Sesión Especial 20

Diseño óptimo de experimentos

Organizadores:

- Edgar Benítez Sastoque (Universidad de Navarra)
- Carlos de la Calle Arroyo (Universidad de Navarra)
- Álvaro Cía Mina (Universidad de Navarra)

Descripción:
En la sesión sobre Diseño Óptimo de Experimentos se abordarán aplicaciones de esta rama, como la optimización del muestreo en experimentos clásicos. También se discuten enfoques novedosos como la teoría de campos múltiples aleatorios y el submuestreo secuencial en línea para modelos de predicción en el contexto de grandes cantidades de datos.

Programa

JUEVES, 25 de enero:

11:30 – 12:00	Irene Mariñas del Collado (Universidad de Oviedo) *Modelling ethanol pharmacokinetics: from single to multiple intakes and D-optimal designs*
12:00 – 12:30	José Guillermo Sánchez León (Universidad de Salamanca) *Optimization of light curves to determine the period-luminosity relation*
12:30 – 13:00	Édgar Benítez Sastoque (Universidad de Navarra) *Multi-Random Field Theory for Optimal Experimental Design with Repeated Measurements*
13:00 – 13:30	Álvaro Cía Mina (Universidad de Navarra) *Sequential Online Subsampling for Prediction with Random Covariates*

Modelling ethanol pharmacokinetics: from single to multiple intakes and D-optimal designs

Irene Mariñas-Collado, Juan M. Rodríguez-Díaz, M. Teresa Santos-Martín

Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática, Universidad de Oviedo

marinasirene@uniovi.es

Abstract: Traditional models for estimating blood alcohol concentration (BAC) tend to oversimplify the complex process of alcohol metabolism and ignore the staggered intake patterns often observed in real life. We propose a new model that takes into account multiple alcohol intake. D-optimal designs are calculated for this model, depending on the time of initial values of the model parameters including the time at which the second intake occurs.

Optimization of light curves to determine the period-luminosity relation

Guillermo Sánchez-León, Juan M. Rodríguez-Díaz

University of Salamanca

guillermo@usal.es

Abstract: Light curves measure the variation in the luminosity of a celestial body. A typical example is the calculation of the rotation periods of a binary system. An eclipsing binary consists of two close stars moving in an orbit such as, in relation to Earth, the light of one can sometimes be hidden behind the other. In an eclipsing binary system we can distinguish the repeating phases of an eclipse: without eclipse, partial eclipse, and total eclipse. The same method can be used to calculate the rotation period of an irregular asteroid whose luminosity varies with rotation. Building an experimental light curve consists of measuring the luminosity at different times. We aim to find an optimal design to choose the best moments where the measurement should be taken. The problem is that the design must take into account that the measurements can only be taken at certain known time intervals (during the night) and that the measurements are also conditioned by meteorological conditions, which is a random component.

Multi-Random Field Theory for Optimal Experimental Design with Repeated Measurements

Edgar Benitez, Jesús López-Fidalgo

DATAI (Tecnun), Universidad de Navarra

ebenitezs@unav.es

Abstract: In 2020, over 19 million cases of cancer were estimated, with 10 million resulting in fatalities. The pursuit of treatments to decelerate tumor progression remains a pivotal research area. Empirical models in pre-clinical stages assess treatment effects, showing results adequately correlating with human exposure at therapeutic dosages. Nonlinear (sigmoidal) models, observing macroscopic behaviors like volume or cell count, are employed [4].

Defining optimal experimental design is crucial for valid parameter estimation inferences at a reduced cost [3]. In nonlinear models, D-optimal designs prove efficient in parameter estimation, presuming independent observations [2, 3]. However, this isn't typically the case in temporal processes like tumor growth.

In longitudinal studies, measuring the same experimental unit over time violates independence assumptions. Hughes-Oliver demonstrated that considering covariance enhances parameter estimation [2]. Nonetheless, integrating correlated structures into growth models is rare [1]. This work aims to evaluate the inclusion of correlation structures in optimal design settings and its impact on parameter estimation performance.

Referencias

[1] J. Lopez Fidalgo, I. M. Ortiz Rodriguez, W. K. Wong (2011). Design issues for population growth models, Journal of Applied Statistics, 38(3), 501-512.

[2] J. Hughes-Oliver (1998). Optimal Designs for Nonlinear Models with Correlated Errors, Lecture Notes-Monograph Series, 34, 163-174.

[3] A. Atkinson, A. Donev, R. Tobias (2007). Optimum experimental designs, Oxford University Press, New York, U. S. A.

[4] M. Simeoni, P. Magni, C. Cammia, G. De Nicolao, V. Croci, E. Pesenti, M. Germani, I. Poggesi, M. Roccheti (2004). Predictive Pharmacokinetic-Pharmacodynamic Modeling of Tumor Growth Kinetics in Xenograft Models after Administration of Anticancer Agents, Cancer Research, 64, 1094-1101.

Sequential Online Subsampling for Prediction with Random Covariates

Álvaro Cía-Mina, Jesús López-Fidalgo

Instituto de Ciencia de los Datos e Inteligencia Artificial (DATAI), Universidad de Navarra

aciamina@unav.es

Abstract: Subsampling is commonly employed to improve computation efficiency in regression models. However, existing methods primarily focus on minimizing errors in estimating parameters, whereas the main practical goal of statistical models often lies in minimizing prediction errors. This study introduces a novel approach to selecting subdata for linear models, which takes into account the distribution of covariates. Our method specifically addresses scenarios with large samples where obtaining labels for the response variable is costly. The introduction of the "J-optimalityçriterion is supported by theoretical justifications and aligned with standard linear optimality criteria. The sequential approach is developed for online data, where the distribution of the convariates is unknown. As expected based on theory, our method demonstrates a reduction in prediction mean squared error compared to existing methods. Through simulations, we present empirical evidence of the performance and potential of our approach in enhancing prediction accuracy.