de variable X es la inscripción o no en la Oficina de Empleo, que se pregunta en la Encuesta de Población Activa a todas las personas mayores de 15 años, y que está siendo considerada en los modelos que están siendo investigados por el INE para la estimación en áreas pequeñas de la tasa de desempleo. Su correspondiente agregado por áreas, procedente de las estadísticas administrativas del INEM es el complemento necesario para la aplicación del estimador. Un ejemplo de variable Z en estudio para ser aplicada en modelos por el INE, es la base imponible del IRPF per cápita en el área, en nivel y en estructura según el origen de la renta (salarios, pensiones, desempleo, rentas del capital...). Para modelos de encuestas económicas, el volumen de negocio total de una actividad en el área, disponible en algunas fuentes externas, es un candidato firme a variable auxiliar (de área o de unidad elemental, según el tipo de encuesta).

En ocasiones, una variable tipo X puede dar mayor rendimiento en los modelos si se considera como Z , debido a que los errores de observación u otras diferencias conceptuales pueden hacer que la característica recogida en la encuesta difiera sustancialmente de la que constituye el agregado de la fuente externa. La decisión de considerar una variable explicativa presente en la encuesta como de tipo X o de tipo Z , es de gran trascendencia para la potencia explicativa de la componente sintética del modelo. Según sea el tipo de variables de las que disponemos en el mundo real de la encuesta y de las fuentes externas que proporcionan información para cada pequeña área, tendremos los diferentes tipos de modelos en los que se basan los estimadores frecuentistas.

Los *modelos de unidad elemental de dos niveles*, con efecto de áreas (índice d) y efecto de individuo (índice i) para explicar una variable y observada en la unidad elemental i del área d , pueden incluir tanto variables del tipo X como del tipo Z (por eso también se denominan ‘modelos mixtos’) según la casuística expuesta anteriormente:

$$y_{id} = Z_d^T \eta + x_{id}^T \beta + u_d + e_{id} \quad \text{con } u_d \approx iid \text{ N}(0, \sigma_u^2); \quad e_{id} \approx iid \text{ N}(0, \sigma_e^2)$$

Sin embargo los modelos de área contemplan solamente el efecto del área:

$$\bar{Y}_d = \bar{Z}_d^T \eta + u_d \quad \text{que se complementa con el modelo para el estimador } \hat{Y}_d = \bar{Y}_d + e_d;$$

$$\text{con } u_d \approx iid \text{ N}(0, \sigma_u^2); \quad e_{id} \approx iid \text{ N}(0, \psi_d);$$

En el ejemplo de la fórmula, el modelo explica el valor medio de la variable objetivo en el área, a partir de las variables auxiliares tipo Z , agregados de área.

Un concepto importante en la elección del modelo es el de grupo de regresión, o conjunto poblacional que engloba a las áreas pequeñas objetivo, y con cuya muestra se estiman los parámetros de los modelos (por ejemplo, una NUT II puede constituir el grupo de regresión para estimar parámetros a aplicar en estimadores al nivel NUT IV).

A la vista de los indicadores de evaluación, si su rendimiento se considera adecuado para estimar los parámetros, de estos modelos se derivan los estimadores sintéticos, como el *sintético de regresión*

$$\hat{Y}_d^{rsyn} = \bar{X}_d^T \hat{\beta} + \bar{Z}_d^T \hat{\eta}$$

Un caso particular de los sintéticos de regresión es el sintético básico, en el que la estructura de una variable discreta en el dominio superior sirve para explicar la estructura en el pequeño dominio incluida en aquella

$$\hat{Y}_d^{syn} = \frac{1}{N_d} \sum_g N_{gd} \hat{Y}_g^{direct}$$

con lo que el estimador en el área se obtiene a partir de las estimaciones de los valores medios en un dominio superior dentro de los grupos g de una variable explicativa discreta (ej. clase de actividad, grupos de edad) cuyos subtotales poblacionales están disponibles también en el área pequeña, a partir de un registro actualizado, un Padrón o un Censo reciente.

Una clase de estimadores muy utilizados para áreas pequeñas es la de los *EBLUP* (*empirical best linear unbiased predictor*)

$$\hat{Y}_d^{eblup} = \lambda_d \hat{Y}_d^{greg} + (1 - \lambda_d) \bar{Y}_d^{rsyn} \quad (\text{con modelo de unidad elemental});$$

En el que interviene el estimador GREG, que se expone más adelante, entre los estimadores asistidos por el modelo. O bien el

$$\hat{Y}_d^{eblup} = \lambda_d \hat{Y}_d^{direct} + (1 - \lambda_d) \bar{Y}_d^{rsyn}, \quad \lambda_d = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\phi}_d} \quad (\text{con modelo de área}).$$

Estos estimadores combinan una componente directa o asistida por modelos, con el fin de asegurar que no se produzcan aumentos sustanciales de sesgo, con la componente sintética, para incorporar la ganancia en precisión debida a la aportación del modelo, cuya potencia explicativa se utiliza para disminuir la variabilidad del estimador combinado. El factor λ_d tiene en cuenta las dos fuentes de variabilidad contempladas en el modelo, e interviene en este tipo de estimadores combinados ponderando cada componente del estimador inversamente a cada aportación a las varianzas σ_u^2 y ϕ_d .

Los modelos no frecuentistas (bayesianos) para estimación en áreas pequeñas, disponibles ya en la literatura, no están siendo objeto de consideración para su desarrollo a corto plazo, aunque es previsible que se vaya a producir un avance en su utilización en los próximos años, también en el campo de la estimación en dominios pequeños.

Los modelos EBLUP parecen en principio los mejores candidatos a consolidarse para la obtención de estimadores de variables de encuestas en las que existe riesgo de que abunden las áreas con escasa muestra para el estimador directo. Un ejemplo serían los indicadores de ingresos o del gasto per capita procedentes del Panel de Hogares o de la EPF (muestras nacionales entre 10 ó 15.000 hogares) a nivel de comarcas (NUT IV) o incluso provincias (NUT III). Sin embargo, para parámetros de encuestas con muestra generosa en un área pequeña (como la EPA), incluso a nivel NUT IV se puede pensar en la eficiencia de estimadores basados en el diseño del tipo *GREG*; es decir,

$$\hat{Y}_d^{greg} = \hat{Y}_d^{direct} + (\bar{X}_d - \bar{X}_d^{direct})^T \hat{\beta}$$

Un caso particular, muy utilizado, de estos estimadores es el postestratificado dentro de área, cuando las variables explicativas son dicotómicas de pertenencia o no a un grupo g de clasificación de una variable discreta, del que se dispone de subtotales en el área (ej. nivel de estudios, edad, etc.)

$$\hat{Y}_d^{pps} = \frac{1}{N_d} \sum_g N_{gd} \hat{Y}_{gd}^{direct}$$

Éstos y otros estimadores (incluidas algunas versiones de la clase SPREE), para diversas opciones y encuestas, han sido investigados en el proyecto internacional de investigación EURAREA:

http://www.statistics.gov.uk/methods_quality/eurarea/

6. Efectos de calidad estadística de la componente sintética del estimador

Los estimadores sintéticos, o la componente sintética de un estimador combinado, tratan de “ganar fuerza” respecto al estimador directo, a partir de fuentes externas, de referencia coetánea o no, y de las estimaciones directas de conjuntos poblacionales ‘próximos’ (en el tiempo, en el espacio o en la analogía frente al fenómeno en estudio). Cuanto mayor sea la variabilidad que explique el modelo subyacente al estimador sintético, mayor reducción de varianza del estimador entre áreas. Es decir, el modelo sintético aporta una parte sustancial de “muestra efectiva” al estimador directo, aumentando ‘virtualmente’ el tamaño muestral en el área. Por ello, este tipo de estimadores es muy sensible a la omisión o a la mala calidad de las variables auxiliares que intervienen en el modelo. En cuanto al sesgo, estos estimadores, condicionados al modelo, son aproximadamente insesgados, pero sin esa condición son sesgados, produciendo sobreestimación en algunas áreas y subestimación en otras, con un error medio entre áreas cercano a cero.

La componente sintética de estos estimadores plantea así la disyuntiva clásica en muestreo relativa a la ganancia en eficiencia aportada en principio por el modelo, pero a cambio de la aparición de sesgos en el área pequeña. Este comportamiento da origen a dos de los aspectos más difíciles de transmitir con transparencia al usuario de estadísticas oficiales que utilicen modelos de este tipo: el suavizado del estimador sintético y el riesgo (mayor que en la estimación directa) de estimaciones outlier o anómalas. En efecto, el estimador sintético, por su propia definición, introduce cierto suavizado sobre la distribución en el muestreo del estimador directo en el área. Basta pensar en un estimador sintético básico que imputa a una comarca la media de la variable objetivo en la comunidad autónoma que la contiene; el riesgo de “suavizado” excesivo viene dado por una excesiva simplificación –omisión– en el número de variables que “explican” la especificidad del área dentro de la comunidad autónoma.

El riesgo de outliers presente en cualquier estimador, puede adquirir especial dimensión en los modelos sintéticos de pequeñas áreas, por la aparición de residuos grandes en la regresión. Determinados grupos de usuarios de estas estadísticas son especialmente vulnerables ante estos errores, que deben ser cuidadosamente controlados. Es el caso de aquellas áreas pequeñas a las que se dan valores inusuales de la variable objetivo de Y , pero que sin embargo presentan valor X o Z dentro de lo esperado en el dominio superior o en los dominios próximos. Por ejemplo, la estimación en un área de gran densidad, con residencias aparentemente secundarias (aunque de hecho se trate de principales), cuando se utilicen como variables explicativas del modelo las procedentes de un registro de población que no recoja bien estos fenómenos.

Un procedimiento, ya clásico, para evaluar un estimador basado en modelos es su comparación con el estimador directo en dominios con muestra suficiente, utilizando métodos de simulación bootstrap o con poblaciones artificiales. Entre los requisitos mínimos para evaluación que se imponen de manera natural en la elección de variables auxiliares y modelos para estimación en áreas pequeñas se encuentra el de que las estimaciones basadas en modelos deben ser consistentes respecto a estimadores insesgados directos, cuando haya muestra suficiente en el área. Además, es lógico aspirar a que un estimador combinado elimine una parte importante de la variabilidad entre áreas explicada por el modelo, por lo que es exigible que su error cuadrático medio sea sustancialmente menor que la variabilidad total del estimador directo en el área:

$$ECM\left(\hat{Y}_d^{com}\right) \ll ECM\left(\hat{Y}_d^{direct}\right)$$

A su vez, se deberá evaluar hasta qué punto la posible aportación al sesgo de la componente sintética del estimador combinado será compensada por la componente ‘directa’. También se suele considerar como criterio de evaluación de la mejora que introduce el estimador con componente sintética, por razones obvias, la aportación o no de una mayor estabilidad temporal del estimador combinado del cambio, respecto a la que proporcionan las estimaciones directas.

7. El papel de una Oficina Central de Estadística en la estimación de áreas pequeñas

La demanda de estadísticas oficiales con un elevado detalle en la estimación en pequeños dominios plantea importantes disyuntivas en el proceso de decisión relativo a la producción/difusión sistemática de estimaciones en esos ámbitos, basadas o asistidas por modelos. Son varias las líneas de actuación necesarias para abordar esta demanda del sistema con criterios de calidad, algunas de ellas ya vigentes en los planes corrientes de los principales sistemas estadísticos internacionales y, en particular, en la Unión Europea y en España, por lo que es previsible que en el medio plazo, si se consiguen superar algunas incertidumbres –o si se encuentra satisfacción y estímulo en el otro lado de la balanza– veamos aparecer paulatinamente los estimadores para áreas pequeñas

basados en modelos, en el ámbito de la UE y, en concreto, en nuestro país.

En primer lugar, la aportación de la estadística oficial a la componente I+D de estimadores en áreas pequeñas debe ser suficientemente decidida e intensa, especialmente por la posibilidad de abastecer a la investigación con los datos primarios de encuestas, poblaciones reales o artificiales, esenciales para la evaluación de modelos y del rendimiento de los estimadores. Los modelos que contienen tiempo-espacio, la estimación de varianzas, el problema de la afijación óptima y la evaluación de modelos con diseños complejos o la consistencia de marginales en la estimación en áreas pequeñas, son campos en los que la teoría o la experimentación tienen todavía mucho que aportar, por lo que es de esperar que el camino ya iniciado tenga continuidad en los próximos años.

Es necesario también que la generalización de la estimación en áreas pequeñas en un país como España, en las que las administraciones territoriales están interesadas ó tienen competencias para producir estadísticas oficiales, se produzca sin despilfarro de recursos o pérdidas de calidad potencial. Cuando se trata de aplicar distintos modelos en distintos territorios no se debe olvidar que algunos modelos ganan fuerza cuanto mayor es el “grupo de regresión” donde se estiman los parámetros del modelo, por lo sería absurdo una auto limitación a priori de los objetivos de calidad. La incorporación de las mejores prácticas internacionales y la armonización de metodologías a nivel nacional son también parte importante del papel a jugar por una oficina central de estadística en este campo.

Un aspecto muy importante para analistas, productores eventuales de estimadores para áreas pequeñas, es la incorporación en las primeras fases del proyecto estadístico de una encuesta por muestreo del objetivo de estimación en pequeños dominios. Un ejemplo sería la afijación por comarcas, NUT IV, en el rediseño de la EPA aplicado en el primer trimestre de este año, a fin de disponer de una suficiente componente “directa” para estimadores combinados, acotando en lo posible el riesgo de sesgos incontrolados. Este tipo de afijaciones ya se han aplicado, por ejemplo, en la EPF por provincias, con el fin de que las estimaciones plurianuales a ese nivel (una forma de “estimador sintético” al fin y al cabo, ganando fuerza en el tiempo) permitan estimar la estructura del consumo para la base del IPC.

Precisamente, tiene mucho interés evaluar el impacto en la eficiencia general del diseño cuando se consideran las áreas pequeñas de una determinada tipología (zonas, distritos, subgrupos de empresas de determinada actividad, hogares de ciertas categorías socioeconómicas, etc.) como unidades de observación o “dominios a priori” en el proceso de afijación/selección de la muestra. Esta evaluación, incorporada en la fase inicial de un proyecto del órgano estadístico, constituye una aportación más de calidad al producto final.

Por supuesto, un papel importantísimo de los órganos centrales del sistema de estadísticas oficiales es asegurar la calidad de las fuentes administrativas potenciales suministradoras de las variables auxiliares de los modelos, alma mater de la eficiencia de los estimadores sintéticos. Todo esfuerzo en este sentido frente a las administraciones responsables del mantenimiento de las bases de datos o registros de educación, salud, prestaciones sociales, mercado de trabajo, renta, actividad económica, etc., es fundamental. La asimilación por parte de las distintas administraciones de la conveniencia de alcanzar una buena calidad del dato primario administrativo en áreas pequeñas, es básica en el proceso.

Es necesario potenciar la calidad de la información geo-referenciada a muy bajo nivel de detalle a fin de posibilitar la utilización de las bases de datos físicos en combinación con otros cualitativos y cuantitativos: sin ir más lejos, la presencia de una coordenada UTM en las bases de datos de la Seguridad Social y de la Agencia Tributaria serían de gran utilidad. La colaboración entre administraciones (central, autonómica) es esencial en la consecución de este objetivo y en general, en la mejora de las diversas fuentes administrativas con potencialidad para su utilización con fines estadísticos. Las propias operaciones exhaustivas del sistema estadístico (censos de población, padrón continuo, Movimiento natural de la Población, etc.) deben conceder la importancia que se merece al detalle geográfico y a la disponibilidad de la información para dominios pequeños.

A nivel más “doméstico” se plantea en las oficinas centrales de estadística el problema de la ubicación de la unidad administrativa encargada de la estimación en áreas pequeñas. ¿Debe ser una actividad centralizada, o cada servicio promotor –encuestas sociales, laborales, de empresas,...- debe disponer de su propia capacidad de producción de estimaciones en áreas pequeñas? Otros problemas, como la ‘normalización’ de la difusión de estimaciones en áreas pequeñas, con presentación de información ‘consistente’ con las estadísticas corrientes, o el suministro de microdatos y metainformación con relevancia y transparencia constituyen también retos a resolver en el contexto de esta nueva tipología de datos.

Corresponsales:

Cristina Jiménez Los Arcos,
cjimenez@ine.es
Instituto Nacional de Estadística

Ana Justel
ana.justel@uam.es
Universidad Autónoma de Madrid

Isabel Molina Peralta
imolina@est-econ.uc3m.es
Universidad Carlos III de Madrid

Javier Yañez Gestoso
jayage@mat.ucm.es
Universidad Complutense de
Madrid

Francisco Callealta Barroso
franciscoj.callealta@uah.es
Universidad de Alcalá de Henares

Fernando Reche Lorite
freche@ual.es
Universidad de Almería

Luis Antonio Sarabia Peinador
lsarabia@ubu.es
Universidad de Burgos

Araceli Tuero
tueroma@unican.es
Universidad de Cantabria

Juan Luis González Caballero
juanluis.gonzalez@uca.es
Universidad de Cádiz

José María Caridad y Ocerín
ccjm@uco.es
Universidad de Córdoba

Miguel González Velasco
mvelasco@unex.es
Universidad de Extremadura

Rocío Raya Miranda
rraya@ugr.es
Universidad de Granada

Beatriz Hernández Jiménez
beatriz.hernandez@dmat.uhu.es
Universidad de Huelva

Emilio Lozano Aguilera
elozano@ujaen.es
Universidad de Jaén
David Alcaide López de Pablo
dalcaide@ull.es
Universidad de la Laguna

Juan Carlos Fillat Ballesteros
juan-carlos.fillat@dmc.unirioja.es
Universidad de la Rioja

María Eva Vallejo Pascual
ddeepv@unileon.es
Universidad de León

Lázaro Cánovas
lcanovas@um.es
Universidad de Murcia

Carmen Morcillo Aixelá
aixela@uma.es
Universidad de Málaga

Susana Montes Rodríguez
montes@uniovi.es
Universidad de Oviedo

María Teresa Santos Martín
maysam@gugu.usal.es
Universidad de Salamanca

Antonio Beato Moreno
beato@us.es
Universidad de Sevilla

M^a Cruz Valsero Blanco
mcruz@eio.uva.es
Universidad de Valladolid

Fernando Plo
fplo@unizar.es
Universidad de Zaragoza

Marc Almiñana Alemany
marc@umh.es
Universidad Miguel Hernandez

María del Pilar Moreno Navarro
mpmornav@upo.es
Universidad Pablo de Olavide

Ana Fernández Militino
militino@unavarra.es
Universidad Pública de Navarra

Antonio Alonso
antonio.alonso@urjc.es
Universidad Rey Juan Carlos

José Antonio Vilar
Fernández
eijoseba@udc.es
Universidade da Coruña

María José Lombardía
Cortiña
mjoselc@usc.es
Universidade de Santiago de
Compostela

Alberto Rodríguez Casal
arodriguez@uvigo.es
Universidade de Vigo

Jordi Ocaña
jocana@ub.edu
Universitat de Barcelona

Vera PAWLOWSKY-Glahn
vera.pawlowsky@ima.udg.es
Universitat de Girona

Carles Capdevila Marques
ccm@matematica.udl.es
Universitat de Lleida

José Domingo Bermúdez
Edo
Jose.D.Bermudez@uv.es
Universitat de Valencia

Pilar Muñoz
pilar.munyo@upc.edu
Universitat Politècnica de
Catalunya

Javier Alcaraz Soria
jalcaraz@eio.upv.es
Universitat Politècnica de
Valencia

Dolores Romero Morales
Dolores.Romero-
Morales@said-business-
school.oxford.ac.uk
University of Oxford