**INTERNATIONAL WORKSHOP**

# Model-Oriented Data Analysis and Optimum Design

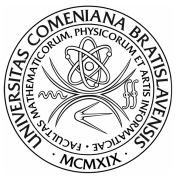
# mODa 12

**ABSTRACTS**

June 23 – 28, 2019, Smolenice Castle, Slovakia

mODa 12
Model-Oriented Data Analysis and Optimum Design
June 23 – 28, 2019
Smolenice Castle, Slovak Republic

LOCAL ORGANIZERS
Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava
Institute of Measurement Science of the Slovak Academy of Sciences





PUBLISHER
Faculty of Mathematics, Physics and Informatics,
Comenius University in Bratislava, Slovakia
Bratislava, June 2019

**mODa board**

Barbara Bogacka
Valerii V. Fedorov
Alessandra Giovagnoli
Radoslav Harman
Joachim Kunert
Jesús López-Fidalgo
Christine Müller
Werner G. Müller
Luc Pronzato
Dariusz Uciński
Dave Woods
Henry P. Wynn
Anatoly A. Zhigljavsky

**Local organizing commitee**

Radoslav Harman
Viktor Witkovský
Lenka Filová
Martina Chvosteková
Eva Gurišová
Andrej Dvurečenskij
Samuel Rosa

**Scientific Programme**

| | | | |
|---|---|---|---|
| **Monday 24 June** | | | |
| 07:30-08:45 | **Breakfast** | 14:00-14:30 | F. Freise |
| 08:45-09:00 | **Opening** | 14:30-15:00 | V. Casero-Alonso |
| 09:00-09:30 | R. A. Bailey | 15:00-15:30 | S. G. M. Biedermann |
| 09:30-10:00 | K. Filipiak | 15:30-16:00 | **Coffee break** |
| 10:00-10:30 | R. Singh | 16:00-16:30 | M. Prus |
| 10:30-11:00 | B. A. Jones | 16:30-17:00 | F. Röttger |
| 11:00-11:30 | **Coffee break** | 17:00-17:30 | K. Mylona |
| 11:30-12:00 | A. Teckentrup | 17:30-17:35 | **mODa13 information** |
| 12:00-12:30 | D. Ginsbourger | 18:30-19:30 | **Dinner** |
| 12:30-13:30 | **Lunch** | | |
| **Tuesday 25 June** | | | |
| 07:30-09:00 | **Breakfast** | 14:00-14:30 | A. Pepelyshev |
| 09:00-09:30 | W. G. Müller | 14:30-15:00 | S. L. Leonov |
| 09:30-10:00 | A. Zhigljavsky | 15:00-15:30 | D. Woods |
| 10:00-10:30 | A. Pázman | 15:30-16:00 | **Coffee break** |
| 10:30-11:00 | **Coffee break** | 16:00-16:30 | W. F. Rosenberger |
| 11:00-11:30 | A. C. Atkinson | 16:30-17:00 | A. C. Lane |
| 11:30-12:00 | M. Radloff | 17:00-17:30 | R.-D. Hilgers |
| 12:00-12:30 | V. Fedorov | 17:30-18:30 | **Poster session** |
| 12:30-12:45 | **Photo** | 18:30-19:30 | **Dinner** |
| 12:45-13:30 | **Lunch** | | |
| **Wednesday 26 June** | | | |
| 07:30-08:30 | **Breakfast** | 11:00-11:30 | A. C. Hooker |
| 08:30-09:00 | T. Mielke | 11:30-12:00 | N. Malevich |
| 09:00-09:30 | A. Ivanova | 12:00-12:30 | L. Pronzato |
| 09:30-10:00 | C. May | 12:30-13:30 | **Lunch** |
| 10:00-10:30 | N. Flournoy | 14:00-17:30 | **Excursion** |
| 10:30-11:00 | **Coffee break** | 18:30-19:30 | **Dinner** |

**Thursday 27 June**

| | | | |
|---|---|---|---|
| 07:30-09:00 | **Breakfast** | 14:00-14:30 | E. Pesce |
| 09:00-09:30 | G. Sagnol | 14:30-15:00 | L. Deldossi |
| 09:30-10:00 | S. Rosa | 15:00-15:30 | H. Wang |
| 10:00-10:30 | S. D. Ahipasaoglu | 15:30-16:00 | **Coffee break** |
| 10:30-11:00 | D. Uciński | 16:00-16:30 | J. Muré |
| 11:00-11:30 | **Coffee break** | 16:30-17:00 | M. Hainy |
| 11:30-12:00 | F. Rapallo | 17:00-17:30 | T. W. Waite |
| 12:00-12:30 | P. P. Goos | 17:30-18:30 | **Future of OD** |
| 12:30-13:30 | **Lunch** | 19:30-21:00 | **Conference dinner** |

**Friday 28 June**

| | | | |
|---|---|---|---|
| 07:30-09:00 | **Breakfast** | 10:30-11:00 | J. López-Fidalgo |
| 09:00-09:30 | H. Großmann | 11:00-11:15 | **Closing** |
| 09:30-10:00 | M. Vandebroek | 11:15-12:00 | **Lunch** |
| 10:00-10:30 | X. Xu | 12:00 | Bus to Bratislava |

**List of Posters**

E. Benková
A. Cakiroglu
M. Chvosteková
C. de la Calle Arroyo
L. Filová
R. Fontana
I. García Camacha Gutiérrez
I. Mariñas-Collado
H. Maruri-Aguilar
S. Pozuelo-Campos
A. Rappold
J. M. Rodríguez-Díaz
T. Schmelter
B. Spangl
M. S. Tackney
V. Witkovský
N. Youssef

# Contents

# Abstracts

## EXACT OPTIMAL DESIGNS, MINIMUM VOLUME ESTIMATOR, AND ROBUST OPTIMIZATION

SELIN D. AHIPASAOGLU*
*Singapore Univ. Tech. and Design*

Calculating an exact D-optimal design over a finite set of points remains a very challenging task even though approximate designs can be found efficiently. We develop a new branch-and-bound framework for solving the exact D-optimal design problem and discuss how it can be implemented efficiently. In each node of the branch-and-bound tree, we solve a generalization of the D-optimal design problem via a simple yet effective first-order algorithm. We show how the generalised design problem can also be used to find good solutions for the Minimum Volume Estimator problem, which itself is not trivial. Finally, we demonstrate the importance of the MVE estimator and exact designs for constructing ellipsoidal uncertainty sets in robust optimization.

## DESIGNS FOR COMPARTMENTAL MODELS WITH A MICHAELIS-MENTEN ELIMINATION RATE

BELMIRO P. M. DUARTE
*Instituto Politécnico de Coimbra, Portugal*
and
ANTHONY C. ATKINSON*
*London School of Economics, UK*

The optimal design of experiments for the parameters of the differential equations arising in many kinetic and pharmacodynamic models requires the numerical maximization of a function of the numerical solutions of sets of differential equations, the solutions depending on the experimental design.

This problem in the design of experiments is solved as a Dynamic Optimization problem, using a simultaneous approach. The time horizon of the experiments is discretized in finite elements and Orthogonal Collocation on Finite Elements is used to parameterize the solution. The solution in each finite element is described by cubic Lagrange polynomials and the control action is represented by piecewise constant polynomials.

We find D-optimum designs for two- and three-compartment models with Michaelis-Menten elimination rate kinetics. We consider three different design problems including

both dynamic and static experiments. Importantly, our algorithm allows the calculation of designs with constraints not only on the design region but also on the concentrations of the components in the compartments and on the rates of change of the design variables and of the resulting concentrations.

## MULTI-PART BALANCED INCOMPLETE-BLOCK DESIGNS

ROSEMARY A. BAILEY*
*University of St Andrews, UK*

Fedorov and Leonov [2] describe cancer trials which allow each medical centre to treat only a limited number of cancer types with only a limited number of drugs. How should such trials be designed?

We specify some desirable combinatorial properties of these designs, and prove some consequences. We give several different constructions. Then we generalize this to three or more factors, such as biomarkers, and show how these methods are related to previous work by Mukerjee [3].

This is joint work with P. J. Cameron, reported in [1].

### References

[1] BAILEY, R. A. and CAMERON, P. J. (2019). Multi-part balanced incomplete-block designs. *Statistical Papers* **60**, 55–76.

[2] FEDOROV, V. V. and LEONOV, S. L. (2019). Combinatorial and model-based methods in structuring and optimizing cluster trials. In: Beckman RA, Antonijevic Z. (eds) *Platform Trials in Drug Development: Umbrella Trials and Basket Trials*, Chapman & Hall/CRC Press, Boca Raton, pp. 265–286.

[3] MUKERJEE, R. (1998). On balanced orthogonal multi-arrays: Existence, construction and application to design of experiments. *Journal of Statistical Planning and Inference* **73**, 149–162.

## MINIMUM-DISTANCE PRIVACY SETS

EVA BENKOVÁ*
*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava*
*Johannes Kepler University Linz*

RADOSLAV HARMAN
*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava*
and

WERNER G. MÜLLER
*Johannes Kepler University Linz*

Utilizing a typology for space-filling into "soft" and "hard" methods, the paper [1] introduces the central notion of "privacy sets" for dealing with the latter. This notion provides a unifying framework for standard designs without replication, Latin hypercube designs, and Bridge designs, among many others. The privacy sets presented here guarantee some minimum distance between any two design points, which spreads out the measurements across the design space in a very natural way. We propose an exchange-type heuristic method dealing with this type of constraints, elaborating on the ideas of [1]. Some of the key steps of the algorithm are based on the computing of Voronoi diagrams and their generalizations.

### References

[1] BENKOVÁ, E., HARMAN, R. and MÜLLER, W. G. (2016). Privacy sets for constrained space-filling. *Journal of Statistical Planning and Inference* **171**, 1–9.

## FRACTIONAL POLYNOMIAL MODELS FOR CONSTRAINED MIXTURE EXPERIMENTS

STEFANIE BIEDERMANN*
*University of Southampton, UK*
RANA KHASHAB
*Umm Al-Qura University, Saudi Arabia*
and
STEVEN GILMOUR
*King's College London, UK*

Experiments involving mixtures are conducted in a variety of areas, for example in food processing or in chemical research. The experimental region is constrained naturally, as the proportions of all ingredients have to sum to one. Additional constraints may arise when there are bounds on the proportions, for example, a cake must contain a minimum percentage of flour to have the right texture and flavour. Khashab et al. [1] propose a new - parsimonious but flexible - class of non-linear models, based on fractional polynomials [2], to fit the data from constrained mixture experiments. In the talk, I will motivate this modelling approach, and will use a number of historical data sets to compare these models with various other models suggested in the literature. I will then present some optimal designs for these models, and will further discuss some general issues related to designing experiments for mixtures.

### References

[1] KHASHAB, R., GILMOUR, S. and BIEDERMANN, S. (2019). Fractional polynomial models for constrained mixture experiments. Working paper.

[2] ROYSTON, P. and ALTMAN, D. G. (1994). Regression using fractional polynomials of continuous covariates. *Journal of the Royal Statistical Society* **43**, 429–467.

## SELECTIVE RECRUITMENT DESIGNS FOR OBSERVATIONAL STUDIES USING ELECTRONIC HEALTH RECORDS

JAMES BARRETT
*King's College London*
AYLIN CAKIROGLU*
*The Francis Crick Institute*
CATEY BUNCE
*King's College London*
ANOOP SHAH
*University College London*
and
SPIROS DENAXAS
*University College London*

Observational studies traditionally wait for participants to accrue at designated study centres. The advent of large-scale electronic health records (EHRs) presents an opportunity to quickly identify eligible individuals and directly invite them to participate in a study. EHRs typically contain data from even millions of individuals, raising the question of how to optimally select a cohort of individuals from a larger group. In this paper we propose a simple selective-recruitment protocol that identifies a cohort of patients in which the covariates of interest have a uniform (or close to uniform) distribution. Using numerical simulations we show that if there are covariate imbalances in the source population, then selectively recruited target populations offer a greater level of statistical power in comparison to randomly selected cohorts. Our protocol can be applied to categorical and continuous covariates, and we propose a pragmatic method for dealing with multiple covariates. We apply our selective-recruitment protocol to a simulated prospective observational study based on an EHR-derived cohort of stable acute coronary disease patients containing cardiovascular risk factors and biomarkers from 82,089 individuals in the U.K. We illustrate that selectively recruited cohorts offer greater statistical power and more accurate parameter estimates than randomly selected cohorts. Thus, when larger-scale EHR sources of eligible patients are available, approaches on a priori identifying optimal cohorts of patients can potentially reduce the amount of resources and time required to recruit individuals to investigator-led studies.

## OPTIMAL EXPERIMENTAL DESIGNS FOR FRACTIONAL POLYNOMIAL MODELS

VÍCTOR CASERO-ALONSO*
*University of Castilla-La Mancha*
JESÚS LÓPEZ-FIDALGO
*University of Castilla-La Mancha and Navarra University*
and
WENG KEE WONG
*UCLA*

Fractional polynomials (FP) have been shown to be more flexible than polynomials for fitting a continuous outcome in the biomedical sciences but design issues for such models are not well addressed. We construct D-optimal designs and I-optimal designs for estimating model parameters and prediction in FP models for single and multi-factor models. We also construct an applet to facilitate users find various types of tailor made optimal designs for his or her FP model. We re-design studies using FP models and show that optimal designs can produce substantial gains in terms of cost and quality of the statistical inference.

## CALIBRATION EXPERIMENT FOR MULTIVARIATE STATISTICAL CALIBRATION

MARTINA CHVOSTEKOVÁ*
*Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovakia*

We numerically investigated the impact of the design of the calibration experiment on the size and confidence level of two types of multiple use confidence regions. The multiple use confidence regions for the multivariate calibration problem are constructed for unknown and possibly different values of an explanatory variable, say $x$, after observing a sequence of observations of a response variable, say $Y_x$, related to the explanatory variable through a multivariate linear regression model. The unknown parameters of the model are estimated based on calibration data from a calibration experiment. It is required that at least $\gamma$ proportion of the multiple use confidence regions constructed by using the same estimated regression curve will contain the corresponding true $x$ value and the probability that the calibration data will provide such coverage is at least $1 - \alpha$.

Our results show that the confidence level and the size of the multiple use confidence regions based on the classical estimator of the unknown explanatory value are affected by the design of the calibration experiment in contrast to the confidence level and the size of the multiple use confidence regions based on the condition of the tolerance region in linear regression.

**Acknowledgements** This research was supported by the Slovak Research and Development Agency, project APVV-15-0295, and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, projects VEGA 2/0054/18 and VEGA 2/0081/19.

*Keywords:* Multivariate Calibration, Multiple-use Confidence Regions

## D- AND I-OPTIMAL DESIGNS FOR THE ANTOINE'S EQUATION IN DISTILLATION EXPERIMENTS

CARLOS DE LA CALLE-ARROYO*
*Universidad de Castilla-La Mancha*
JESÚS LÓPEZ-FIDALGO
*Universidad de Navarra*
and
RODRÍGUEZ-ARAGÓN, LICESIO
*Universidad de Castilla-La Mancha*

In the distillation processes it is very important to know precisely the relationship between temperature and vapor pressure. The vapor pressures not only depend on the temperature but vary enormously for different substances. The Antoine equation is a hyperbolic equation, a class of semi-empirical correlations describing the relationship between temperature and vapor pressure very precisely for certain temperature ranges. In this work, the study of the optimal designs for the estimation of the parameters of Antoine equation is presented. An analytical solution for the D-Optimal designs is shown, as well as numerically calculated designs for different criteria. Of special interest for this equation is the work done in I-Optimality, as there is usually some overlap between models in the boundary regions of the space of the design. The results presented are a first step in our exploration of the Optimal Experimental Design for the Antoine's Equation.

## OPTIMAL DESIGN OF EXPERIMENTS AND MODEL-BASED SURVEY SAMPLING IN BIG DATA

LAURA DELDOSSI*
*Department of Statistical Science, Università Cattolica del Sacro Cuore, Milano*
and
CHIARA TOMMASI
*Department of Economics, Management and Quantitative Methods, University of Milan*

Big Data are generally huge quantities of digital information accrued automatically and/or merged from several sources and rarely result from properly planned population surveys. A Big Dataset is herein conceived as a collection of information concerning a finite population. Since the analysis of an entire Big Dataset can require enormous computational effort, we suggest selecting a sample of observations and using this sampling information to achieve the inferential goal. Instead of the design-based survey sampling approach (which relates to the estimation of summary finite population measures, such as means, totals, proportions) we consider the model-based sampling approach, which involves inference about parameters of a super-population model. This model is assumed to have generated the finite population values, i.e. the Big Dataset. Given a super-population model we can apply the theory of optimal design to draw a sample from the Big Dataset which contains the majority of information about the unknown parameters of interest. In addition, since a Big Dataset might provide poor information despite its size, from the definition of efficiency of a design we suggest a device to measure the quality of the Big Data.

## OPTIMAL DESIGNS FOR VARIOUS LOGISTIC MODELS

VALERII V. FEDOROV*
*ICON plc, North Wales, USA*

After a short review of classical results the presentation will be focused on two four parameter logistic models with continuous responses. The first model contains an additive observational error. In the second one an observational error appears in the exponential term(s). Both models can be viewed as partially nonlinear regression models. This fact simplifies the optimization problem and allows to explicitly find the support points of locally D-optimal designs for cases with a single predictor. The approach is based on the

mapping of the design set to the set of expected responses. This method does not work in multivariate cases as well as in a univariate one but it still leads to some useful insights.

## OPTIMALITY OF BLOCK DESIGNS UNDER THE MODEL WITH THE FIRST-ORDER CIRCULAR AUTOREGRESSION

KATARZYNA FILIPIAK*
*Institute of Mathematics, Poznań University of Technology, Poland*
RAZIEH KHODSIANI
*Department of Mathematical Sciences, Isfahan University of Technology, Iran*
and
AUGUSTYN MARKIEWICZ
*Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poland*

In this talk optimal properties of some circular balanced block designs under the model with circular autoregression of order one are studied. Universal optimality of some balanced block designs with equal block sizes is proven and E-optimality of complete balanced block designs with the number of blocks equal to the number of treatments or the number of treatments reduced by two is shown.

**Acknowledgements** This research is partially supported by Scientific Activities (Poznań University of Technology) No. 04/43/SBAD/0115 (K. Filipiak).

*Keywords:* Circular autoregression of order one, Circular balanced block design, Universal optimality, E-optimality, Nondirectionally neighbor balanced design

### References

[1] FILIPIAK, K., KHODSIANI, R. and MARKIEWICZ, A. (2019). Optimality of block designs under themodel with the first-order circular autoregression. *Statistical Papers* **60**, 77–97.

## R PACKAGE OPTIMALDESIGN

LENKA FILOVÁ*
*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava*
and
RADOSLAV HARMAN
*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava*

We present computational and graphical capabilities of our R ([5]) package OptimalDesign which provides a toolbox for the computation of D-, A- and I-efficient exact and approximate designs of experiments on finite domains, for regression models with real-valued, uncorrelated observations. The package fills a gap in presently available functions for experimental design optimization by implementing several competing algorithms based on significantly different principles, including mathematical programming methods (see, e.g., [3], [6]), iterative approaches ([7], [4]) and search heuristics ([3], [2]). This allows the user to cross-check the quality of the results, and, in many cases, provides efficient design alternatives to choose for the experiment. An important feature of the package is that several implemented procedures allow for multiple linear constraints on the vector of design weights.

*Keywords:* R, Mathematical programming, Design of experiments, Constrained optimization

### References

[1] ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum experimental designs, with SAS*. Oxford University Press, Oxford.

[2] HARMAN, R., BACHRATÁ, A. and FILOVÁ, L. (2016). Construction of efficient experimental designs under multiple resource constraints. *Applied Stochastic Models in Business and Industry* **32**, 3–17.

[3] HARMAN, R. and FILOVÁ, L. (2014). Computing efficient exact designs of experiments using integer quadratic programming. *Computational Statistics & Data Analysis* **71**, 1159–1167.

[4] HARMAN, R., FILOVÁ, L. and RICHTÁRIK, P. (2019). A Randomized Exchange Algorithm for Computing Optimal Approximate Designs of Experiments. *Journal of the American Statistical Association*, to appear.

[5] R DEVELOPMENT CORE TEAM (2011). R: A Language and Environment for Statistical Computing. Foundation for Statistical Computing, Vienna, Austria.

[6] SAGNOL, G. and HARMAN, R. (2015). Computing exact D-optimal designs by mixed integer second order cone programming. *The Annals of Statistics* **43**, 2198–2224.

[7] YU, Y. (2011). D-optimal designs via a cocktail algorithm. *Statistics and Computing* **21**, 475–481.

## ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS WITH SAMPLE SIZE RECALCULATION

SERGEY TARIMA
*Medical College of Wisconsin*
and
NANCY FLOURNOY*
*University of Missouri*

Consider an experiment in which the primary objective is to determine the significance of a treatment effect at a predetermined type I error and statistical power. Assume that the sample size required to maintain these type I error and power rates will be re-estimated at an interim analysis. A secondary objective is to estimate the treatment effect. Our main finding is that the asymptotic distributions of standardized statistics are random mixtures of distributions, which are non-normal with sample size re-estimation (SSR) except under certain model choices. Monte-Carlo simulation studies and an illustrative example highlight the fact that asymptotic distributions of estimators with SSR may differ from the asymptotic distribution of the same estimators without SSR.

## ON THE ABERRATIONS OF MIXED LEVEL ORTHOGONAL ARRAYS WITH REMOVED RUNS

ROBERTO FONTANA*
*DISMA, Dipartimento di Scienze Matematiche, Politecnico di Torino*
and
FABIO RAPALLO
*DISIT, Dipartimento di Scienze e Innovazione Tecnologica, Università del Piemonte Orientale, Alessandria*

Given an Orthogonal Array we analyze the aberrations of the sub-fractions which are obtained by the deletion of some of its points. We provide formulae to compute the Generalized Word-Length Pattern of any sub-fraction. In the case of the deletion of one single point, we provide a simple methodology to find which the best sub-fractions are according to the Generalized Minimum Aberration criterion, [2]. We also study the effect of the deletion of 1, 2 or 3 points on some examples. The methodology does not put any restriction on the number of levels of each factor. It follows that any mixed level Orthogonal Array can be considered.

The presentation is based on [1].

*Keywords:* Orthogonal Arrays, Generalized Word-Length Pattern, Generalized Minimum Aberration criterion, Incomplete designs

## References

[1] FONTANA, R. and RAPALLO, F. (2018). On the aberrations of mixed level orthogonal arrays with removed runs. *Statistical Papers*, 1–15.

[2] XU, H. and WU, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *The Annals of Statistics* **29**, 1066–1077.

## OPTIMAL DESIGNS FOR $K$-FACTOR TWO-LEVEL MODELS WITH FIRST-ORDER INTERACTIONS ON A SYMMETRICALLY RESTRICTED DESIGN REGION

FRITJOF FREISE*
*TU Dortmund University*
and
RAINER SCHWABE
*University of Magdeburg*

We develop $D$-optimal designs for linear models with first-order interactions on a subset of the $2^K$ full factorial design region, when both the number of factors set to the higher level and the number of factors set to the lower level are simultaneously bounded by the same threshold. It turns out that in the case of narrow margins the optimal design is concentrated only on those design points, for which either the threshold is attained or the numbers of high and low levels are as equal as possible. In the case of wider margins the settings are more spread and the resulting optimal designs are as efficient as a full factorial design. These findings also apply to other optimality criteria.

## OPTIMAL-ROBUST SELECTION OF A DIESEL FUEL SURROGATE FOR HOMOGENEOUS CHARGE COMPRESSION IGNITION MODELING

IRENE GARCÍA-CAMACHA GUTIÉRREZ*
*Institute of Applied Mathematics for Science and Engineering (IMACI), University of Castilla-La Mancha, Toledo, Spain*

R. MARTÍN MARTÍN
*Institute of Applied Mathematics for Science and Engineering (IMACI), University of Castilla-La Mancha, Toledo, Spain*

and

J. SANZ ARGENT
*Fundación Valenciaport, Valencia, Spain*

Homogeneous Charge Compression Ignition (HCCI) combustion is a potential candidate for dealing with the stringent regulations on vehicle emissions while keeping a very good energy efficiency. In the development of the HCCI concept, the availability of reliable computer models has proved fundamental due to its lower cost compared with real engine experiments. The main aim of this study is to provide a new methodology which optimally and robustly selects the simulations to be performed for the selection of the most realistic formulation of a diesel surrogate fuel. The problem was posed from an optimal mixture experimental design approach allowing that the response varies over a neighbourhood of the fixed model. This consideration was motivated by the fact that practitioners have little information about model suitability prior to running the experiments, and optimal designs are strongly model-dependent. A complex minimax optimization problem must be solved for which a genetic algorithm was specially developed. Standard mixture designs, the most common choice among engine researchers, are around 25% efficient compared with the designs obtained in this study, when there is deviation from the model. Once an optimal-robust design was selected, we provide the optimal mixture that best reproduces the combustion property to be imitated.

## MODELING AND OPTIMIZING SET FUNCTIONS VIA RKHS EMBEDDINGS

DAVID GINSBOURGER*
*Idiap Research Institute and University of Bern*
POOMPOL BUATHONG
*Idiap Research Institute and Mahidol University*
and
TIPALUCK KRITYAKIERNE
*Mahidol University*

We consider the issue of modeling and optimizing set functions, with a main focus on kernel methods for expensive objective functions taking finite sets as inputs. Based on recent developments on embeddings of probability distributions in Reproducing Kernel Hilbert Spaces, we explore adaptations of Gaussian Process modeling and Bayesian Optimization to the framework of interest. In particular, combining RKHS embeddings and positive definite kernels on Hilbert spaces delivers a promising class of kernels, as illustrated on a challenging mechanical test case (tackled at mODa11) with bi-phasic input media parametrized by inclusion centers.

## A NEW FAMILY OF RESPONSE SURFACE DESIGNS

PETER GOOS*
*KU Leuven*
and
JOSÉ NÚÑEZ ARES
*KU Leuven*

Response surface designs are a core component of the response surface methodology, which is widely used in the context of product and process optimization. In this contribution, we present a new class of 3-level response surface designs, which can be viewed as matrices with entries equal to -1, 0 and +1. Because the new designs are orthogonal for the main effects and exhibit no aliasing between the main effects and the second-order effects (two-factor interactions and quadratic effects), we call them orthogonal minimally aliased response surface designs or OMARS designs. We constructed a catalog of 55,531 OMARS design for 3 to 7 factors using integer programming techniques. Also, we characterized each design in the catalog extensively in terms of estimation and prediction efficiency, power, fourth-order correlations, and projection capabilities, and we identified

interesting designs and investigated trade-offs between the different design evaluation criteria. Finally, we developed a multi-attribute decision algorithm to select designs from the catalog. Important results of our study are that we discovered some novel designs that challenge standard response surface designs and that our catalog offers much more flexibility than the standard designs currently used.

## SOLVING A PRACTICAL DESIGN PROBLEM BY USING THEORY AND HEURISTIC THINKING

HEIKO GROßMANN*
*Otto-von-Guericke-Universität Magdeburg*

Designs for choice experiments with many attributes require the respondents to process a large amount of information when the alternatives in the choice sets are specified by using all available attributes. The resulting complexity may prompt participants to use simplifying decision rules that violate the additivity assumption on the latent utility scale of the choice model which in turn may invalidate the statistical analysis. One approach to mitigating this problem is to use partial profiles which use only some of the attributes to specify the alternatives in each choice set. The design problem considered in this talk is motivated by a project in health sciences where an efficient partial profile design for estimating the main effects and two-factor interactions of eleven two-level attributes had to be constructed. It is shown how a heuristic procedure which is derived from theoretical results can be successfully used to address this type of problem. In order to apply the procedure, some "intuition" is needed to solve a combinatorial problem. It is hoped that participants of the conference may suggest ideas how this intuition can be turned into a formal algorithm.

### References

[1] GROßMANN, H. (2017). Partial-profile choice designs for estimating main effects and interactions of two-level attributes from paired comparison data. *Journal of Statistical Theory and Practice* **11**, 236–253.

[2] GROßMANN, H. (2018+). A practical approach to designing partial-profile choice experiments for estimating main effects and interactions of many two-level attributes from pairs. *Journal of Choice Modelling*, online first.

## OPTIMAL BAYESIAN DESIGN FOR MODELS WITH INTRACTABLE LIKELIHOODS VIA SUPERVISED LEARNING METHODS

MARKUS HAINY*
*Johannes Kepler University Linz, Austria*
DAVID J. PRICE
*University of Melbourne, Australia*
OLIVIER RESTIF
*University of Cambridge, U.K.*
and
CHRISTOPHER DROVANDI
*Queensland University of Technology, Brisbane, Australia*

Optimal Bayesian experimental design is often computationally intensive due to the need to approximate many posterior distributions for datasets simulated from the prior predictive distribution. The issues are compounded further when the statistical models of interest do not possess tractable likelihood functions and only simulation is feasible. We employ supervised learning methods to facilitate the computation of utility values in optimal Bayesian design. This approach requires considerably fewer simulations from the candidate models than previous approaches using approximate Bayesian computation. The approach is particularly useful in the presence of models with intractable likelihoods but can also provide computational advantages when the likelihoods are manageable. We consider the two experimental goals of model discrimination and parameter estimation. The methods are applied to find optimal designs for models in epidemiology and cell biology.

## SHOULD WE INFORM ABOUT ALLOCATION RATIO IN RANDOMIZED CLINICAL TRIALS?

RALF-DIETER HILGERS*
*Department of Medical Statistics, RWTH Aachen University Aachen*
MARTIN MANOLOV
*Department of Medical Statistics, RWTH Aachen University Aachen*
and
NICOLE HEUSSEN
*Department of Medical Statistics, RWTH Aachen University Aachen*

**Background** The informed consent of a clinical trial should include information about the randomization process. Most often the allocation ratio is given to indicate, that the patient has an equal chance to receive one of the treatments. The objective of this paper is to investigate whether unbalanced allocation, which is applied frequently in rare disease clinical trials, per se as well as the knowledge about the allocation ratio boost the effect of allocation bias on the test decision.

**Methods** We consider a 2-arm parallel group, single center randomized clinical trial with continuous endpoint, with no-interim analysis and no adaptation in the randomization process. Generalized versions of complete randomization, random allocation rule, permuted block randomization as well as Efron's Biased coin for unbalanced allocation ratio say n:m are formulated and used to investigate whether the test decision is affected by allocation bias. The decision is quantified by means of the type I error probability for the t-test under misspecification.

**Results** Because this is work in progress, the results could not be confirmed. However, first simulations show, that with increased allocation ratio the impact of allocation bias on the test decision increases, if the allocation ratio is known. If the allocation ratio is concealed, the effect is diminished.

**Conclusion** In terms of allocation bias, the allocation ratio should not be published before the trial is finished, neither in the protocol nor in the informed consent.

*Keywords:* Randomization, Clinical Trial, Allocation bias

## IMPROVEMENTS IN HANDLING LIMIT OF QUANTIFICATION AND/OR CENSORED DATA IN OPTIMAL TRIAL DESIGN

ANDREW C. HOOKER*
*Department of Pharmaceutical Biosciences, Uppsala University*

It is not uncommon to expect some amount of data in a drug trial to be below limit of quantification (BLQ). Typically, this data will just be reported as being BLQ, whereas data above limit of quantification will be reported with a value of the measured variable. Although BLQ data is less informative, it still has information and often cannot be ignored [1]. In the design of trials, previous work has shown that methods considering the probability of an observation being BLQ have better properties than those ignoring this possibility [2]. In that work, the recommended method permuted across all possible iterations of the design points being BLQ, computed the Fisher information matrix (FIM) for each of those permutations assuming that the BLQ data were uninformative, and then took an average of the computed FIMs weighted by the probability for each design permutation to occur. This work proposes improvements to this method, filtering out permutations that are very unlikely, making the method much faster for design evaluation and optimization. Further, the method is extended to handle other types of censoring, such as censoring when observations exceed a given value. Examples in pharmacometric systems are presented using nonlinear mixed-effect models.

### References

[1] BERGSTRAND, M. and KARLSSON, M. O. (2009). Handling Data Below the Limit of Quantification in Mixed Effect Models. *The AAPS Journal* **11** 371–380.

[2] VONG, C., UECKERT, S., NYBERG, J. and HOOKER, A. C. (2012). Handling Below Limit of Quantification Data in Optimal Trial Design. *PAGE. Abstracts of the Annual Meeting of the Population Approach Group in Europe.*

## CLINICAL TRIALS WITH ADAPTIVE ENRICHMENT

ANASTASIA IVANOVA*
*UNC at Chapel Hill*

We consider the problem of estimating the best subgroup and testing for treatment effect in a clinical trial. We propose a three-stage design, where the subgroup is estimated

at the first interim analysis and refined in the second interim analysis. This approach is illustrated by an example of a clinical trial in severe asthma.

## CONSTRUCTION, PROPERTIES, AND ANALYSIS OF GROUP-ORTHOGONAL SUPERSATURATED DESIGNS

BRADLEY JONES*
*JMP Division/SAS*
and
CHRISTOPHER NACHTSHEIM
*University of Minnesota*

In this paper, we propose a new method for constructing supersaturated designs that is based on the Kronecker product of two carefully-chosen matrices. The construction method leads to a partitioning of the columns of the design such that the columns within a group are correlated to the others within the same group, but are orthogonal to any factor in any other group. We refer to the resulting designs as group orthogonal supersaturated designs (GO-SSDs). We leverage this group structure to obtain an unbiased estimate of the error variance, and to develop an effective, design-based model selection procedure. Simulation results show that the use of these designs, in conjunction with our model selection procedure enables the identification of larger numbers of active main effects than have previously been reported for supersaturated designs. The designs can also be used in group screening; however, unlike previous group-screening procedures, with our designs, main effects in a group are not confounded.

## CONDITIONAL INFORMATION AND INFERENCE IN RESPONSE-ADAPTIVE RANDOMIZATION

ADAM LANE*
*Cincinnati Children's Hospital Medical Center*

Response-adaptive randomization refers to a class of designs where the probability a subject is assigned to a treatment is changed throughout the trial based on the accrued responses. Inference following such procedures is an important and difficult question that has been examined extensively. The difficulty arises from the dependence of the treatment sample sizes on the observed responses. Specifically, we consider exact and approximate conditional inference of the maximum likelihood estimate (MLE) for the randomized play the winner (RPW) design. The RPW design is one of the simplest

and widely studied response-adaptive randomized designs. Exact conditional versus unconditional inference is the subject of Wei et al. [1]. Wei et al. (1990) found the unconditional inference to be superior to conditional inference. However, Wei et al. did not consider the conditional MLE. We find heuristically, that the conditional MLE is superior in certain regions of the parameter space.

## References

[1] WEI, L. J., SMYTHE, R. T., LIN, D. Y. and PARK, T. S. (1990). Statistical Inference with Data-Dependent Treatment Allocation Rules. *Journal of the American Statistical Association* **85**, 156–162.

## MODELLING OF CORRELATED RANDOM VARIABLES AND ITS APPLICATION IN CLINICAL TRIALS: EXTREME CORRELATIONS

SERGEI L. LEONOV*
*CSL Behring, 1020 1st Ave, King of Prussia, PA 19406, U.S.A.*

Modelling of correlated random variables is important for evaluating operating characteristics of experimental designs in a variety of fields, including clinical trials with multiple endpoints. There exist efficient algorithms for generating multivariate distributions with given marginals and correlation structure. For model fitting and simulation, it is critical to know the feasible range of pairwise correlations, which can be much narrower than the interval [-1, +1]. We focus on deriving closed-form expressions for extreme correlations for several classes of bivariate distributions that involve both discrete and continuous endpoints.

## MODEL-ROBUST CLASSIFICATION IN ACTIVE LEARNING

JESÚS LÓPEZ-FIDALGO*
*University of Navarra*
JOSÉ ANTONIO MOLER
*Public University of Navarra*
and
DOUGLAS P. WIENS
*University of Alberta*

We aim to develop a theory of model-robust classification, and a methodology for applying this to large data sets such as arise in machine learning. The general idea is that there is a (large) population of explanatory variables, which can be easily sampled. With probability $a(x;t)$, an item with covariates $x$ belongs to group A and with probability $1-a(x;t)$ it belongs to group B. We suppose that the determination of the appropriate group, given $x$, is difficult and expensive, so that the investigator wishes to sample from $x$ in a manner which is more effcient than random sampling (sometimes termed 'passive learning').

## OPTIMAL DESIGN OF INSPECTION TIMES FOR INTERVAL CENSORING

NADJA MALEVICH*
*TU Dortmund University*
and
CHRISTINE H. MÜLLER
*TU Dortmund University*

We treat optimal equidistant and optimal non-equidistant inspection times for interval censoring of exponential distributions. We provide in particular a new approach for determining the optimal non-equidistant inspection times. The resulting recursive formula is related to a formula for optimal spacing of quantiles for asymptotically best linear estimates based on order statistics and to a formula for optimal cutpoints by the discretisation of continuous random variables. Moreover, we show that by the censoring with the optimal non-equidistant inspection times as well as with optimal equidistant inspection times, there is no loss of information if the number of inspections is converging to infinity. Since optimal equidistant inspection times are easier to calculate and easier to handle in practice, we study the efficiency of optimal equidistant inspection times with

respect to optimal non-equidistant inspection times. Moreover, since the optimal inspection times are only locally optimal, we also provide some results concerning maximin efficient designs.

## D-OPTIMAL DESIGNS FOR AFT MODELS WITH TEMPERATURE ACCELERATION FACTOR

IRENE MARIÑAS-COLLADO*
*Dpto. de Estadística, Universidad de Salamanca*
MARÍA J. RIVAS-LÓPEZ
*Dpto. de Estadística, Universidad de Salamanca*
JUAN M. RODRÍGUEZ-DÍAZ
*Dpto. de Estadística, Universidad de Salamanca*
and
M. TERESA SANTOS-MARTÍN
*Dpto. de Estadística, Universidad de Salamanca*

An accelerated life test of a product or material consists on the observation of its failure time when it is subjected to conditions that stress the usual ones. The purpose is to obtain the parameters of the distribution of the time-to-failure for usual conditions through the observed failure-times. Increasing temperature is one of the most commonly used methods to accelerate a failure mechanism. In this case, the accelerated failure process can be described making use of Arrhenius (or Eyring) equations. Some results on D-optimal design are presented for accelerated failure time models when the acceleration factor is temperature.

## LASSO AND MODEL COMPLEXITY

HUGO MARURI-AGUILAR*
*School of Mathematical Sciences, Queen Mary, University of London*

The statistical technique of Lasso (Tibshirani et al, 1996) is built around weighted penalisation of the error term by the absolute sum of coefficients. As the control parameter increases in value, model coefficients shrink progressively towards zero, thus providing the user with a collection of models that start from the ordinary least squares regression model and end with a model with no terms.

This work gravitates around hierarchical squarefree regression models. These models can be seen as simplicial complexes and thus a measure of complexity is given by Betti numbers of the "model/complex". We detail our computations and implementation of the methodology and illustrate our proposal with simulation results and also apply the methodology to a dataset from the literature.

This is joint work with S. Hu (Queen Mary).

## INFERENCE UNDER A TWO-STAGE ADAPTIVE DESIGN FOR NON-LINEAR REGRESSION MODELS

NANCY FLOURNOY
*University of Missouri*
CATERINA MAY*
*Università del Piemonte Orientale*
and
CHIARA TOMMASI
*Università di Milano*

In this work, we assume that a response variable is explained by several controlled explanatory variables through a non-linear regression model. The unknown parameter is the vector of coefficients, and thus it is multidimensional. To collect the responses, we consider a two-stage experimental design; in the first-stage data are observed at some fixed initial design; then the data are used to "estimate" an optimal design at which the second-stage data are observed. Therefore, first and second stage responses are dependent. At the end of the study, the whole set of data is used to estimate the unknown vector of coefficients through maximum likelihood. In practice, it is quite common to take a small pilot sample to demonstrate feasibility. This pilot study provides an initial estimate of unknown parameters which are used to build a second-stage design at which additional data are collected to improve the estimate. Accordingly, we obtain the asymptotic behaviour of the maximum likelihood estimator under the assumption that only the second-stage sample size goes to infinity, while the first-stage sample size is assumed to be fixed. This contrasts with the classical approach in which both the sample sizes are assumed to become large.

## ADAPTIVE DESIGNS FOR DRUG COMBINATION INFORMED BY LONGITUDINAL MODEL FOR THE RESPONSE

TOBIAS MIELKE*
*Janssen, Johnson & Johnson Platz 1, Neuss 41470, Germany*
and
VLADIMIR DRAGALIN
*Janssen, 1400 McKean Rd, Spring House, PA, 19477, USA*

Objectives in Phase II drug combination studies are to estimate the efficacy response surface for the combination of doses of different drugs and to select the most efficient combination for the final Phase III clinical trial. One problem is to find an optimal design that allocates subjects to the dose-combinations which will maximize the information obtained in the trial. Adaptive designs help in these situations to ensure high efficiency of the study design. We are using a binary efficacy endpoint and consider the practical situation when the timing of the endpoint assessment period on the subject level is considerably longer relative to the inter-arrival time of subjects. This poses implementation challenges for the adaptive design. A solution to the adaptive design problem by using time-to-event models as longitudinal model will be presented.

*Keywords:* Adaptive design, Dose combination study, Delayed response

### References

[1] CAI, C., LIU, S. and YUAN, Y. (2014). A Bayesian design for phase II clinical trials with delayed responses based on multiple imputation. *Statistics in Medicine* **33**, 4017–4028.

[2] LIU, S. and NING, J. (2013). A Bayesian Dose-Finding Design for Drug Combination Trials with Delayed Toxicities. *Bayesian Analysis* **8**, 703–722.

[3] ATKINSON, A. C., FEDOROV, V. V., HERZBERG, A. M. and ZHANG, R. (2014). Elemental information matrices and optimal experimental design for generalized regression models. *Journal of Statistical Planning and Inference* **144**, 81–91.

## ON THE EXPERIMENTAL ATTAINMENT OF OPTIMUM CONDITIONS IN AN INJECTION MOULDING PROCESS

MICHAEL BRUMMER
*Johannes Kepler University Linz*
and
WERNER G. MÜLLER*
*Johannes Kepler University Linz*

Depending upon how much trust can be put in the statistical model there exists an inherent need for balancing the "exploration" and "exploitation" characteristics of an experimental design. Recently Nachtsheim and Jones [2] have argued that Box and Wilson's (1951) classic search for optimum operating conditions be complemented by a model estimation element. In our presentation we intend to provide an alternative, which is more strongly oriented towards the focus of optimization and we will exemplify the relative advantages or disadvantages of our proposal on a real experiment, where the input factors melt temperature, mould temperature, holding pressure, cooling time and ejection delay were used to minimize the production cycle time of an injection moulded part.

### References

[1] BOX, G. E. P. and WILSON, K. B. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society. Series B (Methodological)* **13**, 1–45.

[2] NACHTSHEIM, C. and JONES, B. (2018). Design augmentation for response optimization and model estimation. *Quality Engineering* **30**, 38–51.

## EMBEDDING KNOWLEDGE OF THE DESIGN IN KRIGING PREDICTION THROUGH OBJECTIVE BAYESIAN ANALYSIS

JOSEPH MURÉ*
*EDF R&D, 6 quai Watier, 78401 Chatou, France*
*Université Paris Diderot, Bâtiment Sophie Germain, Avenue de France, 75013 Paris, France*

Estimation of Kriging parameters is sensitive to the design used. How can information about the design of experiments be embedded in the prior distribution?

Berger and De Oliveira [1] proposed the reference prior as an "objective" prior distribution in the case of isotropic correlation kernels. This prior naturally takes the design set into account.

The Gibbs reference posterior distribution [2] is an extension to anisotropic correlation kernels that is well suited to Gibbs sampling.

Simulation examples show that a posteriori Predictive intervals at unobserved points have effective coverage close to their theoretical level.

*Keywords:* Design of Experiments, Objective Bayes, Reference prior, Optimal compromise, Gibbs reference posterior

### References

[1] BERGER, J. O., DE OLIVEIRA, V. and SANSÓ, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* **96**, 1361–1374.

[2] MURÉ, J. (2018). Optimal compromise between incompatible conditional probability distributions, with application to Objective Bayesian Kriging. To appear in *ESAIM: Probability and Statistics*, https://doi.org/10.1051/ps/2018023, https://arxiv.org/pdf/1703.07233.

## OPTIMAL RESTRICTED-RANDOMISED DESIGNS ENSURING PRECISE PURE-ERROR ESTIMATION OF THE VARIANCE COMPONENTS

KALLIOPI MYLONA*
*Department of Mathematics, King's College London, United Kingdom*
STEVEN G. GILMOUR
*Department of Mathematics, King's College London, United Kingdom*
and
PETER GOOS
*Faculty of Bioscience Engineering, KU Leuven, Belgium*
*Faculty of Applied Economics & StatUa Center for Statistics, Universiteit Antwerpen, Belgium*

In this work, we present a novel approach to design blocked and split-plot experiments which ensures that the two variance components can be estimated from pure error and guarantees a precise estimation of the response surface model. Our novel approach involves a new Bayesian compound D-optimal design criterion which pays attention to both the variance components and the fixed treatment effects. One part of the compound criterion (the part concerned with the treatment effects) is based on the response surface model of interest, while the other part (which is concerned with pure-error estimates of the variance components) is based on the full treatment model. We demonstrate that our new criterion yields split-plot designs that outperform existing designs from the literature both in terms of the precision of the pure-error estimates and the precision of the estimates of the factor effects.

## DISTRIBUTION OF THE MULTIVARIATE NONLINEAR LS ESTIMATOR UNDER AN UNCERTAIN INPUT

ANDREJ PÁZMAN*
*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Slovakia*

The aim of the talk is to develop further the approach presented in [1] for the computation of the probability density of a least squares estimator for moderate size samples in nonlinear regression. We consider here cases when the variance matrix of observations is not known, hence, it can not be used for the definition of the parameter estimator. We derived "almost exact" results, with a modified and better defined meaning of this concept. Possible applications on three variants of an experiment of heat transfer are indicated.

### References

[1] PÁZMAN, A. (1993). *Nonlinear statistical models.* Kluwer, Dordrecht.

## THE BLUE IN CONTINUOUS-TIME REGRESSION MODELS WITH CORRELATED ERRORS

HOLGER DETTE
*Bochum University*
ANDREY PEPELYSHEV*
*Cardiff University*
and
ANATOLY ZHIGLJAVSKY
*Cardiff University*

In my talk the problem of best linear unbiased estimation is investigated for continuous-time regression models. We present several general statements concerning the explicit form of the best linear unbiased estimator (BLUE), in particular when the error process is a smooth process with one or several derivatives of the response process available for construction of the estimators. We derive the explicit form of the BLUE for many

specific models including the cases of continuous autoregressive errors of order two and integrated error processes (such as integrated Brownian motion).

## LARGE DATASETS, BIAS AND MODEL ORIENTED OPTIMAL DESIGN OF EXPERIMENTS

ELENA PESCE*
*Department of Mathematics, University of Genoa, Italy*
and
EVA RICCOMAGNO
*Department of Mathematics, University of Genoa, Italy*

Statistical methods have been developed for the analysis of Big Datasets, which use the full available data. In contrast other authors argue on the advantages of inference statements based on a well-chosen subset of the big dataset. We review recent literature that proposes to adapt ideas from classical model based optimal design of experiments to problems of data selection of large datasets. Special attention is given to bias reduction and to protection against confounders. Some new results are presented. Theoretical and computational comparisons are made.

## EFFICIENCY ANALYSIS FOR THE MISSPECIFICATION PROBLEM OF RESPONSE PROBABILITY DISTRIBUTION IN OPTIMAL EXPERIMENTAL DESIGN

SERGIO POZUELO-CAMPOS*
*University of Castilla-La Mancha*
MARIANO AMO-SALAS
*University of Castilla-La Mancha*
and
VÍCTOR CASERO-ALONSO
*University of Castilla-La Mancha*

In the Optimal Experimental Design theory, a homoscedastic normally distributed response is generally assumed. In some works, other probability distributions for the response are considered. Here, we analyze, in terms of efficiency, the influence of a misspecification of the response probability distribution. We rely on the elemental information matrix, which allows us to generalize Fisher's information matrix, to obtain

approximate optimal designs for responses with any probability distribution. We have proved some theoretical results when the true distribution is the Gamma or Poisson distribution, but it is considered an equivalent heteroscedastic Normal distribution. Finally, we have extended this analysis to a 4 parameters Hill model, which explain the effect of a drug in the cellular growth.

*Keywords:* D-optimality, D-efficiency, Elemental information matrix, Approximate design, Hill model

## BREGMAN DIVERGENCES BASED ON OPTIMAL DESIGN CRITERIA AND SIMPLICIAL MEASURES OF DISPERSION

LUC PRONZATO*

*CNRS, UCA, Laboratoire I3S, UMR 7172; 2000, route des Lucioles, Les Algorithmes, bât. Euclide B, 06900 Sophia Antipolis, France*

HENRY P. WYNN

*London School of Economics, Houghton Street, London, WC2A 2AE, UK*

and

ANATOLY ZHIGLJAVSKY

*School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, CF24 4YH, UK*

In previous work the authors defined the $k$-th order simplicial distance between probability distributions which arises naturally from a measure of dispersion based on the squared volume of random simplices of dimension $k$. This theory is embedded in the wider theory of divergences and distances between distributions which includes Kullback-Leibler, Jensen-Shannon, Jeffreys-Bregman divergence and Bhattacharyya distance. A general construction is given based on defining a directional derivative of a function $\phi$ from one distribution to the other whose concavity or strict concavity influences the properties of the resulting divergence. For the normal distribution these divergences can be expressed as matrix formula for the (multivariate) means and covariances. Optimal experimental design criteria contribute a range of functionals applied to non-negative, or positive definite, information matrices. Not all can distinguish normal distributions but sufficient conditions are given. The $k$-th order simplicial distance is revisited from this aspect and the results are used to test empirically the identity of means and covariances.

*Keywords:* Simplicial distances, Bregman divergence, Optimal design criteria, Burbea-Rao divergence, Energy statistic

## OPTIMAL DESIGNS FOR MINIMAX-CRITERIA IN RANDOM COEFFICIENT REGRESSION MODELS

MARYNA PRUS*
*Otto von Guericke University Magdeburg*

We consider minimax-optimal designs for the prediction of individual parameters in random coefficient regression models. We focus on the minimax-criterion, which minimizes the "worst case" for the basic criterion with respect to the covariance matrix of random effects. We discuss particular models: linear and quadratic regression, in detail.

## LOCALLY $D$-OPTIMAL DESIGNS FOR A WIDER CLASS OF NON-LINEAR MODELS ON THE $K$-DIMENSIONAL BALL WITH APPLICATIONS TO LOGIT AND PROBIT MODELS

MARTIN RADLOFF*
*Institute for Mathematical Stochastics, Otto-von-Guericke-University, Universitätsplatz 2, 39106 Magdeburg, Germany*
and
RAINER SCHWABE
*Institute for Mathematical Stochastics, Otto-von-Guericke-University, Universitätsplatz 2, 39106 Magdeburg, Germany*

In this presentation we extend the results of [1], which could be applied for example to Poisson regression, negative binomial regression and proportional hazard models with censoring, to a wider class of non-linear multiple regression models. This includes the binary response models with logit and probit link besides others. For this class of models we derive (locally) $D$-optimal designs when the design region is a $k$-dimensional ball. For the corresponding construction we make use of the concept of invariance and equivariance in the context of optimal designs as in our previous paper. In contrast to the former results the designs will not necessarily be exact designs in all cases. Instead approximate designs can appear. These results can be generalized to arbitrary ellipsoidal design regions.

### References

[1] RADLOFF, M. and SCHWABE, R. (2018). Locally $D$-optimal Designs for Non-linear Models on the $k$-dimensional Ball. arXiv:1806.00275.

## CIRCUITS IN EXPERIMENTAL DESIGN. THEORY AND ALGORITHMS

FABIO RAPALLO*
*Department DISIT, Università del Piemonte Orientale, Italy*

In the framework of factorial designs, a statistical (linear) model is defined through an appropriate model matrix, which depends on the experimental runs and encodes their geometric structure. In this talk we discuss some properties of the circuit basis of the model matrix in connection with two well-known properties of the designs: robustness and D-optimality.

We introduce an exchange-type algorithm for choosing robust designs based on the intersections between the design and the supports of the circuits. Using some simulations, we study the properties of such an algorithm, and we show that it generates a good design in few iterations. Finally, we modify the algorithm by using only a special class of circuits, namely the basic moves, widely used in Algebraic Statistics. This allows us to apply the algorithm also in large-sized problems.

This is joint work with Henry Wynn (CATS, London School of Economics, UK).

### References

[1] FONTANA, R., RAPALLO, F. and ROGANTIN, M. P. (2014). A Characterization of Saturated Designs for Factorial Experiments. *Journal of Statistical Planning and Inference* **147**, 204–211.

[2] FONTANA, R. and RAPALLO, F. (2018). Simulations on the combinatorial structure of D-optimal designs. In: J. Pilz, D. Rasch, V.B. Melas, K. Moder (Eds.) *Statistics and Simulation*, Springer, New York, 343–353

## COPULA-BASED ROBUST OPTIMAL BLOCK DESIGNS

ANDREAS RAPPOLD*
*Johannes Kepler University Linz*
WERNER G. MÜLLER
*Johannes Kepler University Linz*
and
DAVID WOODS
*University of Southampton*

Blocking is often used to reduce known variability in designed experiments by collecting together homogeneous experimental units. A common modeling assumption for such experiments is that responses from units within a block are dependent. Accounting for such dependencies in both the design of the experiment and the modeling of the resulting data when the response is not normally distributed can be challenging, particularly in terms of the computation required to find an optimal design. The application of copulas and marginal modeling provides a computationally efficient approach for estimating population-average treatment effects. Motivated by an experiment from materials testing, we develop and demonstrate designs with blocks of size two using copula models. Such designs are also important in applications ranging from microarray experiments to experiments on human eyes or limbs with naturally occurring blocks of size two. We present a methodology for design selection, make comparisons to existing approaches in the literature, and assess the robustness of the designs to modeling assumptions.

## EFFICIENT DESIGNS IN TWO-WAY-COVARIANCE MULTIRESPONSE MODELS

JUAN M. RODRÍGUEZ DÍAZ*
*University of Salamanca*
and
GUILLERMO SÁNCHEZ-LEÓN
*University of Salamanca*

When the interest is in observing several responses, the assumption of a double covariance structure (between different responses observed at the same point, and between the same type of response observed at different points) increases the difficulty to obtain efficient designs for the knowledge of the model. Analytical results can help very much to

this purpose for certain covariance structures. Theoretical results can as well be obtained for compositional response models, that can be seen as special cases of multiresponse models. Examples of application for a chemical reaction and a compositional response model will be shown. Stepping out of the comfortable assumptions for the covariance structure makes the problem much more difficult, opening new lines of research.

*Keywrods:* Compositional data, Covariance matrix , Exponential covariance, Linear models, Optimal design of experiments

### References

[1] RODRÍGUEZ-DÍAZ J. M. and SÁNCHEZ-LEÓN, G. (2019). Optimal designs for multiresponse models with double covariance structure. *Chemometrics and Intelligent Laboratory Systems* in press, doi.org/10.1016/j.chemolab.2019.03.009.

## EQUIVALENCE OF WEIGHTED AND PARTIAL OPTIMALITY OF EXPERIMENTAL DESIGNS

SAMUEL ROSA*

*Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava*

Morgan and Wang [1] proposed the weighted optimality criteria for experimental design to allow one to place various emphasis on different parameters of interest. Later, this approach was generalized in [3] to allow for various weights for the functions of model parameters. However, differential emphasis on parameter functions can also be expressed by considering the well-developed optimality criteria for estimating a parameter system of interest (the partial optimality criteria), see [2]. We prove that the approaches of weighted optimality and partial optimality are in fact equivalent for any eigenvalue-based optimality criterion [4]. Therefore, the large body of theoretical and computational results on partial optimality can be employed when various weights are placed on the (functions of) parameters of interest.

### References

[1] MORGAN, J. P. and WANG X (2010). Weighted optimality in designed experimentation. *Journal of the American Statistical Association* **105**, 1566–1580.

[2] PUKELSHEIM, F. (1993). Optimal design of experiments. Wiley, New York.

[3] STALLINGS, J. W. and MORGAN, J. P. (2015). General weighted optimality of designed experiments. *Biometrika* **102**, 925–935.

[4] ROSA S. (2019). Equivalence of weighted and partial optimality of experimental designs. *Metrika*, Online first.

## RANDOMIZATION-BASED INFERENCE AND THE CHOICE OF RANDOMIZATION PROCEDURES

WILLIAM F. ROSENBERGER*
*George Mason University*
DIANE USCHNER
*George Mason University*
and
YANYING WANG
*George Mason University*

In testing the significance of treatment effects in randomized clinical trials (RCTs), randomization-based inference is distinguished from population-based parametric and nonparametric inference, such as the t-test or permutation tests, taking into account three properties: preservation of type I error rate, relation of power to the randomization procedure, and flexibility in choosing the test statistic. In this paper, we revisit rationale of the properties and provide justification through simulations. We propose that the choice of randomization procedures and the analysis of RCTs can be facilitated by the application of randomization-based inference.

## RHOMBOID DESIGNS FOR LINEAR REGRESSION WITH CORRELATED RANDOM COEFFICIENTS

ULRIKE GRAßHOFF
*Humboldt-Universität zu Berlin*
HEINZ HOLLING
*Universität Münster*
FRANK RÖTTGER*
*Otto-von-Guericke-Universität Magdeburg*
and
RAINER SCHWABE
*Otto-von-Guericke-Universität Magdeburg*

We study a linear regression model $Y_i(x_i) = f(x_i)^T b_i$ on the hypercube with an intercept where $b_i \sim N(\beta, D)$ and all $Y_i(x_i)$ are independent, which means that there is only one observation per realisation of $b_i$. The parameter to be estimated is the population

location parameter $\beta$, while the dispersion matrix $D$ is assumed to be known. We further assume that the structure of $D$ displays an independent intercept and a completely symmetric covariance matrix for the random main effects. Through a model transformation and the introduction of rhomboid designs, we see that the Kiefer-Wolfowitz equivalence theorem characterizes the situations when the optimality regions of these designs are either algebraic varieties intersected with trivial constraints or semi-algebraic sets, i.e if we can restrict the investigation to the vanishing set of a collection of polynomials or if we need to take polynomial inequalities into account. In fact, it shows that this distinction depends on certain properties of the design points. Consequently we then discuss up to dimension 4, for which covariance matrices an optimal rhomboid design is supported either completely on the vertices of the hypercube or has support points in the interior. Furthermore, we conjecture a similar result for arbitrary dimension.

## AN UNEXPECTED CONNECTION BETWEEN BAYES A-OPTIMAL DESIGNS AND THE GROUP LASSO

GUILLAUME SAGNOL*
*TU Berlin*
and
EDOUARD PAUWELLS
*Toulouse 3 – Université Paul Sabatier*

We show that the $A$-optimal design optimization problem over $m$ design points in $\mathbb{R}^n$ is equivalent to minimizing a quadratic function plus a group lasso sparsity inducing term over $nm$ real matrices. This observation allows to describe several new algorithms for $A$-optimal design based on splitting and block coordinate decomposition. These techniques are well known and proved powerful to treat large scale problems in machine learning and signal processing communities. The proposed algorithms come with rigorous convergence guaranties and convergence rate estimate stemming from the optimization literature. Performances are illustrated on synthetic benchmarks and compared to existing methods for solving the optimal design problem.

## AN INTERACTIVE TOOL TO APPLY OPTIMAL EXPERIMENTAL DESIGN IN PLANNING DOSE FINDING STUDIES

THOMAS SCHMELTER*
*Bayer AG*
KATRIN KETTELHAKE
*Bayer AG*
KATRIN ROTH
*Bayer AG*
FRANCO MENDOLIA
*Bayer AG*
and
CHRISTOPH NEUMANN
*Bayer AG*

On behalf of Bayer's biostatistics innovation group on dose finding, we are presenting an interactive tool that applies the theory of optimal experimental design to facilitate the planning of dose finding studies. Dose finding studies aim at estimating the dose response relationship and thereby finding the optimal therapeutic dose. Often the shape of the dose response curve is unknown, let alone the parameters for the dose response model. While the MCP-Mod approach already provides an excellent theoretical framework for planning and analyzing dose finding studies under model uncertainty, we have created an interactive visual tool using the Shiny framework and the DoseFinding package in R. Using this interactive application, potential dose response curves can be discussed with non-statistical colleagues, and the impact of changes to the optimal design can be visualized instantaneously.

## PERFORMANCE OF SUPERSATURATED DESIGNS FROM AN ANALYSIS PERSPECTIVE

RAKHI SINGH*
*TU Dortmund, Germany*
and
JOHN STUFKEN
*University of North Carolina Greensboro, USA*
*Arizona State University, USA*

Supersaturated designs (SSDs) are useful for factor screening when only a small number of factors are active, which holds under the effect sparsity assumption. For two-level level-balanced SSDs, the $E(s^2)$-criterion chooses a design that minimizes the non-orthogonality among all columns of a design. Jones and Majumdar [1] proposed the $UE(s^2)$-criterion, dropped the requirement of level-balance, and provided simple constructions of $UE(s^2)$-optimal designs for each parameter set. They also showed that there exist many $UE(s^2)$-optimal designs given a parameter set. Given the two criteria, it becomes important to understand if either of these two criteria give designs that perform well from an analysis perspective. Various analysis methods are available for SSDs and it has been observed previously that Dantzig Selector works the best. Since SSDs are eventually used to detect active factors, in this work, we compare designs under different criteria based on their effectiveness for identifying active factors. We observe that the criteria defined above are not sufficiently discriminatory when comparing designs from an analysis perspective and we propose a new discriminatory criterion based on resolution-rank of designs. We also obtain several designs with higher resolution ranks than the ones available in the literature.

*Keywords:* Active factor, Factor sparsity, Hadamard matrix, Screening design

### References

[1] JONES, B. and MAJUMDAR, D. (2014). Optimal supersaturated designs. *Journal of the American Statistical Association* **109**, 1592–1600.

## OPTIMUM DESIGN OF EXPERIMENTS BASED ON PRECISION REQUIREMENTS

BERNHARD SPANGL*

*Institute of Applied Statistics and Computing, University of Natural Resources and Life Sciences, Vienna, Austria*

LENKA FILOVÁ

*Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia*

and

RADOSLAV HARMAN

*Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Bratislava, Slovakia*

Let us assume that the observed response of each trial of an experiment depends on a design point $x$ chosen from a design space $\mathfrak{X}$. Let $\Xi_n^E$ denote the set of all exact experimental designs of size $n$ on $\mathfrak{X}$. We first aim at finding the minimum experimental size $n^*$ subject to some precision requirements $\delta$:

$$n^* \in \arg\min\{n : \exists\, \boldsymbol{\xi} \in \Xi_n^E \text{ with } \Psi(\boldsymbol{\xi}) \leq \delta\},$$

where $\Psi(\boldsymbol{\xi})$ is a precision function, such that smaller values of $\Psi$ indicate greater precision. Then we compute the optimum precision design $\boldsymbol{\xi}^*$ as

$$\boldsymbol{\xi}^* \in \arg\min\{\Psi(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi_{n^*}^E \text{ with } \Psi(\boldsymbol{\xi}) \leq \delta\}.$$

We will discuss natural choices of the precision function $\Psi$ for different types of statistical inference, i.e., estimation and testing. Next, we will suggest algorithms for constructing optimal or nearly-optimal precision designs. Finally, we will give some examples and back our proposal by Monte Carlo simulations.

## SEQUENTIAL DESIGN FOR PERSONALIZED MEDICINE

MIA S. TACKNEY*
*Social Statistics and Demography, Social Sciences - University of Southampton, Southampton, SO17 1BJ, United Kingdom*

DAVE WOODS
*Mathematical Sciences - University of Southampton, Southampton, SO17 1BJ, United Kingdom*

KIM MAY LEE
*MRC Biostatistics Unit, School of Clinical Medicine - Cambridge Institute of Public Health, Forvie Site, Robinson Way Cambridge Biomedical Campus, Cambridge, CB2 0SR, United Kingdom*

ILYA SHPITSER
*Department of Computer Science - Johns Hopkins University, 160 Malone Hall, 3400 North Charles Street Baltimore, MD 21218-2608, USA*
and

PETER SMITH
*Social Statistics and Demography, Social Sciences - University of Southampton, Southampton, SO17 1BJ, United Kingdom*

Advances in genomics are making it possible to personalize medicine so that treatments are tailored to patients' genetic information. For example, cancers can now be characterized at the molecular level, and treatments can be targeted at specific genetic and biological mechanisms (biomarkers). Personalized clinical trials need to be able to identify effective treatment-biomarker combinations. We demonstrate a method of designing a sequential experiment with an adaptive treatment allocation scheme which seeks to both find effective treatments and estimate the corresponding treatment-biomarker interactions. This method weights the probabilities of treatment assignment according to an optimality criterion which takes into account the biomarkers, treatment assignment and response of the patients in the trial so far. We provide examples of both myopic and non-myopic sequential strategies. In the former, decisions on which treatment to apply to the current patient ignore any potential information about future patients; in the latter, we account for potential treatment allocations to future patients when choosing the treatment for the current patient. We describe some computational challenges in implementing the non-myopic method and, through simulation studies, we describe possible settings in which it may provide benefit over the myopic approach.

### References

[1] ATKINSON, A. C. (1982). Optimum biased coin designs for sequential clinical trials with prognostic factors. *Biometrika* **69**, 61–67.

[2] LEE, K. M. and WATSON, J. (2019). Design of experiments for a confirmatory trial of precision medicine. *Journal of Statistical Planning and Inference* **199**, 179–187.

## ANALYSIS OF HIERARCHICAL GAUSSIAN PROCESS REGRESSION

ARETHA TECKENTRUP*
*University of Edinburgh*
and
ANDREW STUART
*Caltech*

We consider hierarchical Gaussian process regression, where hyper-parameters appearing in the mean and covariance structure of the Gaussian process emulator are a-priori unknown, and are learnt from the data, along with the posterior mean and covariance. We work in the framework of empirical Bayes, where a point estimate of the hyper-parameters is computed, using the data, and then used within the standard Gaussian process prior to posterior update. Using results from scattered data approximation, we provide a convergence analysis of the method.

## MAJORIZATION-MINIMIZATION ALGORITHM FOR D-OPTIMAL SENSOR SELECTION WITH CORRELATED OBSERVATIONS

DARIUSZ UCIŃSKI*
*University of Zielona Góra, Poland*
and
MACIEJ PATAN
*University of Zielona Góra, Poland*

The problem of sensor selection for parameter estimation of spatiotemporal systems with correlated measurement noise is considered. Since in the examined setting the correlation structure of the noise is not known exactly, the ordinary least squares method is supposed to be used for estimation and the determinant of the covariance matrix of the resulting estimator is adopted as the measure of estimation accuracy. This design criterion is to be minimized by choosing a set of spatiotemporal measurement locations from among a given finite set of candidate locations. To make the problem computationally tractable for large sensor networks, its relaxed formulation is considered. As the resulting problem is nonconvex, a majorization-minimization algorithmic framework is employed. Thus, at each iteration, a convex tangent surrogate function that majorizes the original nonconvex design criterion is minimized using simplicial decomposition. This results in a sequence of iterates which monotonically reduce the value of the original nonconvex

design criterion. The approximate design produced in this manner then forms a basis for computation of the appropriate exact design using the branch-and-bound technique.

## DESIGNING AND CONDUCTING DISCRETE CHOICE EXPERIMENTS WITH THE R-PACKAGE IDEFIX

FRITS TRAETS
*KU Leuven*
and
MARTINA VANDEBROEK*
*KU Leuven*

Discrete choice experiments are widely used in a broad area of research fields to capture the preference structure of respondents. The design of such experiments will determine to a large extent the accuracy with which the preference parameters can be estimated. This presentation presents a new R-package, called idefix, which enables users to generate optimal designs for discrete choice experiments based on the multinomial logit model. In addition, the package provides the necessary tools to set up online surveys with the possibility of making use of the individual adaptive sequential Bayesian design approach for estimating the mixed logit model. After data collection the package can be used to transform the data into the necessary format in order to use existing estimation software in R.

## BAYESIAN DESIGN OF PHYSICAL EXPERIMENTS FOR NONLINEAR AND COMPUTATIONAL MODELS

TIM WAITE*
*University of Manchester*
DAVE WOODS
*University of Southampton*
and
YIOLANDA ENGLEZOU
*KIOS Research Center, University of Cyprus*

We discuss Bayesian decision-theoretic optimal design of physical experiments for parameter estimation in nonlinear models, including models that incorporate an expensive computer simulator.

In the Bayesian approach, one key challenge is the presence of analytically intractable nested integrals in the expected utility (e.g. expected Shannon information gain) of any proposed design. We propose new Monte Carlo approaches for approximate numerical integration of the expected utility. These methods combine features of importance sampling and Laplace approximations, and give reduced bias and computational expense compared to several existing methods.

Another challenge is that, when the model incorporates an expensive computer simulator, it is prohibitively costly to use an expected utility estimate that relies on direct evaluations of the simulator. Hence in order to perform design optimization for the physical experiment, and also to conduct subsequent inference, one must use a computationally cheap surrogate model in place of the simulator. We accomplish this using a Gaussian process emulator built with pre-existing training data from a computer experiment, thereby extending the analysis framework from the calibration literature to the design problem.

The proposed fully Bayesian framework enables uncertainty about the simulator output at untested input combinations to be incorporated when designing the physical experiment.

## OPTIMAL SUBSAMPLING FOR SOFTMAX REGRESSION

YAQIONG YAO
*University of Connecticut*
and
HAIYING WANG*
*University of Connecticut*

To meet the challenge of massive data, Wang et al. (2018) developed an optimal subsampling method for logistic regression. The purpose of this paper is to extend their method to softmax regression, which is also called multinomial logistic regression and is commonly used to model data with multiple categorical responses. We first derive the asymptotic distribution of the general subsampling estimator, and then derive optimal subsampling probabilities under the A-optimality criterion and the L-optimality criterion with a specific L matrix. Since the optimal subsampling probabilities depend on the unknowns, we adopt a two-stage adaptive procedure to address this issue and use numerical simulations to demonstrate its performance. We derive optimal subsampling probabilities for both sampling with replacement and Poisson subsampling.

*Keywords:* Massive data, Sampling with replacement, Poisson subsampling

## DEVELOPMENT OF THE CHARACTERISTIC FUNCTIONS TOOLBOX FOR MATLAB

VIKTOR WITKOVSKÝ*
*Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia*

The exact distributions of many estimators and test statistics can be specified by their characteristic functions. In particular, the exact null and/or null-distribution of the most common likelihood ratio based test statistics in multivariate analysis. However, derivation of the exact distribution functions by using the analytical inversion of the Fourier transform is frequently too complicated. Here we present development of the repository of characteristic functions and tools for their combinations and numerical inversion in MATLAB. The Characteristic Functions Toolbox (CharFunTool), available at https://github.com/witkovsky/CharFunTool, consists of a set of algorithms for evaluating the selected characteristic functions and algorithms for numerical inversion of the combined and/or compound characteristic functions, used to evaluate the cumulative distribution function (CDF), the probability density function (PDF), and/or the quantile function (QF). The toolbox comprises different inversion algorithms, including those based on the Gil-Pelaez inversion formulae in combination with the simple trapezoidal quadrature rule, or other more sofisticated quadratures and advanced acceleration methods, used for computing the required Fourier transform integrals of oscillatory functions. New methods and algorithms are being continuously developed and implemented into the Toolbox. We shall present current status and basic functionality of the toolbox and illustrate its usage for selected problems of statistical inference.

**Acknowledgements** This research was supported by the Slovak Research and Development Agency, project APVV-15-0295, and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, projects VEGA 2/0054/18 and VEGA 2/0081/19.

## DESIGN OF EXPERIMENTS WITH FUNCTIONAL INDEPENDENT VARIABLES

DAVID WOODS*
*University of Southampton*
and
MARIA ADAMOU
*University of Southampton*

In this talk, some novel methodology will be presented for the optimal design of experiments when at least one independent variable is a function (e.g. of time) and can be varied continuously during a single run of the experiment. Hence, finding a design becomes a question of choosing functions to define this variation for each run in the experiment. The work is motivated by, and applied to, experiments in the pharmaceutical industry.

## OPTIMAL AND ROBUST ACTIVE LEARNING

XIAOJIAN XU*
*Brock University*
and
ANTHONY GRECO
*Brock University*

In this paper, we discuss the common nature of active learning and optimal designs. The techniques developed in optimal and robust regression designs can be implemented for optimally selecting the training data in an active learning process. We consider approximate linear regression models and weighted least squares estimation. Both optimal weighting schemes and optimal design for selecting the training data are investigated for various scenarios. Analytical forms of optimal design densities with given weights, robust design densities with optimal weight functions are derived for general linear regression. The simulation results and comparison study using practical examples indicate improved efficiencies.

## BAYESIAN AND NON-BAYESIAN OPTIMAL DESIGN FOR THE LOGISTIC MIXED EFFECT MODEL

NOHA YOUSSEF*
*American University in Cairo*

Finding an a Bayesian or a Non Bayesian optimal design for a Logistic mixed effects model involves the computation of Fisher information matrix. The calculations of the second partial derivatives and expectations are required, however, they are not always analytically tractable. Therefore, in this study we suggest the use of two different strategies to simplify the calculations. The first is the integrated Laplace approximation (INLA) and the second is to use the observed Fisher information matrix rather than the expected one. The two suggested strategies are compared with other existing methods in the literature of optimal designs for the generalized mixed effects models.

## ENERGY MINIMIZATION FOR SPACE-FILLING DESIGNS

LUC PRONZATO
*CNRS-UNSA, Sophia Antipolis, France*
and
ANATOLY ZHIGLJAVSKY*
*Cardiff university, UK*

A standard objective in computer experiments is to approximate the behaviour of an unknown function on a compact domain from a few evaluations inside the domain. When little is known about the function, space-filling design is advisable: typically, points of evaluation spread out across the available space are obtained by minimizing a geometrical (for instance, minimax-distance) or a discrepancy criterion measuring distance to uniformity. In this talk we outline connections between design for integration (quadrature design), construction of the (continuous) BLUE for the location model, and minimization of energy (kernel discrepancy) for signed measures. Integrally strictly positive definite kernels define strictly convex energy functionals, with an equivalence between the notions of potential and directional derivative, showing the strong relation between discrepancy minimization and more traditional design of optimal experiments. In particular, kernel herding algorithms, which are special instances of vertex-direction methods used in optimal design, can be applied to the construction of point sequences with suitable space-filling properties.

**List of Participants**

| | | |
|---|---|---|
| Selin D. | Ahipasaoglu | ahipasa@gmail.com |
| Giacomo | Aletti | giacomo.aletti@unimi.it |
| Mariano | Amo-Salas | Mariano.Amo@uclm.es |
| Anthony C. | Atkinson | a.c.atkinson@lse.ac.uk |
| Rosemary A. | Bailey | rab24@st-andrews.ac.uk |
| Eva | Benková | Eva.Benkova@fmph.uniba.sk |
| Stefanie G. M. | Biedermann | s.biedermann@soton.ac.uk |
| Wolfgang | Bischoff | wolfgang.bischoff@ku.de |
| Barbara | Bogacka | b.bogacka@qmul.ac.uk |
| Aylin | Cakiroglu | Aylin.cakiroglu@crick.ac.uk |
| Víctor | Casero-Alonso | victormanuel.casero@uclm.es |
| Martina | Chvosteková | martina.chvostekova@savba.sk |
| Carlos | de la Calle Arroyo | carlos.callearroyo@uclm.es |
| Laura | Deldossi | laura.deldossi@unicatt.it |
| Valerii V. | Fedorov | v.v.fedorov2011@gmail.com |
| Katarzyna | Filipiak | katarzyna.filipiak@put.poznan.pl |
| Lenka | Filova | filova@fmph.uniba.sk |
| Nancy | Flournoy | flournoyn@missouri.edu |
| Roberto | Fontana | roberto.fontana@polito.it |
| Fritjof | Freise | fritjof.freise@tu-dortmund.de |
| Irene | García Camacha Gutiérrez | irene.garciacamacha@uclm.es |
| David | Ginsbourger | ginsbourger@idiap.ch |
| Peter P. | Goos | peter.goos@kuleuven.be |
| Heiko | Großmann | heiko.grossmann@ovgu.de |
| Markus | Hainy | markus.hainy@jku.at |
| Radoslav | Harman | harman@fmph.uniba.sk |
| Ralf-Dieter | Hilgers | rhilgers@ukaachen.de |
| Andrew C. | Hooker | andrew.hooker@farmbio.uu.se |
| Anastasia | Ivanova | aivanova@bios.unc.edu |
| Bradley A. | Jones | Bradley.jones@jmp.com |
| Joachim | Kunert | joachim.kunert@udo.edu |
| Adam C. | Lane | adam.lane@cchmc.org |
| Sergei L. | Leonov | Sergei.Leonov@cslbehring.com |
| Jesús | López-Fidalgo | fidalgo@unav.es |
| Nadja | Malevich | nadja.malevich@tu-dortmund.de |
| Irene | Mariñas-Collado | irenemc@usal.es |
| Hugo | Maruri-Aguilar | H.Maruri-Aguilar@qmul.ac.uk |
| Caterina | May | caterina.may@uniupo.it |

| | | |
|---|---|---|
| Tobias | MIELKE | tmielke1@its.jnj.com |
| Werner G. | MÜLLER | werner.mueller@jku.at |
| Christine H. | MÜLLER | cmueller@statistik.tu-dortmund.de |
| Joseph | MURÉ | joseph.mure@edf.fr |
| Kalliopi | MYLONA | kalliopi.mylona@kcl.ac.uk |
| Christopher J. | NACHTSHEIM | nacht001@umn.edu |
| Andrej | PÁZMAN | Andrej.Pazman@fmph.uniba.sk |
| Andrey | PEPELYSHEV | pepelyshevan@cardiff.ac.uk |
| Elena | PESCE | pesce@dima.unige.it |
| Sergio | POZUELO-CAMPOS | Sergio.Pozuelo@uclm.es |
| Luc | PRONZATO | pronzato@i3s.unice.fr |
| Maryna | PRUS | maryna.prus@ovgu.de |
| Martin | RADLOFF | martin.radloff@ovgu.de |
| Fabio | RAPALLO | fabio.rapallo@uniupo.it |
| Andreas | RAPPOLD | andreas.rappold@gmx.net |
| Juan M. | RODRÍGUEZ-DÍAZ | juanmrod@usal.es |
| Samuel | ROSA | samuel.rosa@fmph.uniba.sk |
| William F. | ROSENBERGER | wrosenbe@gmu.edu |
| Frank | RÖTTGER | frank.roettger@ovgu.de |
| Guillaume | SAGNOL | sagnol@math.tu-berlin.de |
| Thomas | SCHMELTER | thomas.schmelter@bayer.com |
| Rainer | SCHWABE | rainer.schwabe@ovgu.de |
| Rakhi | SINGH | rakhi.singh@tu-dortmund.de |
| Bernhard | SPANGL | bernhard.spangl@boku.ac.at |
| Katarína | STERNMÜLLEROVÁ | katarina.sternmullerova@fmph.uniba.sk |
| Mia S. | TACKNEY | M.S.Tackney@soton.ac.uk |
| Aretha | TECKENTRUP | a.teckentrup@ed.ac.uk |
| Paula C. | TRANDAFIR | camelia@unavarra.es |
| Michaela | TUROŠÍKOVÁ | michaela.turosikova@gmail.com |
| Dariusz | UCIŃSKI | d.ucinski@issi.uz.zgora.pl |
| Pablo | URRUCHI | pablo_urruchi@hotmail.com |
| Martina | VANDEBROEK | martina.vandebroek@kuleuven.be |
| Timothy W. | WAITE | timothy.waite@manchester.ac.uk |
| HaiYing | WANG | haiying.wang@uconn.edu |
| Viktor | WITKOVSKÝ | witkovsky@savba.sk |
| Dave | WOODS | D.Woods@southampton.ac.uk |
| Xiaojian | XU | xxu@brocku.ca |
| Elham | YOUSEFI | elham.yousefi@jku.at |
| Noha A. | YOUSSEF | nayoussef@aucegypt.edu |
| Anatoly | ZHIGLJAVSKY | Zhigljavskyaa@cardiff.ac.uk |