technische universität dortmund

SFB 823

16th Workshop on Quality Improvement Methods

Hotel Esplanade
Burgwall 3
D-44135 Dortmund
Tel: +49 231 58530

June 1st and 2nd, 2018

Probabilistic construction and properties of gamma processes and extensions

Fr 14:10

Sophie Mercier

Université de Pau et des Pays de l'Adour

Reliability theory is devoted to the study of lifetimes of industrial systems, together with the evolution of their deterioration level over time. In this context, the standard gamma process is widely used for modeling deterioration which is accumulating over time (non decreasing deterioration level). The point of this presentation is to make some review on the probabilistic construction of the gamma process through series representation and on its jumps structure. Some univariate extensions will next be reviewed, which can help to overcome some possibly restrictive properties of the gamma process in an applicative context.

On a goodness-of-fit test for the gamma process

Fr 14:55

Christian Paroissin

Université de Pau et des Pays de l'Adour

Degradation models are more and more studied and used in practice. Most of these models are based on Lévy processes. For such models, estimation methods have been proposed. These models are also considered for developing complex and efficient maintenance policies. However, a main issue remains: goodness-of-fit (GoF) test for these models. In this talk, we propose a GoF test for the homogeneous gamma process under a general sampling scheme.

Preventive maintenance in degradation processes

Fr 16:00

Waltraud Kahle

Otto-von-Guericke University

In the talk, we consider the Wiener process with drift for degradation modelling. We assume that a failure occurs when the degradation process reaches a given level h for first time. Regularly, inspections are carried out, and the level of degradation is measured. Based on these measurements we consider the following problems:

- Optimal replacement: If the degradation at some inspection point is larger than a predefined level a, a preventive renewal is carried out. There are three possible kinds of costs: costs of inspection, costs of (preventive) renewal, and costs of a failure. The aim is to find an optimal distance between inspection points as well as an optimal level a which minimizes the long-run costs of the system.
- Influence of preventive maintenance to residual life: Here we consider an imperfect maintenance where the system is not renewed but the degradation level is decreased. We compare these methods with Kijima incomplete repair models. Especially, we find the connection between the decrease of the degradation level and the decrease of the virtual age in Kijima models.

Nonparametric Bayes estimation of a system's reliability function

Fr 16:45

AKM Fazlur Rahman and Edsel A. Peña

University of Alabama at Birmingham and University of South Carolina

Coherent systems abound in many disciplines: engineering, industry, etc. It is of high importance that we be able to infer about the failure characteristics of complex systems in order that the quality of such systems could be enhanced. In this talk we present a Bayesian approach to statistical inference, specifically estimation, of a coherent systems components and system reliability functions. When observing a coherent system, at system failure, some of the components might have already failed prior to system failure, while others will be right-censored by the system failure. Our Bayesian approach relies on assigning Partition-Based Dirichlet Measures (PBDM), introduced by Sethuraman and Hollander in a JSPI 2009 article, as prior measures on the components system reliability functions, and then utilizing the systems structure function to obtain an estimator of the systems reliability function from the components reliability functions estimators. The PBDM is particularly appropriate due to the afore-mentioned data accrual on the components when observing a coherent system. We evaluate the resulting estimator of the system reliability function in a frequentist manner and compare its performance with the nonparametric estimator proposed by Doss, Freitag and Proschan (AoS, 1989) and also with the estimator based only on the system failure times. We will demonstrate that though our estimator is more biased than the DFP estimator, it performs better with respect to the root mean squared error (RMSE).

Inference with sequential order statistics

Fr 17:50

Udo Kamps

RWTH Aachen

Sequential order statistics (SOSs) have been introduced as an extension of common order statistics to model successive failures, e.g., in a k-out-of-n system or in other coherent and load sharing systems, where a component failure may cause a change of the underlying hazard rates of remaining system components. SOSs can also be applied to model an alternative step-stress experiment in accelerated life testing. In the model of common order statistics, the hazard rate is supposed to remain the same for all component failures.

The model of SOSs is introduced along with some structural properties (e.g., Markov property, distribution theory, exponential family), and relations to other models of (ascendingly) ordered random variables, such as record values, are shown. Different censored multiple sample set-ups w.r.t. underlying distributions and unknown parameters are considered, and respective univariate and multivariate statistical methods (parameter estimation, confidence regions, tests) are summarized. Some procedures are illustrated via simulations and real data.

Stochastic control principles in antithetic variables and multi-fidelity simulation

Fr 18:35

Henry Wynn

London School of Economics

The principal of merging different sources of information in an optimal way is fundamental to statistics. Perhaps the earliest mention is stratified sampling giving rise to Neymann allocation. In Robust Engineering design strategies are used to create devices which meet target specifications with small variability. In portfolio theory one seeks an optimal trade off between high yield of a portfolio and volatility. Recently, so-called multi level simulation deals with the optimal combination of high fidelity but expensive simulations (solvers) and lower level fidelity but faster simulations. This theory is closely related to the classical "antithetic variable" method, which keeps means or biases fixed while decreasing variability. We suggest that these problems can be conveniently cast in an elementary quadratic stochastic control framework, giving access to some important control principles. The particular application is to multi-scale emulation in computer experiments.

Monitoring compositional data using multivariate EWMA

Sa 9:15

Philippe Castagliola

Université de Nantes

Recently, the monitoring of compositional data by means of control charts has been investigated in the Statistical Process Control literature. In this presentation, we develop a Phase II Multivariate Exponentially Weighted Moving Average (MEWMA) control chart, for the continuous surveillance of compositional data based on a transformation into coordinate representation. We use a Markov chain approximation to determine the performance of the proposed multivariate control chart. The optimal MEWMA smoothing constants, control limits and out-of-control Average Run Lengths have been computed for different combinations of the in-control Average Run Lengths and the number of variables. Several tables are presented and enumerated to show the statistical performance of the proposed control chart. An example illustrates the use of this chart on an industrial problem from a plant in Europe.

The steady-state average run length - modeling and calculation

Sa 10:00

Sven Knoth

Helmut-Schmidt-Universität

At first sight, the definition of a reasonable steady-state measure of control chart detection performance seems to be both straightforward and well established. Classical results are provided by Darroch and Seneta (1965), Roberts (1966), Taylor (1968) and Crosier (1986). Moreover, there are some relationships to change point detection delay measures introduced in Pollak and Siegmund (1975) and Pollak (1985), and more recently in Pollak and Tartakovsky (2009). Here, we want to illustrate how careless several steady-state measures (average run length = ARL, in particular) are used within SPC (statistical process control) literature. Moreover, some useful integral equations (several of them are known) are provided and numerically solved. In particular, univariate and multivariate EWMA (exponentially weighted moving average) control charts are analyzed.

References

[1] Crosier, R.B. (1986). A new two-sided cumulative quality control scheme. *Technometrics* **28**(3), 187–194.

[2] Darroch, J. N., Seneta, E. (1965). On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability* **2**(1), 88–100.

[3] Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Stat.* **13**(1), 206–227.

[4] Pollak, M., Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Stat.* **3**(6), 1267–1282.

[5] Pollak, M., Tartakovsky, A. G. (2009). Optimality properties of the Shiryaev-Roberts procedure. *Statistica Sinica* **19**(4), 1729–1739.

[6] Roberts, S.W. (1966). A comparison of some control chart procedures. *Technometrics* **8**(3), 411–430.

[7] Taylor, H.M. (1968). The Economic Design of Cumulative Sum Control Charts. *Technometrics* **10**(3), 479–488.

Generalized minimum aberration orthogonal arrays Sa 11:00

Roberto Fontana

Politecnico di Torino

We present a methodology and an algorithm to find an Orthogonal Array (OA), of given size and strength, that satisfies the Generalized Minimum Aberration (GMA) criterion, i.e. the lexicographic minimization of the Generalized Word Length Pattern. The methodology is based on the joint use of polynomial counting functions, complex coding of levels and algorithms for quadratic optimization and puts no restriction on the number of levels of each factor, Fontana (2017). We will also briefly discuss some recent developments which are based on a joint work with Ulrike Grömping (Beuth University of Applied Sciences, Berlin, Germany).

In the second part of the presentation, which is based on a joint work with Fabio Rapallo (University of Piemonte Orientale, Alessandria, Italy), we focus on situations where budget constraints or time limitations may occur after the definition of the design, or even when the experiments are running, thus leading to an incomplete design. In such a situation, it is relevant not only to choose an OA with good properties, but also to define an order of the design points, so that the experimenter can stop the sequence of runs and loose as less information as possible. While OAs with added runs are well studied, see for instance Chatzopoulos et al. (2011) and the references therein, less has been done in the case of OAs with removed runs. Some results in this direction can be found in Butler and Ramos (2007). In this work we study how the Generalized Word Length Pattern changes when one or more design points are removed from an OA.

References

[1] Fontana, R. (2017). Generalized minimum aberration mixed-level orthogonal arrays: A general approach based on sequential integer quadratically constrained quadratic programming. *Communications in Statistics-Theory and Methods* **46**(9), 4275–4284.

[2] Chatzopoulos, S.A., Kolyva-Machera, F., Chatterjee, K. (2011). Optimality results on orthogonal arrays plus p runs for s^m factorial experiments. *Metrika* **73**, 385--394.

[3] Butler, N.A., Ramos, V.M. (2007). Optimal additions to and deletions from two-level orthogonal arrays. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69**, 51–61.

An algorithm for generating good mixed level factorial designs

Sa 11:45

Ulrike Grömping

Beuth University of Applied Sciences

Mixed integer programming, implemented with R package **DoE.MIParray** using commercial optimizers Gurobi or Mosek, is applied for the creation of "good" mixed level factorial designs, where "good" refers to (possibly partial) generalized minimum aberration, as introduced by Xu and Wu (2001). The algorithm is presented in Grömping and Fontana (2018); it improves the algorithm of Fontana (2017) by exploiting coding invariance results from Grömping (2018), incorporating lower bounds from Grömping and Xu (2014) and Liu and Lin (2009), and potentially reducing optimization to short word lengths only.

Usefulness of the algorithm is demonstrated on the biotechnological experiment presented in Vasilev et al. (2014) whose (already quite reasonable) design could have been substantially improved.

In addition to this algorithmic approach, a catalog-based approach is discussed for the creation of strong orthogonal arrays based on *optimized* replacement of a multilevel column in a catalogued strong parent array by a (full) factorial experiment (terminology as used by Kuhfeld 2009).

References

[1] Fontana, R. (2017). Generalized minimum aberration mixed-level orthogonal arrays: A general approach based on sequential integer quadratically constrained quadratic programming. *Communications in Statistics-Theory and Methods* **46**(9), 4275–4284.

[2] Grömping, U. (2018). Coding Invariance in Factorial Linear Models and a New Tool for Assessing Combinatorial Equivalence of Factorial Designs. *Journal of Statistical Planning and Inference* **193**(1), 1–14.

[3] Grömping, U., Fontana, R. (2018). An Algorithm for Generating Good Mixed Level Factorial Designs. *Reports in Mathematics, Physics and Chemistry*, Report 1/2018, Dep. II, Beuth University of Applied Sciences Berlin, 1–23.

[4] Grömping, U., Xu, H. (2014). Generalized Resolution for Orthogonal Arrays. *The Annals of Statistics* **42**(3), 918–939.

[5] Kuhfeld, W. (2010). Orthogonal Arrays.
`http://support.sas.com/techsup/technote/ts723.html`.

[6] Liu, M.Q., Lin, D.K.J. (2009). Construction of Optimal Mixed-Level Supersaturated Designs. *Statistica Sinica* **19**, 197–211.

[7] Vasilev, N., Schmitz, Ch., Grömping, U., Fischer, R., Schillberg, S. (2014). Assessment of Cultivation Factors that Affect Biomass and Geraniol Production in Transgenic Tobacco Cell Suspension Cultures. *PLoS ONE* **9**(8:e104620), 1–7.

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

Optimal design of experiments for large datasets Sa 14:00

Elena Pesce and Eva Riccomagno
Università degli Studi di Genova

We refer on three papers that propose 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. Theoretical and computational comparisons are made.

References

[1] Wiens, D.P. (2018). I-robust and D-robust designs on a finite design space. *Statistics and Computing* **28**(2), 241–258.

[2] Drovandi, C.C., Holmes, C., McGree, J.M., Mengersen, K., Richardson, S., Ryan, E.G. (2017). Principles of Experimental Design for Big Data Analysis. *Statistical Science* **32**(3), 385–404.

[3] Pesce, E., Riccomagno, E., Wynn, H.P. (2017). Passive and active observation: experimental design issues in big data, `arXiv:1712.06916 [stat.ME]`

Screening non-normally distributed responses Sa 14:45

John Tyssedal

Norges teknisk-naturvitenskapelige universitet

Screening is about separating out the subset of active factors from the others, and two-level designs have for long been considered to be the preferred choices for screening. Responses have often been assumed to be normally distributed, models assumed to be linear and the analysis, although not always trivial, has benefited from the widely developed theory of linear models. However, there are a lot of practical situations where the assumptions of linear models and normally distributed responses are not even close to be valid. Typical examples are count data, often modelled as binomial or Poisson distributed. If the data are counts of rare events, the normal distribution is clearly not a good approximation, and for these two distributions, the variances and expectations are in all cases closely related. Another challenge is the choice of screening design for such responses and for nonlinear models an optimal design will not only depend on knowing the model, but also the distribution, the size of the parameters and for a GLM modelling also the link function. In this paper we test out the screening performances for three popular screening designs, a definite screening design a minimum resolution IV design and a Plackett-Burman design when four distributions, two binomial, one gamma and one Poisson are chosen for the response values. For each distribution, we test out and compare if it is best to use the raw data, a variance stabilizing transformation of the data or perform a generalized linear modelling.