Biblioteca del Instituto de Estadística y Cartografía de Andalucía

Resúmenes de revistas
Enero-febrero 2018

JUNTA DE ANDALUCIA

Instituto de Estadística y Cartografía de Andalucía
CONSEJERÍA DE ECONOMÍA, INNOVACIÓN, CIENCIA Y EMPLEO

PRESENTACIÓN

El presente boletín de resúmenes tiene una periodicidad mensual y con él la Biblioteca del Instituto de Estadística y Cartografía de Andalucía pretende dar a conocer a los usuarios de una forma detallada el contenido de las revistas especializadas que entran en su colección. Se trata de un complemento al boletín de novedades de publicaciones seriadas ya que en él se incluyen los resúmenes de cada uno de los artículos que aparecen publicados en los diferentes números de las revistas en el idioma original de las mismas.

Los resúmenes de este boletín corresponden a las revistas que han ingresado en la Biblioteca del Instituto de Estadística y Cartografía de Andalucía durante los meses de enero y febrero de 2018 y que pueden consultarse gratuitamente en sus instalaciones en la siguiente dirección:

Instituto de Estadística y Cartografía de Andalucía
Pabellón de Nueva Zelanda
C/Leonardo Da Vinci, n. 21. Isla de La Cartuja
41071 - SEVILLA
E-mail: biblio.ieca@juntadeandalucia.es
Teléfono: 955 033 800
Fax: 955 033 816

Horario de atención al público:
Lunes y martes: de 9:00h a 14:00h. y de 16:00 a 19:00 h.
Miércoles, jueves y viernes: de 9:00h a 14:00h.
Horario de verano (del 15 de junio al 15 de septiembre), Semana Santa, Feria de Sevilla y Navidad (del 24 de diciembre al 6 de enero): de lunes a viernes de 9:00h. a 14:00h.

AH: Andalucía en la historia, ISSN 1695-1956
Número 54 (octubre – diciembre 2016)

Resumen
Coincidiendo con la celebración del tercer centenario del nacimiento de Carlos III (1716-1788), *Andalucía en la Historia* publica un monográfico que pone de relieve los logros del rey ilustrado en nuestra comunidad. No sin afrontar tensiones y resistencias, el reformismo impulsado por el monarca y sus ministros sembró las semillas que generaron un nuevo modelo político, económico, social y cultural en la región andaluza. Las reformas borbónicas, iniciadas por Felipe V y reforzadas por Carlos III —reorganización administrativa de los municipios, iniciativa repobladora de Sierra Morena, tratados de libre comercio, impulso de las universidades y de las sociedades económicas de amigos del país, etc.— hicieron posible la modernización de la región, haciéndola avanzar en un proceso que conduciría a la aparición de una nueva época, cuya eclosión puede simbolizarse en las Cortes de Cádiz. Coordinado por Carlos Martínez Shaw, catedrático de Historia Moderna de la Universidad Nacional de Educación a Distancia, el dosier cuenta con un destacado elenco de conocidos especialistas.

Resumen
Sin un Homero que narrara su epopeya, la historia de los fenicios en la Península Ibérica ha quedado muchas veces en un segundo plano. Pero gracias a la investigación arqueológica, la cultura fenicia está siendo desenterrada y en esta tarea Andalucía desempeña un papel trascendental por la cantidad y la calidad de sus yacimientos arqueológicos.

Resumen
Los dominicos llegaron a Andalucía como unos repobladores más hace ocho siglos. Su labor catequética, la predicación y su particular forma de vivir el mensaje cristiano hicieron de la Orden de Predicadores uno de los referentes en la creación de devociones, lugares de culto y centros de enseñanza en Andalucía.

Resumen
A finales del siglo XV se dio un particular florecimiento de las actividades pirático-corsarias, durante el cual los comerciantes del Reino de Sevilla, que recorrían costas y bahías para llevar y traer productos de los mercados europeos, pugnaron por prosperar por encima de este clima generalizado de violencia.

Resumen
En prueba y reconocimiento a todos los servicios prestados, la Corona otorgó al ilustrado andaluz —académico, militar y administrador— Francisco de Saavedra la Orden de Carlos III, la condecoración con mayor predicamento de la historia española.

Resumen
El accidente nuclear de Palomares fue, a nivel global, el más importante hasta Chernóbil (1986). Sin embargo, es uno de los sucesos más oscuros y desconocidos del franquismo, plagado de falsos mitos y leyendas.

Resumen
Melchor Rodríguez es una de las figuras más representativas de una corriente anarquista que tuvo en la Guerra Civil la prueba más dura a la que se puede enfrentar un libertario: defender la vida de sus enemigos acérrimos. Salvó a miles de personas de derechas durante la Guerra Civil, más que el famoso Schindler alemán.

Cartographic journal, The, ISSN 0008-7041
Volume 54, number 3 (august 2017)

Abstract
Historical enclosure era property-related maps can tell us a great deal about the life and times of communities in the past. This study offers a unique approach to studying the historical landscape by applying GIS techniques to the examination of an eighteenth-century English village. Using novel GIS applications relying on historical maps, the study explores various aspects of the village's physical and social characteristics. In doing so, the study forges effective linkages between cultural and landscape variables to reveal aspects of the historical landscape in eighteenth-century Britain previously inaccessible to researchers. This, in turn, provides a much more comprehensive and sophisticated template for future use by historical geographers in a number of contexts.

Abstract
The long sequence of Pigot's plans of Manchester and Salford is used to test the concept that the dates of churches and chapels can be used as a valuable indicator of the completeness of the coverage of large-scale nineteenth-century town plans. The approach appears to hold some promise and suggests that Pigot's plans were surprisingly comprehensive. This may reflect not merely his drawing on existing surveys but, more interestingly, may be the incidental product of collecting data for his town directories. The methodology could usefully be extended to explore the value of directory plans of other towns.

Abstract
The paper describes a localization of Müller's maps of regions of Bohemia from 1712 to 1718. Original maps represent the territories within regional boundaries in approximate scale 1: 100 000. It is relatively problematic to extract spatial information from the maps based on precise geodetic control and well-known cartographic projection. A different approach must be chosen in case of old maps without geodetic control and identifiable cartographic projection. In such a case the identical points whose coordinates in the reference coordinate system are known must be identified in the old map and their cartometric coordinates measured. This is also the case of manuscript Müller's maps. For creation of a transformation key the suitable input data must be selected. As the most frequented features on these maps are settlements it was decided to use this part of planimetric component. Several ways how to use the settlements for transformation were explored in order to find out the most appropriate way of localization of these rare old maps. For

purpose of old maps localization the database of settlements (DBS) was used. This database is based on the Territorial Identification Register of Basic Settlement Units (TIR-BSU) which has been created in 1992–2004 and contains current coordinates of settlements. Furthermore, after transformation, the analysis of the visualization accuracy of watercourses was done.

Exploiting Illusory Grid Lines for Object-Location Memory Performance in Urban Topographic Maps P. 242-253

Frank Dickmann, Dennis Edler, Anne-Kathrin Bestgen & Lars Kuchinke

Abstract

A good seabed representation is one of the important characteristics of any navigational chart. Along with depth contours and coloured depth areas, soundings are used for this task. All the soundings on a navigational chart are selected for a reason. Soundings contribute to the navigational chart safety aspect by alerting to all the threats and dangers. They also show all the attributes of a seabed relief without overcrowding it, thus maintaining the overall chart quality. Soundings are selected from a hydrographic survey and since it consists of a vast number of data, the process of sounding selection is a challenging and demanding task. It requires experience and knowledge from the nautical cartographer and is mostly done manually. Some types of software nowadays provide an automatic selection feature. This paper analyses a process of automatic sounding selection in the dKart Editor software. On the Croatian side of the Adriatic Sea, Šibenski Kanal (Šibenik channel) and Kanal Sv. Ante (St. Ante's channel) are used as the study area. A hydrographic survey of the area represents the input data. The official navigational chart of the surveyed area is used as the basis for determining three different sets of parameters for the selection process. After the selection, obtained results are assessed based on geometrical accuracy and on the conservation level of navigational safety. For geometrical accuracy, the best results were produced by the third set that was divided in two subsets for each channel. It was determined that the nature of the seabed relief had an impact on the selection process. The same set had the best result for navigational safety assessment but it was concluded that all the sets undermined the aspect. Because of these crucial shortcomings noticed in all the tested sets of parameters, the feature is considered inadequate for serious usage as a completely automatic tool for the process of sounding selection on navigational charts.

A Comparative Study of Various Supervised Learning Approaches to Selective Omission in a Road Network P. 254-264

Qi Zhou & Zhilin Li

Abstract

On the basis of initial studies devoted to a better understanding of how the public user (a pedestrian in the city) perceives cartographic symbols in the mobile augmented reality system, we present an attempt to determine the threshold values of differentiation for three visual variables. The variables of *size*, *transparency*, *and focus* were implemented into image point symbols representing five types of objects. The set of symbols was designed in accordance with the rules of cartographic design taking into consideration an analysis of 19 professional tourist works. The symbols were presented on the screen of a mobile device in a system imitating the augmented reality system against four different backgrounds: white, a wall, and two typical urban landscapes. The results of an internet survey conducted using a tablet at four locations in Poznan (Poland) allowed us to determine the following: threshold differentiation values for the analysed variables, indication of the dependence on the type of background displayed on the mobile device in augmented reality, and the advantage of using a combination of visual variables.

Adaptive Multi-Scale Population Spatialization Model Constrained by Multiple Factors: A Case Study of Russia P. 265-282

Lujin Hu, Zongyi He & Jiping Liu

Abstract

Dot maps are one of the best ways to visualize absolute values in thematic cartography. Dots represent quantitative data on a map. Population is often used in this type of representation. This paper presents a population dot density map for

the year 2011 on two scales: (1) for mainland Portugal, and (2) for the Lisbon and Oporto regions. We have used dots with constant values and sizes at the most detailed statistical level (i.e. statistical subsection) for localities with less than 5000 inhabitants, and proportional circles for localities with more than 5000 inhabitants. These two scales of analysis coupled with two cartographic representation techniques used on a single map allow for a clear reading of the distribution of population.

Estadística española : revista del Instituto Nacional de Estadística, ISSN 0014-1151
Número 192 (enero-abril 2017)

Resumen
El objetivo de este trabajo es evaluar la hipótesis causal de que el estatus socioeconómico de los estudiantes y la actitud de los estudiantes hacia la matemática son factores que determinan en gran medida los resultados académicos de los estudiantes costarricenses, medido a través del resultado en la prueba de alfabetización matemática de PISA 2012. Para esto, se define un modelo de medida de los constructos latentes y se estima el modelo estructural, tanto desde el enfoque clásico como desde el enfoque Bayesiano, para comparar ambos tipos de estimaciones resaltando las bondades del método Bayesiano para este tipo de modelos.

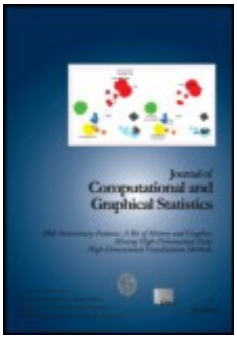
Abstract
Many modern scholars think that the term "histogram" is related to the word "history". Recent work in the field of the history of statistics has only increased this misunderstanding. The etymology is incorrect for several reasons: first, the word "histogram" does not share a stem with "history"; second, we show that Karl Pearson, who coined the term, used a clearly defined method to designate recently devised graphs. His method, inspired by Levasseur's work, excludes any relation to the word "history" and confirms the link to the Greek *ἱστός* (mast). Some clarifications of Pearson's use of the term are given in the paper.

Resumen
Muchos especialistas contemporáneos, dentro y fuera del campo de la estadística, piensan que el término "histograma" está relacionado con el sustantivo "historia", una confusión que agravan algunos trabajos recientes en el campo de la historia de la estadística. Esta etimología es necesariamente errónea por varias razones: en primer lugar, desde un punto de vista etimológico "histograma" no puede compartir la raíz de "historia". Pero lo más relevante es que sabemos que Karl Pearson, que fue quien acuñó el término y contaba con una buena preparación filológica, había diseñado un método muy preciso para la designación de los gráficos que se habían diseñado recientemente para visualizar datos estadísticos. Su método, inspirado en el trabajo de Levasseur, excluye cualquier relación con "historia" y confirma su relación con el sustantivo griego "ἱστός" ("mástil"). Se hacen algunas clarificaciones sobre el uso que Pearson daba al término estudiado.

Resumen
En el presente artículo estudiamos las propiedades del estimador "cuasidesviación estándar muestral" como estimador de la "desviación estándar poblacional" cuando el diseño muestral es el muestreo aleatorio simple con reemplazamiento de tamaño fijo, así como cuando este tamaño muestral tiende a infinito.

Abstract
To avoid biased results, sample units must be included in the sample in the correct proportion. Sample weights are intended to correct potential disproportions observed in survey sample data. While their use is widely accepted to estimate population descriptive statistics, their role to estimate causal effects is not clear. This paper analyzes when and how to use weights, considering wages in Colombia as an example, providing a procedure for selecting the final weight components based on empirical evidence. Results indicate that weights are required for descriptive statistics to resemble the population ones. However, several coefficients obtained from weighted and unweighted wage equations show no significant differences.

Resumen
A fin de evitar resultados sesgados, las unidades muestrales deben ser incluidas en la muestra en la proporción correcta. Los pesos muestrales se utilizan para corregir posibles desproporciones, frecuentes en datos muestrales. Mientras su uso es ampliamente aceptado para estimar estadísticas descriptivas de la población, su papel en la estimación de efectos causales no es claro. Este trabajo analiza cuándo y cómo utilizar dichos pesos, considerando salarios en Colombia como ejemplo, proporcionando un procedimiento para seleccionar los componentes del peso final basado en la evidencia empírica. Los resultados indican que los pesos muestrales son necesarios para que las estadísticas descriptivas de la muestra se asemejen a las de la población. Sin embargo, varios coeficientes obtenidos a partir de ecuaciones de salarios ponderados y no ponderados no muestran diferencias significativas.



Journal of computational and graphical statistics, ISSN 1061-8600
Volume 26, number 1 (2017)

Abstract
Data with multivariate count responses frequently occur in modern applications. The commonly used multinomial-logit model is limiting due to its restrictive mean-variance structure. For instance, analyzing count data from the recent RNA-seq technology by the multinomial-logit model leads to serious errors in hypothesis testing. The ubiquity of overdispersion and complicated correlation structures among multivariate counts calls for more flexible regression models. In this article, we study some generalized linear models that incorporate various correlation structures among the counts. Current literature lacks a treatment of these models, partly because they do not belong to the natural exponential family. We study the estimation, testing, and variable selection for these models in a unifying framework. The regression models are compared on both synthetic and real RNA-seq data. Supplementary materials for this article are available online.

Abstract
In many atmospheric and earth sciences, it is of interest to identify dominant spatial patterns of variation based on data observed at p locations and n time points with the possibility that $p > n$. While principal component analysis (PCA) is commonly applied to find the dominant patterns, the eigenimages produced from PCA may exhibit patterns that are too noisy to be physically meaningful when p is large relative to n. To obtain more precise estimates of eigenimages, we propose a regularization approach incorporating smoothness and sparseness of eigenimages, while accounting for their orthogonality. Our method allows data taken at irregularly spaced or sparse locations. In addition, the resulting optimization problem can be solved using the alternating direction method of multipliers, which is easy to implement, and applicable to a large spatial dataset. Furthermore, the estimated eigenfunctions provide a natural basis for representing the underlying spatial process in a spatial random-effects model, from which spatial covariance function estimation and spatial prediction can be efficiently performed using a regularized fixed-rank kriging method. Finally, the effectiveness of the proposed method is demonstrated by several numerical examples.

Abstract
Sufficient dimension reduction (SDR) is a paradigm for reducing the dimension of the predictors without losing regression information. Most SDR methods require inverting the covariance matrix of the predictors. This hinders their use in the analysis of contemporary datasets where the number of predictors exceeds the available sample size and the predictors are highly correlated. To this end, by incorporating the seeded SDR idea and the sequential dimension-reduction framework, we propose a SDR method for high-dimensional data with correlated predictors. The performance of the proposed method is

studied via extensive simulations. To demonstrate its use, an application to microarray gene expression data where the response is the production rate of riboflavin (vitamin B_2) is presented.

Variational Approximations for Generalized Linear Latent Variable Models P. 35-43

Francis K. C. Hui, David I. Warton, John T. Ormerod, Viivi Haapaniemi & Sara Taskinen

Abstract

Generalized linear latent variable models (GLLVMs) are a powerful class of models for understanding the relationships among multiple, correlated responses. Estimation, however, presents a major challenge, as the marginal likelihood does not possess a closed form for nonnormal responses. We propose a variational approximation (VA) method for estimating GLLVMs. For the common cases of binary, ordinal, and overdispersed count data, we derive fully closed-form approximations to the marginal log-likelihood function in each case. Compared to other methods such as the expectation-maximization algorithm, estimation using VA is fast and straightforward to implement. Predictions of the latent variables and associated uncertainty estimates are also obtained as part of the estimation process. Simulations show that VA estimation performs similar to or better than some currently available methods, both at predicting the latent variables and estimating their corresponding coefficients. They also show that VA estimation offers dramatic reductions in computation time particularly if the number of correlated responses is large relative to the number of observational units. We apply the variational approach to two datasets, estimating GLLVMs to understanding the patterns of variation in youth gratitude and for constructing ordination plots in bird abundance data. R code for performing VA estimation of GLLVMs is available online. Supplementary materials for this article are available online.

A Marginal Sampler for σ-Stable Poisson–Kingman Mixture Models P.44-53

María Lomelí, Stefano Favaro & Yee Whye Teh

Abstract

We investigate the class of σstable Poisson–Kingman random probability measures (RPMs) in the context of Bayesian nonparametric mixture modeling. This is a large class of discrete RPMs, which encompasses most of the popular discrete RPMs used in Bayesian nonparametrics, such as the Dirichlet process, Pitman–Yor process, the normalized inverse Gaussian process, and the normalized generalized Gamma process. We show how certain sampling properties and marginal characterizations of σstable Poisson–Kingman RPMs can be usefully exploited for devising a Markov chain Monte Carlo (MCMC) algorithm for performing posterior inference with a Bayesian nonparametric mixture model. Specifically, we introduce a novel and efficient MCMC sampling scheme in an augmented space that has a small number of auxiliary variables per iteration. We apply our sampling scheme to a density estimation and clustering tasks with unidimensional and multidimensional datasets, and compare it against competing MCMC sampling schemes. Supplementary materials for this article are available online.

Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis P. 54-65

Zhiqiang Tan

Abstract

Consider the two problems of simulating observations and estimating expectations and normalizing constants for multiple distributions. First, we present a self-adjusted mixture sampling method, which accommodates both adaptive serial tempering and a generalized Wang–Landau algorithm. The set of distributions are combined into a labeled mixture, with the mixture weights depending on the initial estimates of log normalizing constants (or free energies). Then, observations are generated by Markov transitions, and free energy estimates are adjusted online by stochastic approximation. We propose two stochastic approximation schemes by Rao–Blackwellization of the scheme commonly used, and derive the optimal choice of a gain matrix, resulting in the minimum asymptotic variance for free energy estimation, in a simple and feasible form. Second, we develop an offline method, locally weighted histogram analysis, for estimating free energies and expectations, using all the simulated data from multiple distributions by either self-adjusted mixture sampling or other sampling algorithms. This method can be computationally much faster, with little sacrifice of statistical efficiency, than a global method currently

used, especially when a large number of distributions are involved. We provide both theoretical results and numerical studies to demonstrate the advantages of the proposed methods.

Imposing Minimax and Quantile Constraints on Optimal Matching in Observational Studies P. 66-78

Paul R. Rosenbaum

Abstract

Modern methods construct a matched sample by minimizing the total cost of a flow in a network, finding a pairing of treated and control individuals that minimizes the sum of within-pair covariate distances subject to constraints that ensure distributions of covariates are balanced. In aggregate, these methods work well; however, they can exhibit a lack of interest in a small number of pairs with large covariate distances. Here, a new method is proposed for imposing a minimax constraint on a minimum total distance matching. Such a match minimizes the total within-pair distance subject to various constraints including the constraint that the maximum pair difference is as small as possible. In an example with 1391 matched pairs, this constraint eliminates dozens of pairs with moderately large differences in age, but otherwise exhibits the same excellent covariate balance found without this additional constraint. A minimax constraint eliminates edges in the network, and can improve the worst-case time bound for the performance of the minimum cost flow algorithm, that is, a better match from a practical perspective may take less time to construct. The technique adapts ideas for a different problem, the bottleneck assignment problem, whose sole objective is to minimize the maximum within-pair difference; however, here, that objective becomes a constraint on the minimum cost flow problem. The method generalizes. Rather than constrain the maximum distance, it can constrain an order statistic. Alternatively, the method can minimize the maximum difference in propensity scores, and subject to doing that, minimize the maximum robust Mahalanobis distance. An example from labor economics is used to illustrate. Supplementary materials for this article are available online.

Sampling for Conditional Inference on Contingency Tables P. 79-87

Robert D. Eisinger & Yuguo Chen

Abstract

We propose new sequential importance sampling methods for sampling contingency tables with given margins. The proposal for each method is based on asymptotic approximations to the number of tables with fixed margins. These methods generate tables that are very close to the uniform distribution. The tables, along with their importance weights, can be used to approximate the null distribution of test statistics and calculate the total number of tables. We apply the methods to a number of examples and demonstrate an improvement over other methods in a variety of real problems. Supplementary materials are available online.

Circulant Embedding of Approximate Covariances for Inference From Gaussian Data on Large Lattices P. 88-97

Joseph Guinness & Montserrat Fuentes

Abstract

Recently proposed computationally efficient Markov chain Monte Carlo (MCMC) and Monte Carlo expectation–maximization (EM) methods for estimating covariance parameters from lattice data rely on successive imputations of values on an embedding lattice that is at least two times larger in each dimension. These methods can be considered exact in some sense, but we demonstrate that using such a large number of imputed values leads to slowly converging Markov chains and EM algorithms. We propose instead the use of a discrete spectral approximation to allow for the implementation of these methods on smaller embedding lattices. While our methods are approximate, our examples indicate that the error introduced by this approximation is small compared to the Monte Carlo errors present in long Markov chains or many iterations of Monte Carlo EM algorithms. Our results are demonstrated in simulation studies, as well as in numerical studies that explore both increasing domain and fixed domain asymptotics. We compare the

exact methods to our approximate methods on a large satellite dataset, and show that the approximate methods are also faster to compute, especially when the aliased spectral density is modeled directly. Supplementary materials for this article are available online.

An Inversion-Free Estimating Equations Approach for Gaussian Process Models P. 98-107

Mihai Anitescu, Jie Chen & Michael L. Stein

Abstract

One of the scalability bottlenecks for the large-scale usage of Gaussian processes is the computation of the maximum likelihood estimates of the parameters of the covariance matrix. The classical approach requires a Cholesky factorization of the dense covariance matrix for each optimization iteration. In this work, we present an estimating equations approach for the parameters of zero-mean Gaussian processes. The distinguishing feature of this approach is that no linear system needs to be solved with the covariance matrix. Our approach requires solving an optimization problem for which the main computational expense for the calculation of its objective and gradient is the evaluation of traces of products of the covariance matrix with itself and with its derivatives. For many problems, this is an $O(n \log n)$ effort, and it is always no larger than $O(n^2)$. We prove that when the covariance matrix has a bounded condition number, our approach has the same convergence rate as does maximum likelihood in that the Godambe information matrix of the resulting estimator is at least as large as a fixed fraction of the Fisher information matrix. We demonstrate the effectiveness of the proposed approach on two synthetic examples, one of which involves more than 1 million data points.

Bayesian and Maximum Likelihood Estimation for Gaussian Processes on an Incomplete Lattice P. 108-120

Jonathan R. Stroud, Michael L. Stein & Shaun Lysen

Abstract

This article proposes a new approach for Bayesian and maximum likelihood parameter estimation for stationary Gaussian processes observed on a large lattice with missing values. We propose a Markov chain Monte Carlo approach for Bayesian inference, and a Monte Carlo expectation-maximization algorithm for maximum likelihood inference. Our approach uses data augmentation and circulant embedding of the covariance matrix, and provides likelihood-based inference for the parameters and the missing data. Using simulated data and an application to satellite sea surface temperatures in the Pacific Ocean, we show that our method provides accurate inference on lattices of sizes up to 512 × 512, and is competitive with two popular methods: composite likelihood and spectral approximations.

Bayesian Model Assessment in Joint Modeling of Longitudinal and Survival Data With Applications to Cancer Clinical Trials P. 121-133

Danjie Zhang, Ming-Hui Chen, Joseph G. Ibrahim, Mark E. Boye & Wei Shen

Abstract

Joint models for longitudinal and survival data are routinely used in clinical trials or other studies to assess a treatment effect while accounting for longitudinal measures such as patient-reported outcomes. In the Bayesian framework, the deviance information criterion (DIC) and the logarithm of the pseudo-marginal likelihood (LPML) are two well-known Bayesian criteria for comparing joint models. However, these criteria do not provide separate assessments of each component of the joint model. In this article, we develop a novel decomposition of DIC and LPML to assess the fit of the longitudinal and survival components of the joint model, separately. Based on this decomposition, we then propose new Bayesian model assessment criteria, namely, ΔDIC and ΔLPML, to determine the importance and contribution of the longitudinal (survival) data to the model fit of the survival (longitudinal) data. Moreover, we develop an efficient Monte Carlo method for computing the conditional predictive ordinate statistics in the joint modeling setting. A simulation study is conducted to examine the empirical performance of the proposed criteria and the proposed methodology is further applied to a case study in mesothelioma. Supplementary materials for this article are available online.

Computationally Efficient Changepoint Detection for a Range of Penalties P. 134-143

Kaylea Haynes, Idris A. Eckley & Paul Fearnhead

Abstract

In the multiple changepoint setting, various search methods have been proposed, which involve optimizing either a constrained or penalized cost function over possible numbers and locations of changepoints using dynamic programming. Recent work in the penalized optimization setting has focused on developing an exact pruning-based approach that, under certain conditions, is linear in the number of data points. Such an approach naturally requires the specification of a penalty to avoid under/over-fitting. Work has been undertaken to identify the appropriate penalty choice for data-generating processes with known distributional form, but in many applications the model assumed for the data is not correct and these penalty choices are not always appropriate. To this end, we present a method that enables us to find the solution path for all choices of penalty values across a continuous range. This permits an evaluation of the various segmentations to identify a suitable penalty choice. The computational complexity of this approach can be linear in the number of data points and linear in the difference between the number of changepoints in the optimal segmentations for the smallest and largest penalty values. Supplementary materials for this article are available online.

Principal Nested Spheres for Time-Warped Functional Data Analysis P. 144-151

Qunqun Yu, Xiaosun Lu & J. S. Marron

Abstract

There are often two important types of variation in functional data: the horizontal (or phase) variation and the vertical (or amplitude) variation. These two types of variation have been appropriately separated and modeled through a domain warping method (or curve registration) based on the Fisher–Rao metric. This article focuses on the analysis of the horizontal variation, captured by the domain warping functions. The square-root velocity function representation transforms the manifold of the warping functions to a Hilbert sphere. Motivated by recent results on manifold analogs of principal component analysis, we propose to analyze the horizontal variation via a principal nested spheres approach. Compared with earlier approaches, such as approximating tangent plane principal component analysis, this is seen to be an efficient and interpretable approach to decompose the horizontal variation in both simulated and real data examples.

Interweaving Markov Chain Monte Carlo Strategies for Efficient Estimation of Dynamic Linear Models P. 152-159

Matthew Simpson, Jarad Niemi & Vivekananda Roy

Abstract

In dynamic linear models (DLMs) with unknown fixed parameters, a standard Markov chain Monte Carlo (MCMC) sampling strategy is to alternate sampling of latent states conditional on fixed parameters and sampling of fixed parameters conditional on latent states. In some regions of the parameter space, this standard data augmentation (DA) algorithm can be inefficient. To improve efficiency, we apply the interweaving strategies of Yu and Meng to DLMs. For this, we introduce three novel alternative DAs for DLMs: the scaled errors, wrongly scaled errors, and wrongly scaled disturbances. With the latent states and the less well known scaled disturbances, this yields five unique DAs to employ in MCMC algorithms. Each DA implies a unique MCMC sampling strategy and they can be combined into interweaving and alternating strategies that improve MCMC efficiency. We assess these strategies using the local level model and demonstrate that several strategies improve efficiency relative to the standard approach and the most efficient strategy interweaves the scaled errors and scaled disturbances. Supplementary materials are available online for this article.

How Many Communities Are There? P. 171-181

D. Franco Saldaña, Yi Yu & Yang Feng

Abstract

Stochastic blockmodels and variants thereof are among the most widely used approaches to community detection for social networks and relational data. A stochastic blockmodel partitions the nodes of a network into disjoint sets, called communities. The approach is inherently related to clustering with mixture models; and raises a similar model selection problem for the number of communities. The Bayesian information criterion (BIC) is a popular solution, however, for stochastic blockmodels, the conditional independence assumption given the communities of the endpoints among different edges is usually violated in practice. In this regard, we propose composite likelihood BIC (CL-BIC) to select the number of communities, and we show it is robust against possible misspecifications in the underlying stochastic blockmodel assumptions. We derive the requisite methodology and illustrate the approach using both simulated and real data. Supplementary materials containing the relevant computer code are available online.

Efficient Computation of the Joint Sample Frequency Spectra for Multiple Populations P. 182-194

John A. Kamm, Jonathan Terhorst & Yun S. Song

Abstract

A wide range of studies in population genetics have employed the sample frequency spectrum (SFS), a summary statistic which describes the distribution of mutant alleles at a polymorphic site in a sample of DNA sequences and provides a highly efficient dimensional reduction of large-scale population genomic variation data. Recently, there has been much interest in analyzing the joint SFS data from multiple populations to infer parameters of complex demographic histories, including variable population sizes, population split times, migration rates, admixture proportions, and so on. SFS-based inference methods require accurate computation of the expected SFS under a given demographic model. Although much methodological progress has been made, existing methods suffer from numerical instability and high computational complexity when multiple populations are involved and the sample size is large. In this article, we present new analytic formulas and algorithms that enable accurate, efficient computation of the expected joint SFS for thousands of individuals sampled from hundreds of populations related by a complex demographic model with arbitrary population size histories (including piecewise-exponential growth). Our results are implemented in a new software package called *momi* (MOran Models for Inference). Through an empirical study, we demonstrate our improvements to numerical stability and computational complexity.

An Augmented ADMM Algorithm With Application to the Generalized Lasso Problem P. 195-204

Yunzhang Zhu

Abstract

In this article, we present a fast and stable algorithm for solving a class of optimization problems that arise in many statistical estimation procedures, such as *sparse fused lasso over a graph*, *convex clustering*, and *trend filtering*, among others. We propose a so-called augmented *alternating direction methods of multipliers* (ADMM) algorithm to solve this class of problems. Compared to a standard ADMM algorithm, our proposal significantly reduces the computational cost at each iteration while maintaining roughly the same overall convergence speed. We also consider a new varying penalty scheme for the ADMM algorithm, which could further accelerate the convergence, especially when solving a sequence of problems with tuning parameters of different scales. Extensive numerical experiments on the sparse fused lasso problem show that the proposed algorithm is more efficient than the standard ADMM and two other existing state-of-the-art specialized algorithms. Finally, we discuss a possible extension and some interesting connections to two well-known algorithms. Supplementary materials for the article are available online.

Fast Tree Inference With Weighted Fusion Penalties P. 205-216

Julien Chiquet, Pierre Gutierrez & Guillem Rigaill

Abstract

Given a dataset with many features observed in a large number of conditions, it is desirable to fuse and aggregate conditions that are similar to ease the interpretation and extract the main characteristics of the data. This article presents a multidimensional fusion penalty framework to address this question when the number of conditions are large. If the fusion penalty is encoded by an ℓ_q-norm, we prove for uniform weights that the path of solutions is a tree that is suitable for interpretability. For the ℓ_1 and ℓ_∞-norms, the path is piecewise linear and we derive a homotopy algorithm to recover exactly the whole tree structure. For weighted ℓ_1-fusion penalties, we demonstrate that distance-decreasing weights lead to balanced tree structures. For a subclass of these weights that we call "exponentially adaptive," we derive an $\mathcal{O}(n\log(n))$ homotopy algorithm and we prove an asymptotic oracle property. This guarantees that we recover the underlying structure of the data efficiently both from a statistical and a computational point of view. We provide a fast implementation of the homotopy algorithm for the single feature case, as well as an efficient embedded cross-validation procedure that takes advantage of the tree structure of the path of solutions. Our proposal outperforms its competing procedures on simulations both in terms of timings and prediction accuracy. As an example we consider phenotypic data: given one or several traits, we reconstruct a balanced tree structure and assess its agreement with the known taxonomy. Supplementary materials for this article are available online.

Discrete Approximation of a Mixture Distribution via Restricted Divergence P. 217-222

Christian Röver & Tim Friede

Abstract

Mixture distributions arise in many application areas, for example, as marginal distributions or convolutions of distributions. We present a method of constructing an easily tractable discrete mixture distribution as an approximation to a mixture distribution with a large to infinite number, discrete or continuous, of components. The proposed DIRECT (divergence restricting conditional tesselation) algorithm is set up such that a prespecified precision, defined in terms of Kullback–Leibler divergence between true distribution and approximation, is guaranteed. Application of the algorithm is demonstrated in two examples. Supplementary materials for this article are available online.

Accurate Small Tail Probabilities of Sums of iid Lattice-Valued Random Variables via FFT P. 223-229

Huon Wilson & Uri Keich

Abstract

Accurately computing very small tail probabilities of a sum of independent and identically distributed lattice-valued random variables is numerically challenging. The only general purpose algorithms that can guarantee the desired accuracy have a quadratic runtime complexity that is often too slow. While fast Fourier transform (FFT)-based convolutions have an essentially linear runtime complexity, they can introduce overwhelming roundoff errors. We present sisFFT (segmented iterated shifted FFT), which harnesses the speed of FFT while retaining control of the relative error of the computed tail probability. We rigorously prove the method's accuracy and we empirically demonstrate its significant speed advantage over existing accurate methods. Finally, we show that sisFFT sacrifices very little, if any, speed when FFT-based convolution is sufficiently accurate to begin with. Supplementary material is available online.

Revista de fomento social, ISSN 0015-6043
Volumen 73/1, número 289 (2018)

Resumen
El trabajo es una dimensión humana esencial. Para muchas personas es su forma principal, si no única, de participación en la renta y riqueza del sistema productivo, de realización personal y de inserción social. El trabajo es mucho más que mercancía,aunque también se regule en el mercado laboral. El editorial parte de una consideración personalista del trabajo humano en la perspectiva de su contribución al bien común. En la segunda parte, descriptiva, el editorial trata de las características estructurales del mercado de trabajo en España. Tras esta presentación, en el apartado tercero se analiza el comportamiento coyuntural del mercado laboral durante la crisis. En el apartado cuarto,se trata de las dos recientes reformas llevadas a cabo por los gobiernos españoles de Zapatero (2010) y de Rajoy (2012). En las conclusiones, tras una breve consideración sobre el cambio epocal de la realidad del trabajo humano, se subrayan algunos aspectos sobre la cultura social, su base moral y la ética el trabajo, el papel de la educación y las nuevas formas de pensamiento sobre el trabajo.

Resumen
Este artículo, basado en el texto de la lección inaugural del curso 2017-2018 de la Universidad Loyola Andalucía, profundiza en una de las líneas de investigación prioritarias en psicología y neurociencia a nivel internacional: el impacto de la calidad de las relaciones humanas en la salud y en la mejora social. Bajo el título: Relaciones humanas de calidad: contexto de salud y libertad se comparten los principales hallazgos científicos centrados en cómo las relaciones humanas violentas perjudican la salud mental y física, incluso el desarrollo cerebral, y cómo, al contrario, las relaciones humanas de calidad como la amistad, garantizan un desarrollo cognitivo y emocional integral, saludable y positivo no sólo para uno mismo sino también para la humanidad.

Resumen
El autor, diputado del grupo Alianza progresista de socialistas y demócratas en el Parlamento europeo (2009¿2010 y desde 2014) parte de la reciente recuperación europea tras los años de fatiga atravesados por diversas brechas (norte¿sur, ste¿oeste),aprovechadas por poderosas fuerzas antieuropeas y por diversas contradicciones nacionales. Por primera vez en esa ¿historia de éxito¿ que ha sido el proceso europeo de construcción e integración, fallaba la lógica de aprovechar la crisis para avanzar, pues la narrativa europeísta y el espíritu de unidad en la paz y en el progreso que siempre animó a Europa han estado muy ausentes. Las crisis no son un obstáculo insalvable para Europa, que avanza precisamente a partir de las soluciones encontradas para aquéllas. Como decía Monnet, ¿Europa

se forja en las crisis¿. El autor las describe, así como el contexto y los signos de respuesta, y describe cinco pilares para el futuro de Europa en la gobernanza de la política económica, en el pilar social, en la resolución de la crisis migratoria, en la integración de la seguridad y la defensa europeas y en el refuerzo del mercado interior para lograr una economía competitiva. La crisis que ha supuesto el brexit para el proceso de construcción europea sirve como punto de apoyo para exponer la idea federalista europea, tras la que el autor describe y defiende una nueva narrativa europea.

I was a stranger and you welcomed meThe response of the Christian civil society to refugee protection in Europe P. 91-114

Amaya Valcárcel Silvela

Resumen

El propósito de este artículo es explicar la doctrina sobre la protección de los refugiados desde la perspectiva del pensamiento social católico y ofrecer un análisis comparativo de cómo las diferentes iniciativas inspiradas en los valores cristianos han respondido a los desafíos y oportunidades que plantean los refugiados. Finalmente, el texto ofrece algunas recomendaciones sobre las diferentes formas de avance. Este estudio compara algunas de estas iniciativas en toda Europa, centrándose particularmente en cinco países: Portugal, Francia, España, Bélgica e Italia. El estudio trata de responder a través de la visión tanto de aquellos que han sido alojados como personas desplazadas por la fuerza y como la de aquellos que los han albergado, así como desde la perspectiva de aquellos que unen comunidades y actúan juntos en nombre de los migrantes forzosos.

Migraciones, derechos humanos y vulnerabilidad P. 115-140

Carlos Arce Jiménez

Resumen

Los objetivos de este artículo son analizar desde una perspectiva multidisciplinar los retos de la movilidad humana en la era de la globalización, las vulneraciones de Derechos Humanos derivadas de las disfunciones de los cauces por los que debe desenvolverse y apuntar alternativas a este contexto. Indagamos en las causas de las migraciones contemporáneas, prestando especial atención al concepto de vulnerabilidad, a los conflictos y a los espacios de vulneración generalizada de los Derechos Humanos. Se realiza una revisión crítica de los instrumentos políticos–jurídicos que regulan a nivel global y regional la protección internacional y la libertad de circulación que deberían facilitar una movilidad humana por vías legales y seguras, y las vulneraciones de derechos en los flujos migratorios en tránsito y en frontera.

La antropología de Gaudium et spes P. 141-168

Mathias Nebel

El autor, teniendo en cuenta la sobreabundancia de estudios aparecidos hasta la fecha, y especialmente en el cincuentenario de la constitución pastoral conciliar Gaudium et spes, que han fijado la historia de la redacción y de los debates en los años conciliares, y puesto que vivimos y pensamos en contextos históricos muy diferentes, adopta una perspectiva fenomenológica y hermenéutica. El fracaso del propósito de diálogo con el mundo que ha intentado la Iglesia, a diferencia de lo que ocurría en los años 60 del siglo pasado, define la actual situación. Siendo el tono y el argumento del documento conciliar explícitamente teológico, su coherencia también lo es abiertamente y no pretende en ningún momento entablar un diálogo desde la razón secularizada, la lectura y recepción que sin embargo se han hecho por muchos para entablar un diálogo desde la razón natural se ha convertido en la historia de una infidelidad, pues los padres conciliares adoptaron una perspectiva netamente teológica para entablar ese diálogo con el mundo, ya que el argumento teológico debe ser propuesto como parte de la razón pública, pues, aunque su racionalidad sea transcendente no deja de ser comunicable e inteligible por otros interlocutores. El autor sospecha que el fracaso del diálogo secularizado puede deberse a la autocensura del discurso religioso en el terreno de lo público. En la segunda parte el autor subraya la necesidad del juicio moral y critica su ausencia en el discurso público y en la ética teológica,

juicio necesario para poder humanizar el mundo. Sólo mediante el discernimiento ético abierto por el discurso teológico sobre el mundo podremos dar razón del sentido de nuestra esperanza, lo que plantea la necesidad de rehabilitar el valor y el status de ese juicio ético.

Technometrics, ISSN 0040-1706
Volume 59, number 1 (february 2017)

Abstract

In deterministic computer experiments, it is often known that the output is a monotonic function of some of the inputs. In these cases, a monotonic metamodel will tend to give more accurate and interpretable predictions with less prediction uncertainty than a nonmonotonic metamodel. The widely used Gaussian process (GP) models are not monotonic. A recent article in *Biometrika* offers a modification that projects GP sample paths onto the cone of monotonic functions. However, their approach does not account for the fact that the GP model is more informative about the true function at locations near design points than at locations far away. Moreover, a grid-based method is used, which is memory intensive and gives predictions only at grid points. This article proposes the weighted projection approach that more effectively uses information in the GP model together with two computational implementations. The first is isotonic regression on a grid while the second is projection onto a cone of monotone splines, which alleviates problems faced by a grid-based approach. Simulations show that the monotone B-spline metamodel gives particularly good results. Supplementary materials for this article are available online.

Abstract

When fitting complex models, such as finite element or discrete event simulations, the experiment design should exhibit desirable properties of both projectivity and orthogonality. To reduce experimental effort, sequential design strategies allow experimenters to collect data only until some measure of prediction precision is reached. In this article, we present a batch sequential experiment design method that uses sliced full factorial-based Latin hypercube designs (sFFLHDs), which are an extension to the concept of sliced orthogonal array-based Latin hypercube designs (OALHDs). At all stages of the sequential design, good univariate stratification is achieved. The structure of the FFLHDs also tends to produce uniformity in higher dimensions, especially at certain stages of the design. We show that our batch sequential design approach has good sampling and fitting qualities through both empirical studies and theoretical arguments. Supplementary materials are available online.

Abstract

Most procedures that have been proposed to identify dispersion effects in unreplicated factorial designs assume that location effects have been identified correctly. Incorrect identification of location effects may impair subsequent identification of dispersion effects. We develop a method for joint identification of location and dispersion effects that

can reliably identify active effects of both types. A normal-based model containing parameters for effects in both the mean and variance is used. Parameters are estimated using maximum likelihood, and subsequent effect selection is done using a specially derived information criterion. An exhaustive search through a limited version of the space of possible models is conducted. Both a single-model output and model averaging are considered. The method is shown to be capable of identifying sensible location-dispersion models that are missed by methods that rely on sequential estimation of location and dispersion effects. Supplementary materials for this article are available online.

Abstract
Modeling a response over a nonconvex design region is a common problem in diverse areas such as engineering and geophysics. The tools available to model and design for such responses are limited and have received little attention. We propose a new method for selecting design points over nonconvex regions that is based on the application of multidimensional scaling to the geodesic distance. Optimal designs for prediction are described, with special emphasis on Gaussian process models, followed by a simulation study and an application in glaciology. Supplementary materials for this article are available online.

Abstract
Recent work in two-level screening experiments has demonstrated the advantages of using small foldover designs, even when such designs are not orthogonal for the estimation of main effects (MEs). In this article, we provide further support for this argument and develop a fast algorithm for constructing efficient two-level foldover (EFD) designs. We show that these designs have equal or greater efficiency for estimating the ME model versus competitive designs in the literature and that our algorithmic approach allows the fast construction of designs with many more factors and/or runs. Our compromise algorithm allows the practitioner to choose among many designs making a trade-off between efficiency of the main effect estimates and correlation of the two-factor interactions (2FIs). Using our compromise approach, practitioners can decide just how much efficiency they are willing to sacrifice to avoid confounded 2FIs as well as lowering an omnibus measure of correlation among the 2FIs.

Abstract
Computer experiments based on mathematical models are powerful tools for understanding physical processes. This article addresses the problem of kriging-based optimization for deterministic computer experiments with tunable accuracy. Our approach is to use multi-fidelity computer experiments with increasing accuracy levels and a nonstationary Gaussian process model. We propose an optimization scheme that sequentially adds new computer runs by following two criteria. The first criterion, called EQI, scores candidate inputs with given level of accuracy, and the second criterion, called EQIE, scores candidate combinations of inputs and accuracy. From simulation results and a real example using finite element analysis, our method outperforms the expected improvement (EI) criterion that works for single-accuracy experiments. Supplementary materials for this article are available online.

Two-Level Designs to Estimate All Main Effects and Two-Factor Interactions P. 69-79

Pieter T. Eendebak & Eric D. Schoen

Abstract

We study the design of two-level experiments with N runs and n factors large enough to estimate the interaction model, which contains all the main effects and all the two-factor interactions. Yet, an effect hierarchy assumption suggests that main effect estimation should be given more prominence than the estimation of two-factor interactions. Orthogonal arrays (OAs) favor main effect estimation. However, complete enumeration becomes infeasible for cases relevant for practitioners. We develop a partial enumeration procedure for these cases and we establish upper bounds on the D-efficiency for the interaction model based on arrays that have not been generated by the partial enumeration. We also propose an optimal design procedure that favors main effect estimation. Designs created with this procedure have smaller D-efficiencies for the interaction model than D-optimal designs, but standard errors for the main effects in this model are improved. Generated OAs for 7–10 factors and 32–72 runs are smaller or have a higher D-efficiency than the smallest OAs from the literature. Designs obtained with the new optimal design procedure or strength-3 OAs (which have main effects that are not correlated with two-factor interactions) are recommended if main effects unbiased by possible two-factor interactions are of primary interest. D-optimal designs are recommended if interactions are of primary interest. Supplementary materials for this article are available online.

Calibration of Stochastic Computer Simulators Using Likelihood Emulation P. 80-92

Jeremy E. Oakley & Benjamin D. Youngman

Abstract

We calibrate a stochastic computer simulation model of “moderate” computational expense. The simulator is an imperfect representation of reality, and we recognize this discrepancy to ensure a reliable calibration. The calibration model combines a Gaussian process emulator of the likelihood surface with importance sampling. Changing the discrepancy specification changes only the importance weights, which lets us investigate sensitivity to different discrepancy specifications at little computational cost. We present a case study of a natural history model that has been used to characterize UK bowel cancer incidence. Datasets and computer code are provided as supplementary material.

Nonstationary Gaussian Process Models Using Spatial Hierarchical Clustering from Finite Differences P. 93-101

Matthew J. Heaton, William F. Christensen & Maria A. Terres

Abstract

Modern digital data production methods, such as computer simulation and remote sensing, have vastly increased the size and complexity of data collected over spatial domains. Analysis of these large spatial datasets for scientific inquiry is typically carried out using the Gaussian process. However, nonstationary behavior and computational requirements for large spatial datasets can prohibit efficient implementation of Gaussian process models. To perform computationally feasible inference for large spatial data, we consider partitioning a spatial region into disjoint sets using hierarchical clustering of observations and finite differences as a measure of dissimilarity. Intuitively, directions with large finite differences indicate directions of rapid increase or decrease and are, therefore, appropriate for partitioning the spatial region. Spatial contiguity of the resulting clusters is enforced by only clustering Voronoi neighbors. Following spatial clustering, we propose a nonstationary Gaussian process model across the clusters, which allows the computational burden of model fitting to be distributed across multiple cores and nodes. The methodology is primarily motivated and illustrated by an application to the validation of digital temperature data over the city of Houston as well as simulated datasets. Supplementary materials for this article are available online.

Anomaly Detection in Images With Smooth Background via Smooth-Sparse Decomposition P. 102-114

Hao Yan, Kamran Paynabar & Jianjun Shi

Abstract

In various manufacturing applications such as steel, composites, and textile production, anomaly detection in noisy images is of special importance. Although there are several methods for image denoising and anomaly detection, most of these perform denoising and detection sequentially, which affects detection accuracy and efficiency. Additionally, the low computational speed of some of these methods is a limitation for real-time inspection. In this article, we develop a novel methodology for anomaly detection in noisy images with smooth backgrounds. The proposed method, named smooth-sparse decomposition, exploits regularized high-dimensional regression to decompose an image and separate anomalous regions by solving a large-scale optimization problem. To enable the proposed method for real-time implementation, a fast algorithm for solving the optimization model is proposed. Using simulations and a case study, we evaluate the performance of the proposed method and compare it with existing methods. Numerical results demonstrate the superiority of the proposed method in terms of the detection accuracy as well as computation time. This article has supplementary materials that includes all the technical details, proofs, MATLAB codes, and simulated images used in the article.

Estimation of Field Reliability Based on Aggregate Lifetime Data P. 115-125

Piao Chen & Zhi-Sheng Ye

Abstract

Because of the exponential distribution assumption, many reliability databases recorded data in an aggregate way. Instead of individual failure times, each aggregate data point is a summation of a series of collective failures representing the cumulative operating time of one component position from system commencement to the last component replacement. The data format is different from traditional lifetime data and the statistical inference is challenging. We first model the individual component lifetime by a gamma distribution. Confidence intervals for the gamma shape parameter can be constructed using a scaled χ^2 approximation to a modified ratio of the geometric mean to the arithmetic mean, while confidence intervals for the gamma rate and mean parameters, as well as quantiles, are obtained using the generalized pivotal quantity method. We then fit the data using the inverse Gaussian (IG) distribution, a useful lifetime model for failures caused by degradation. Procedures for point estimation and interval estimation of parameters are developed. We also propose an interval estimation method for the quantiles of an IG distribution based on the generalized pivotal quantity method. An illustrative example demonstrates the proposed inference methods. Supplementary materials for this article are available online.

Technometrics, ISSN 0040-1706
Volume 59, number 4 (november 2017)

Abstract
Large-scale data analysis problems have become increasingly common across many disciplines. While large volume of data offers more statistical power, it also brings computational challenges. The orthogonalizing expectation–maximization (EM) algorithm by Xiong et al. is an efficient method to deal with large-scale least-square problems from a design point of view. In this article, we propose a reformulation and generalization of the orthogonalizing EM algorithm. Computational complexity and convergence guarantees are established. The reformulation of the orthogonalizing EM algorithm leads to a reduction in computational complexity for least-square problems and penalized least-square problems. The reformulation, named the GOEM (generalized orthogonalizing EM) algorithm, can incorporate a wide variety of convex and nonconvex penalties, including the lasso, group lasso, and minimax concave penalty penalties. The GOEM algorithm is further extended to a wider class of models including generalized linear models and Cox's proportional hazards model. Synthetic and real data examples are included to illustrate its use and efficiency compared with standard techniques. Supplementary materials for this article are available online.

Abstract
Partial least squares (PLS) is a prominent solution for dimension reduction and high-dimensional regressions. Recent prevalence of multidimensional tensor data has led to several tensor versions of the PLS algorithms. However, none offers a population model and interpretation, and statistical properties of the associated parameters remain intractable. In this article, we first propose a new tensor partial least-squares algorithm, then establish the corresponding population interpretation. This population investigation allows us to gain new insight on how the PLS achieves effective dimension reduction, to build connection with the notion of sufficient dimension reduction, and to obtain the asymptotic consistency of the PLS estimator. We compare our method, both analytically and numerically, with some alternative solutions. We also illustrate the efficacy of the new method on simulations and two neuroimaging data analyses. Supplementary materials for this article are available online.

Abstract
Group sparse approaches to regression modeling are finding ever increasing utility in an array of application areas. While group sparsity can help assess certain data structures, it is desirable in many instances to also capture element-wise sparsity. Recent work exploring the latter has been conducted in the context of l_2/l_1 penalized regression in the form of the sparse group lasso (SGL). Here, we present a novel model, called the sparse group elastic net (SGEN),

which uses an l_∞/l_1/ridge-based penalty. We show that the l_∞-norm, which induces group sparsity is particularly effective in the presence of noisy data. We solve the SGEN model using a coordinate descent-based procedure and compare its performance to the SGL and related methods in the context of hyperspectral imaging in the presence of noisy observations. Supplementary materials for this article are available online.

Split-Plot and Multi-Stratum Designs for Statistical Inference P. 446-457

Luzia A. Trinca & Steven G. Gilmour

Abstract

It is increasingly recognized that many industrial and engineering experiments use split-plot or other multi-stratum structures. Much recent work has concentrated on finding optimum, or near-optimum, designs for estimating the fixed effects parameters in multi-stratum designs. However, often inference, such as hypothesis testing or interval estimation, will also be required and for inference to be unbiased in the presence of model uncertainty requires pure error estimates of the variance components. Most optimal designs provide few, if any, pure error degrees of freedom. Gilmour and Trinca (2012 Gilmour, S. G., and Trinca, L. A. (2012), "Optimum Design of Experiments for Statistical Inference" (with discussion), *Applied Statistics*, 61, 345–401.[Crossref], [Web of Science ®], [Google Scholar]) introduced design optimality criteria for inference in the context of completely randomized and block designs. Here these criteria are used stratum-by-stratum to obtain multi-stratum designs. It is shown that these designs have better properties for performing inference than standard optimum designs. Compound criteria, which combine the inference criteria with traditional point estimation criteria, are also used and the designs obtained are shown to compromise between point estimation and inference. Designs are obtained for two real split-plot experiments and an illustrative split–split-plot structure. Supplementary materials for this article are available online.

Bayesian Design of Experiments Using Approximate Coordinate Exchange P. 458-470

Antony M. Overstall & David C. Woods

Abstract

The construction of decision-theoretical Bayesian designs for realistically complex nonlinear models is computationally challenging, as it requires the optimization of analytically intractable expected utility functions over high-dimensional design spaces. We provide the most general solution to date for this problem through a novel approximate coordinate exchange algorithm. This methodology uses a Gaussian process emulator to approximate the expected utility as a function of a single design coordinate in a series of conditional optimization steps. It has flexibility to address problems for any choice of utility function and for a wide range of statistical models with different numbers of variables, numbers of runs and randomization restrictions. In contrast to existing approaches to Bayesian design, the method can find multi-variable designs in large numbers of runs without resorting to asymptotic approximations to the posterior distribution or expected utility. The methodology is demonstrated on a variety of challenging examples of practical importance, including design for pharmacokinetic models and design for mixed models with discrete data. For many of these models, Bayesian designs are not currently available. Comparisons are made to results from the literature, and to designs obtained from asymptotic approximations. Supplementary materials for this article are available online.

Robust Parameter Designs in Computer Experiments Using Stochastic Approximation P. 471-483

Weijie Shen

Abstract

Robust parameter designs are widely used to produce products/processes that perform consistently well across various conditions known as noise factors. Recently, the robust parameter design method is implemented in computer experiments. The structure of conventional product array design becomes unsuitable due to its extensive number of

runs and the polynomial modeling. In this article, we propose a new framework robust parameter design via stochastic approximation (RPD-SA) to efficiently optimize the robust parameter design criteria. It can be applied to general robust parameter design problems, but is particularly powerful in the context of computer experiments. It has the following four advantages: (1) fast convergence to the optimal product setting with fewer number of function evaluations; (2) incorporation of high-order effects of both design and noise factors; (3) adaptation to constrained irregular region of operability; (4) no requirement of statistical analysis phase. In the numerical studies, we compare RPD-SA to the Monte Carlo sampling with Newton–Raphson-type optimization. An "Airfoil" example is used to compare the performance of RPD-SA, conventional product array designs, and space-filling designs with the Gaussian process. The studies show that RPD-SA has preferable performance in terms of effectiveness, efficiency and reliability.

Phase I Distribution-Free Analysis of Multivariate Data P. 484-495

Giovanna Capizzi & Guido Masarotto

Abstract

In this study, a new distribution-free Phase I control chart for retrospectively monitoring multivariate data is developed. The suggested approach, based on the multivariate signed ranks, can be applied to individual or subgrouped data for detection of location shifts with an arbitrary pattern (e.g., isolated, transitory, sustained, progressive, etc.). The procedure is complemented with a LASSO-based post-signal diagnostic method for identification of the shifted variables. A simulation study shows that the method compares favorably with parametric control charts when the process is normally distributed, and largely outperforms other multivariate nonparametric control charts when the process distribution is skewed or heavy-tailed. An R package can be found in the supplementary material.

Statistical Process Control for Latent Quality Characteristics Using the Up-and-Down Test P. 496-507

Dongdong Xiang, Fugee Tsung & Xiaolong Pu

Abstract

In many applications, the quality characteristic of a product is continuous but unobservable, for example, the critical electric voltage of electro-explosive devices. It is often important to monitor a manufacturing process of a product with such latent quality characteristic. Existing approaches all involve specifying a fixed stimulus level and testing products under that level to collect a sequence of response outcomes (zeros or ones). Appropriate control charts are then applied to the collected binary data sequence. However, these approaches offer limited performance. Moreover, the collected dataset provides little information for troubleshooting when an out-of-control signal is triggered. To overcome these limitations, this article introduces the up-and-down test for collecting data and proposes a new control chart based on this test. Numerical studies show that the proposed chart is able to detect any shifts effectively and is robust in many situations. Finally, an example involving real manufacturing data is given to demonstrate the use of our proposed chart.

A Vine Copula Model for Predicting the Effectiveness of Cyber Defense Early-Warning P. 508-520

Maochao Xu, Lei Hua & Shouhuai Xu

Abstract

Internet-based computer information systems play critical roles in many aspects of modern society. However, these systems are constantly under cyber attacks that can cause catastrophic consequences. To defend these systems effectively, it is necessary to measure and predict the effectiveness of cyber defense mechanisms. In this article, we investigate how to measure and predict the effectiveness of an important cyber defense mechanism that is known as *early-warning*. This turns out to be a challenging problem because we must accommodate the *dependence* among certain four-dimensional time series. In the course of using a dataset to demonstrate the prediction methodology, we

discovered a new *nonexchangeable* and *rotationally symmetric* dependence structure, which may be of independent value. We propose a new vine copula model to accommodate the newly discovered dependence structure, and show that the new model can predict the effectiveness of early-warning more accurately than the others. We also discuss how to use the prediction methodology in practice.

Hierarchical Spatially Varying Coefficient Process Model P. 521-527

Heeyoung Kim & Jaehwan Lee

Abstract

The spatially varying coefficient process model is a nonstationary approach to explaining spatial heterogen-eity by allowing coefficients to vary across space. In this article, we develop a methodology for generalizing this model to accommodate geographically hierarchical data. This article considers two-level hierarchical structures and allow for the coefficients of both low-level and high-level units to vary over space. We assume that the spatially varying low-level coefficients follow the multivariate Gaussian process, and the spatially varying high-level coefficients follow the multivariate simultaneous autoregressive model that we develop by extending the standard simultaneous autoregressive model to incorporate multivariate data. We apply the proposed model to transaction data of houses sold in 2014 in a part of the city of Los Angeles. The results show that the proposed model predicts housing prices and fits the data effectively.

Minimum Distance Estimation for the Generalized Pareto Distribution P. 528-541

Piao Chen, Zhi-Sheng Ye & Xingqiu Zhao

Abstract

The generalized Pareto distribution (GPD) is widely used for extreme values over a threshold. Most existing methods for parameter estimation either perform unsatisfactorily when the shape parameter k is larger than 0.5, or they suffer from heavy computation as the sample size increases. In view of the fact that $k > 0.5$ is occasionally seen in numerous applications, including two illustrative examples used in this study, we remedy the deficiencies of existing methods by proposing two new estimators for the GPD parameters. The new estimators are inspired by the minimum distance estimation and the *M*-estimation in the linear regression. Through comprehensive simulation, the estimators are shown to perform well for all values of k under small and moderate sample sizes. They are comparable to the existing methods for $k < 0.5$ while perform much better for $k > 0.5$.