

## **Sesiones invitadas**

# STATISTICAL ANALYSIS OF MICROARRAY DATA

GEOFF MCLACHLAN

DNA microarray technology has made it possible to examine the expression of thousands of genes over multiple developmental stages or different experimental conditions. Microarrays are now the most widely used tool in genomics to efficiently measure an organism's gene expression levels. For example, clinical heterogeneity and progression of many diseases, such as cancer, reflect the activity of different genes that command biological processes at the cellular level. The analysis of microarray gene expression data is thus a very important topic in many areas, including medical, veterinary, and agricultural fields. Traditional statistical methodology has to be re-evaluated and modified to carry out the main analyses required for microarray data. Such data consist of several thousands of genes measured over, say, tens or hundreds of tissues.

In our first lecture, we consider an important and common question in microarray experiments, namely the detection of genes that are differentially expressed in tissue samples across a number of specified classes. These classes may correspond to tissues (cells) that are at different stages in some process, in distinct pathological states, or under different experimental conditions. As this problem concerns the selection of significant genes from a large pool of candidate genes, it is usually carried out within the framework of multiple hypothesis testing. A plethora of methods to detect differential gene expression have been proposed. In our presentation, we will concentrate on a rather simple approach based on mixture models, which allow the estimation of the implied false discovery rate and other rates such as the sensitivity.

In the second lecture, we consider another important problem with the analysis of microarray data. It concerns the grouping of the genes over several tissues in multiple development stages or different experimental conditions such as in time-course experiments. A general goal common to many of these experiments is to characterize temporal patterns of gene expression within various biological conditions and to group genes by these patterns. These groups have the potential to provide insight into the biological function of genes.

In the third lecture, attention is centred on the classification of the tissue samples on the basis of the genes in both an unsupervised and a supervised context. In the former context, an aim might be cluster the tissues on the basis of the genes in order to investigate the existence of classes and subclasses of various forms of cancer. In the latter context, an aim might be to develop a prediction rule to be used as a guide to therapeutic interventions for treating diseases of patients on the basis of their gene-expression signature. Unfortunately, most standard cluster and discriminant techniques cannot be applied directly as the number of observations (tissues) is small compared relative to the number of available variables (genes). We present a mixture model-based approach to the unsupervised problem based on extensions of the EM-MIX program. For the supervised problem, we adopt a support vector machine as our classifier and we demonstrate the care that needs to be taken to avoid misleading biases in estimating the associated error rates for a classifier based on a reduced set of genes.

DEPARTMENT OF MATHEMATICS; UNIVERSITY OF QUEENSLAND; ST. LUCÍA;  
BRISBANE; AUSTRALIA 4072

*E-mail address:* `gjm@maths.uq.edu.au`

## OPTIMIZATION METHODS IN BIOMEDICINE

PANOS M. PARDALOS

In recent years optimization has been widely used in many problems in biomedicine. These problems are inherently complex and very difficult to solve. In this talk we are going to focus on global optimization techniques (multi-quadratic 0-1 integer programming) in computational neurosciences and biclustering (nonlinear fractional 0-1 integer programming) based data mining approaches in cancer research. In addition, several other applications will be briefly discussed.

CENTER FOR APPLIED OPTIMIZATION, DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING, UNIVERSITY OF FLORIDA, 303 WEIL HALL, PO BOX 116595, GAINESVILLE, FL 32611-6595

*E-mail address:* `pardalos@ufl.edu`

# OPTIMIZACIÓN DE SERVICIOS DE URGENCIAS

VLADIMIR MARIANOV

Se aborda el tema del diseño de los servicios de emergencias médicas, desde el punto de vista de la localización y dimensionamiento de los recursos. Éstos incluyen los centros de atención de urgencia, el equipamiento mayor y las ambulancias. Adicionalmente a las consideraciones de minimización de costo de estos servicios, en su diseño se debe considerar la calidad de servicio, que entre otros aspectos incluye el tiempo de atención -en el caso de las ambulancias- o el tiempo que las personas permanecen en espera en el caso de los centros de urgencias. Por otra parte, el propio modelamiento del problema tiene ya sus dificultades, como por ejemplo la determinación de qué tiempos de respuesta son adecuados, o cuál debe ser la distribución probabilística de la demanda, entre otros aspectos. Se hace un análisis de las dificultades del problema, así como un resumen de los modelos y procedimientos de solución propuestos en la literatura.

PONTIFICIA UNIVERSIDAD CATLICA DE CHILE, FACULTAD DE INGENIERÍA, DEPARTAMENTO DE INGENIERÍA ELÉCTRICA, AVDA. VICUNA MACKENNA 4860, SAN JOAQUÍN, SANTIAGO-CHILE

*E-mail address:* marianov@ing.puc.cl

# PREPARACIÓN DE DATOS PARA ENCUESTAS DE EMPRESAS: EL MANUAL EDIMBUS

BEAT HULLIGER, DANIEL KILCHMANN

Raw data of business surveys often contains errors and missing values. Their detection and treatment is expensive and long. Many methods and approaches for particular problems exist but few comprehensive manuals are available. The EDIMBUS project, a joint effort of the National Statistical Institutes of Italy (ISTAT), Netherlands (CBS) and Switzerland (SFSO) with support from Eurostat, developed a Recommended Practices Manual (RPM) which treats the design of the process of editing and imputation, its place in the design of a survey and presents recommended methods. The main objectives of editing and imputation, to inform on the quality of a survey and to prepare the data for analysis, are explained. Advantages and limitations of established methods for the detection and treatment of errors and missing values are discussed and recommendations on their use are given. The methods range from computer aided interactive treatment to automatic procedures and from micro to macro level. Testing of the editing and imputation process and monitoring its application by quality indicators as well as the documentation of editing and imputation are important parts of the EDIMBUS-RPM.

The talk will give an overview over the EDIMBUS manual and its main features. Selective editing, which aims at concentrating efforts on the most influential errors, and outlier detection and treatment are discussed in more detail. The relation between the score functions of selective editing, the influence of observations and outliers is explored.

UNIVERSITY OF APPLIED SCIENCES NORTHWESTERN SWITZERLAND AND SWISS  
FEDERAL STATISTICAL OFFICE

*E-mail address:* `beat.hulliger@fhnw.ch`

# **PREDICCIÓN SECUENCIAL BAJO INFORMACIÓN INCOMPLETA**

**GABOR LUGOSI**

En esta charla se resumen algunas técnicas de la predicción secuencial cuando el predictor tiene acceso limitado a las instancias pasadas de la sucesión que se predice. Un ejemplo clásico es el problema de bandido de  $N$  brazos. Se presentan varios métodos aleatorizados y se investigan condiciones de consistencia. Algunos de los resultados presentados se han obtenido en colaboración con Nicolo Cesa-Bianchi, Gilles Stoltz y Shie Mannor.

DEPARTMENT OF ECONOMICS, POMPEU FABRA UNIVERSITY, RAMON TRIAS  
FARGAS 25-27, 08005, BARCELONA, SPAIN

*E-mail address:* [gabor.lugosi@gmail.com](mailto:gabor.lugosi@gmail.com)