

Abstracts

ROeS 2007 BERN

	Page
<u>Session 1: Selection procedures and bias</u>	5
David Draper: Model Mis-Specification Bias and Model Uncertainty: Bayesian Solutions	5
Amy Racine: Bayesian Model Averaging	6
Michael Schemper, Samo Wakounig and Georg Heinze: Weighted Estimation in Cox Regression Revisited	7
Franz König, Werner Brannath, Martin Posch and Peter Bauer: Adaptive Designs and Treatment Selection	8
<u>Session 2a: Free Papers (I)</u>	9
Andreas Krause and Jose Pinheiro: Modeling and Simulation to Adjust P-values in Presence of a Regression to the Mean Effect	9
Jouni Kerman, Beat Neuenschwander, Michael Branson: Bayesian Monitoring of Drug Safety Data	10
Georg Ferber: The Maximum Mean Difference - Statistical Problems in Assessing Cardiac Safety	11
Petra Ofner-Kopeinig, Maximilian Errath and Andrea Berghold: Simulation Results of Randomization Methods Using Randomizer	13
M. Ahmad Rauf: Analysis of High Dimensional Repeated Measures Designs: Two-Sample Case	14

<u>Session 2b: Young Statisticians (II)</u>	15
Paola Berchialla: Bayesian Network for Pathological Scarring Due to Burn Injuries.....	15
Natasa Kejzar: Probabilistic Inductive Classes of Graphs.....	16
Marek Sedlacik: Use of ROC Curves in Diagnostics of Glaucoma.....	17
Matthias Templ: Problems and Possibilities of Cluster Analysis: Application on Geochemical Data.....	18
Lucie Doudova: Statistical Inference on Negative Binomial Parameters Estimation.....	19
Jaromir Tonner: Bayesian Approach to Estimation of Time-Variable Parameters.....	20
Sonja Zehetmayer: Repeated significance tests controlling the False Discovery Rate.....	22
<u>Session 3: Methods in meta analysis</u>	23
Alex Sutton: Evidence Synthesis and Improving Efficiency in How We Evaluate Healthcare.....	23
Mathias Egger: Meta-Analysis: Perspective from Clinical Epidemiology.....	24
Marc Buyse: Why Meta-Analyses Should be Based on Individual Patient Data.....	25
Martina Mittlböck, Harald Heinzl: Assessing Heterogeneity in Meta Analyses.....	26

<u>Session 4a: Free Papers (II)</u>	27
Gerta Rücker, G. Schwarzer, J.R. Carpenter: Sample Size Distribution - an Overlooked Source of Spurious Bias in Funnel Plots....	27
Klaas Prins: Modeling and Simulation using metadata: Away from inventing the wheel again.....	29
Martin Posch: Adaptive Truncated Sequential Tests and the Bonferroni Procedure.....	31
Ronald Gieschke, Paul Jordan: Use of Stochastic Differential Equations for Pharmacokinetic Modelling.....	32
Valentin Rousson: A mixed approach for proving non-inferiority with respect to binary endpoints.....	33
<u>Session 4b: Young Statisticians (II)</u>	34
Michael Vock: One-Sided Confidence Regions for a Multivariate Location Parameter.....	34
Michaela Paul: Surveillance of Infectious Disease Data using Cumulative Sum Methods.....	35
Andrea Riebler: Bayesian Variable Selection for Detecting Adaptive Genomic Differences among Populations.....	36
Antonia Zapf: Area under the curve (AUC) adjusted for covariates.....	37
Nicolas Fournier: Gene expression data, survival analysis and power.....	38
Frank Konietzschke: Nonparametric Analysis of Clustered-Data in Diagnostic Trials.....	39
Kaspar Rufibach: On a new approach to regression with ordinal explanatory variables.....	40

<u>Session 5: Statistical dose finding strategies</u>	41
Peter Müller: Bayesian Decision Theoretic Dose Finding: A Principled Decision Theoretic Approach to Dose Finding, Including Simulation Based Sequential Design.....	41
Holger Dette, Frank Bretz, Jose Pinheiro and Andrey Pepelyshev: Optimal Designs for Dose Finding Studies	42
Ulrich Beyer: Dose Finding Supported by Trial Simulation: An Example from the Practice.....	43
Frank Bretz: Adaptive Dose Ranging Studies.....	44
<u>Session 6: Statistical Methods in genetics and proteomics</u>	45
Paul Eilers: Statistical Classification for Reliable High-Volume Genetic.....	45
Darlene Goldstein: Gene Expression Signatures in Human Disease: An Illustration with Genomic Data on Breast Cancer.....	46
Alexandra Goll: Two-Stage Designs for Proteomic and Gene Expression Studies Applying Methods Differing in Costs.....	47
Michael G. Schimek: False Discovery Rate Procedures in Practice: A Critical Appraisal.....	48

Session 1: Selection procedures and bias

Model Mis-Specification Bias and Model Uncertainty: Bayesian Solutions

David Draper

University of California, Santa Cruz, USA

In a typical application of the statistical paradigm, there's some quantity Q about which I'm at least partially uncertain, and I wish to quantify my uncertainty about Q , for the purpose of (a) sharing this information with other people or (b) helping myself or others to make a choice in the face of this uncertainty. Uncertainty quantification is usually based on a probability model M , which relates Q to known quantities (such as data values D); M will in turn be based on assumptions and judgments on my part about how Q and D are related, but I'm not always certain about the "right" assumptions and judgments to make. To be completely honest, then, I have to acknowledge two sources of uncertainty: I'm uncertain about Q , and I'm also uncertain about how to quantify my uncertainty about Q . This second source is model uncertainty; and if I get the model "wrong," the result will typically be model mis-specification bias.

In this talk I'll speak to two questions: why model uncertainty and model mis-specification bias matter, and what to do about them. Topics to be addressed, in the context of one or more applied examples, will include

- (1) What is a statistical model (the standard parametric answer; de Finetti's predictive answer based on exchangeability);
- (2) Where do models come from (in practice many specification details often come from looking at the data);
- (3) Consequences of failure to acknowledge model uncertainty and model mis-specification bias (if the right price is not paid in (2), the result will often be poor calibration of inferential and predictive statements); and
- (4) Methods for coping with model uncertainty and model mis-specification bias (Bayesian cross-validation, Bayesian model averaging, Bayesian nonparametric methods).

Dose Finding under order constraint using Bayesian Model Averaging in Clinical Development

Amy Racine, David Ohlssen

Novartis Pharma AG, Basel

The Bayesian approach to dose finding in clinical trials has a modeling and a decision-making component. The former necessitates a flexible class of dose-response models (parametric or nonparametric), whereas the latter relies on decisions that are based on appropriate inferential summaries and can be either informal (posterior- and/or predictive-based) or formal, i.e., fully decision-analytic using utilities. Dose-response modeling can be done in various ways, ranging from simple models with a small number of parameters (typical for early development trials) to high-dimensional or non-parametric models. Often in early clinical development, little is known about the possible shape of the dose response relationship within the studied dose range (typical 3 to 4 active doses), except it is monotone. If there are d dose levels (including placebo) to be studied, there are $2^d (=K)$ models describing the possible dose response relationships (models). Bayesian model averaging (BMA) [1] is possible strategy that tries to estimate the dose response for decision making. The basis is a set of K simple dose-response models M_k with model parameters (mean response for each dose level and variance) θ_k ($k=1, K$). Then, starting from prior model probabilities (“weights”) π_k as well as prior distributions on the model-specific parameters θ_k , standard Bayesian inference leads to posterior updates of the unknown parameters π_k and θ_k [2]. Bayesian model averaging is conceptually straight-forward and can be seen as a special case of a hierarchical model. The approach is well-suited for situations where the quantity of interest is model-independent, such as in dose-finding studies where the objective is to find a dose fulfilling a certain pre-specified criterion, e.g. the effect of dose exceeding the clinical relevant threshold. Bayesian model averaging generalizes model selection strategies and has the advantage of weighting the candidate models in an appropriate (data-dependent and therefore evidence-based) way.

We shall introduce Bayesian model averaging using a motivating example in drug development, in which the choice of prior weight and the prior of the model parameter and implementation methodologies will be discussed. Results will be compared to that of a classical approach.

Reference

[1] J.A. Hoeting, D. Madigan, A.E. Raftery and C.T. Volinsky “Bayesian Model Averaging; A tutorial” *Statistical Science*, 1999 Vol 14 No. 4 382-417.

[2] D. Draper “Assessment and Propagation of Model Uncertainty” *JRSS B* 1995 Vol 57 No. 1 45-97.

Weighted Estimation in Cox Regression Revisited

Michael Schemper, Samo Wakounig and Georg Heinze

Section of Clinical Biometrics, Core Unit for Medical Statistics and Informatics
Medical University of Vienna, Austria; michael.schemper@meduniwien.ac.at

Often the effect of at least one of the prognostic factors included in a Cox regression model (CR) changes over time, which violates the proportional hazards assumption of this model. As a consequence, the *average relative risk* (over time) for such a prognostic factor is under- or overestimated and the efficiency of parameter estimates decreases. While there are several methods to appropriately cope with non-proportional hazards (cf., e.g., Lehr and Schemper, 2007), in particular by including parameters for time-dependent effects, weighted estimation in Cox regression (WCR) is a parsimonious alternative without additional parameters. Furthermore, WCR extends the univariate tests by Breslow and by Prentice to multiple covariates and interactions as does CR for Mantel's test. Also WCR is more robust to outlying observations than CR. The methodology has been presented by Schemper (1992) and by Sasieni (1993) but has fallen into oblivion despite its usefulness for samples with a limited number of events and/or in the presence of many covariates.

Therefore, in this presentation we revisit WCR, define the concept of an *average hazard ratio* (cf. Kalbfleisch and Prentice, 1981) in mathematical and intuitive terms and deal with the use of Schoenfeld residuals and time-dependent effects terms within WCR. Finally, we report results of recent empirical investigations comparing WCR and CR in the statistical analysis of a colon cancer study of the Vienna Medical University, and in a comprehensive Monte Carlo study of efficiency and bias. The conclusions from the empirical investigations permit us to recommend WCR, which is implemented in a SAS macro and in an upcoming R package available at: www.muw.ac.at/msi/biometrie/programs.

The project is supported by grant P18553-N13 of the Austrian Science Fund.

References

- Kalbfleisch, J.D. and Prentice, R.L. (1981). Estimation of the average hazard ratio. *Biometrika*, **68**, 105-112.
- Lehr, S. and Schemper, M. (2007). Parsimonious analysis of time-dependent effects in the Cox model. *Statistics in Medicine*, **26**, 2686-2698.
- Sasieni, P. (1993). Maximum weighted partial likelihood estimators of the Cox model. *Journal of the American Statistical Association*, **88**, 144-152.
- Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *The Statistician*, **41**, 455-465.

Keywords average Hazard Ratio, Proportional Hazards Model, Survival Analysis, Time-Dependent Effect.

Adaptive Designs and Treatment Selection

Franz König, Werner Brannath, Martin Posch, Peter Bauer

Section of Medical Statistics, Medical University of Vienna, Spitalgasse 23,
1090 Vienna, Austria

Clinical trials incorporating treatment selection at pre-specified interim analyses allow to integrate two clinical studies into a single, confirmatory study. A standard application is the combination of a (late) clinical phase II study (focusing on treatment selection) with a (early) phase III study (confirmatory testing of the selected treatments) into a single adaptive seamless design. Note that when analyzing the data of the selected treatments in a conventional way may bias the test statistic and inflate the multiple type I error rate. Adaptive (flexible) designs based on the closure principle allow the selection of treatments in the interim analysis still controlling the multiple type I error rate. The construction of simultaneous confidence intervals in such adaptive designs is an open research topic. Here we will give some suggestions how to plan adaptive seamless designs with simultaneous confidence intervals and compare their statistical properties. Additionally, we investigate the bias and mean squared error of naïve estimates.

Session 2a: Free Papers (I)

Modeling and Simulation to Adjust p-values in Presence of a Regression to the Mean Effect

Andreas Krause*, Jose Pinheiro**

*Pharsight Corporation, **Novartis Pharmaceuticals

Exploratory analysis methods, such as data mining, make it more common than ever to analyze data in a retrospective fashion, for example, to detect subgroups of interest. Such post-hoc approaches can easily be misleading, with an “effect” being found that is nothing more than an artifact of the data. One reason for such a finding – which is actually nonexistent – can be the well-known phenomenon of regression to the mean. It poses a challenge that, if ignored, can generate incorrect claims of significance.

We illustrate the regression to the mean phenomenon by an illustrative example and propose an often easily implemented general approach to adjust such findings for potential confounding effects. The approach relies on the derivation of a reference distribution for calculated p-values, based on modeling and simulation.

The methodology allows for complex models to be used, because the simulations do not require that one works out the correlation structure of the parameter estimates analytically to adjust the regression to the mean effect. It also allows for a straightforward assessment of the goodness of the fit, following the idea of posterior predictive checks.

We illustrate the approach with the adjustment of the result from a recent study on patients suffering from Alzheimer’s disease for which an unplanned subgroup selection was used after study completion.

Bayesian Monitoring of Drug Safety Data

Jouni Kerman, Beat Neuenschwander, Michael Branson
Novartis Pharma AG, Basel

Adverse events that are recorded during clinical trials are a cause of concern. Methods to monitor these events are needed. Since a treatment is often used in several trials it should be of interest to try to combine all of the information available to estimate the true adverse event rate ratio (with respect to the rate in the placebo arm), providing us a more realistic estimate than by looking at each trial separately. In other words, it is plausible that the observations from similar trials should be also exchangeable to some extent. However, pooling all data together ignores between-trial variation and treats the trials to be essentially identical, producing estimates with too narrow confidence intervals. Multilevel modeling (random and fixed effects meta-analysis) provides a method to synthesise the observed adverse events to form more realistic estimates. By modeling the between-trial variation, we construct multilevel estimates that offer a better picture of the safety profile of a drug than one would obtain by simply pooling of event counts over trials, or by looking at each trial separately. The approach yields also predictions of adverse event rate ratios for future trials.

The Maximum Mean Difference - Statistical Problems in Assessing Cardiac Safety

Georg Ferber
Novartis Pharma AG, Basel

The guidance E14 "The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs" of the International Conference for Harmonisation probably contains one of the first attempts to assess clinical safety with confirmatory statistical methods. It asks for a "Thorough QT/QTc study", which is a placebo and active controlled pharmacodynamic trial with the QT interval (corrected and uncorrected) measured at a number of time points. It defines "*a negative 'thorough QT/QTc study'*" as "[...] *one in which the upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug on the QTc interval excludes 10 ms.*"

Taken literally, this calls for a noninferiority test on the maximum over a number of time points. In practice, though, this condition has been replaced by an intersection-union test, requesting that for each of the time points over which the maximum is to be taken, a conventional confidence interval should exclude the critical value of 10 ms. However, attempts have been made to address the request of the guideline literally. In particular, Eaton et al (2006) approximated the maximum by a k-norm and taking the limit $k \rightarrow \infty$ and Boos et al. (2007) investigate, among others, bootstrap approaches.

A point that proved to be central to the statistical debate accompanying the finalisation of the guideline was the size of the "bias" introduced by this endpoint, that is the difference between the "expected (with respect to the study population) maximum (over the timepoints) of the mean difference" and the "Maximum of the expected mean difference". In fact, in the trivial case that the test drug has no effect whatsoever, the former will be positive, whereas the latter obviously is nought. The size and behaviour of this bias has been subject to a number of empirical investigations. Understanding this behaviour and supporting it by real data have influenced the definition of the margin for the noninferiority test, which was moved from 7.5 ms to 8 and finally to 10 ms.

References:

Boos DD, Hoffman D, Kringle R, Zhang J: New confidence bounds for QT studies. *Statistics in medicine*, electronically published Feb 2007.

Eaton ML, Muirhead RJ, Manusco JY, Kolluri S: A confidence interval for the maximal mean QT interval change caused by drug effect. *Drug information journal* **40** (2006), pp 267 - 271

International conference on harmonisation: The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. Step 4, May 2005
(http://www.ich.org/UrlGrpServer.jsr?@_ID=475&@_TEMPLATE=272#E14)

Simulation Results of Randomization Methods Using Randomizer

Petra Ofner-Kopeinig, Maximilian Errath and Andrea Berghold

Institute for Medical Informatics, Statistics and Documentation, Medical
University of Graz, Austria

In randomized clinical trials patients are allocated to different treatments at random to avoid biases. Restricted and unrestricted randomization methods can be used. When patients are allocated completely at random (unrestricted randomization) this can result in an undesirable imbalance in the numbers of patients allocated to the treatment groups, especially in small trials. To avoid imbalance of patient numbers in treatment groups restricted randomization methods, such as permuted block randomization, minimization, biased coin randomization or big stick randomization methods are used.

The Randomizer is a web-based patient randomization service for multi-center clinical trials, developed at the Institute for Medical Informatics, Statistics and Documentation at Medical University of Graz. The application provides unrestricted randomization as well as 5 different restricted randomization methods. Software validation was performed according to the FDA guidelines; GCP-compliance was confirmed in 2005 by the Austrian Agency for Health and Food Safety ([AGES](#)).

The randomization methods implemented in the Randomizer were validated using the integrated simulation tool, a part of the software for generating static randomization lists and testing trial designs. Tests were performed by determining the relative frequencies of the occurrence of absolute differences in treatment group sizes. To estimate probabilities of occurrence with an accuracy of 1 % and a confidence interval of 95 %, each simulation was performed 10000 times. For each randomization method different settings have been simulated. The balance behaviors will be discussed and are compared with theoretical results.

Analysis of High Dimensional Repeated Measures Designs: Two-Sample Case

M. Rauf Ahmad and E. Brunner
University of Göttingen, Germany

We consider two groups of independent subjects, of sizes n_1 and n_2 , where each subject in each group is repeatedly observed at d fixed time points. This type of design set up can be formulated for several possible variants of analyses, including a two-sample multivariate analysis and profile analysis. When $d < n$, there are well-established methods available for the analysis of such a design set up under the assumption of multivariate normality. Recently, there have been serious attempts to propose the methods of analysis for the case $d > n$. The need arises in a variety of research disciplines, more frequently in biostatistics and statistical genetics.

We propose a high dimensional version of the ANOVA-type statistic (see e.g. Brunner and Puri, 2001), based on the idea of Box-approximation (Box, 1954). Quadratic and bilinear forms are used to develop the components of the test statistic the distribution of which is approximated by a χ^2 distribution. The estimators, thus defined, are unbiased and ratio-consistent irrespective of the dimension d . The developed statistic can be used to test the simple hypothesis of the equality of two multivariate mean vectors or can also be used to test profile hypotheses.

Simulations are carried out for several parameter settings which show that the type I error rate is well maintained even for the sample size as small as 10. Moreover, the procedure does not depend on the underlying structure of the covariance matrix. The statistic is compared to other recently developed statistics for the same set up. The application of the new procedure is illustrated using real data examples from medicine.

Keywords: Repeated measures, High dimensional, Quadratic and bilinear forms

References

1. Box, G. E. P.(1954). Some theorems on quadratic forms applied in the study of analysis of variance problems II: Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics*, 25(3), 484-498.
2. Brunner, E. and M. L. Puri(2001). Nonparametric methods in factorial designs. *Statistical Papers*, 42, 1-52.
3. Srivastava, M. S.(2007). Multivariate theory for analyzing high dimensional data. *Journal of Japanese Statistical Society*, 37, To appear.
4. Srivastava, M. S. and Y. Fujikoshi (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis*, 97, 1927- 1940.

Session 2b: Young Statisticians (I)

Bayesian Network for Pathological Scarring Due to Burn Injuries

Paola Berchiolla¹, Maurizio Stella², Ezio Nicola Gangemi², Dario Gregori¹

¹Department of Public Health and Microbiology, University of Torino

²Department of Plastic and Reconstructive Surgery – Burn Center,
Traumatological Center, Torino

The development of hypertrophic scars is a major problem following burn injuries. The healing process varies greatly among patients, and the risk of a bad scar evolution is unpredictable. It also results in the cost it imposes to the burnt patients' quality of life.

Only few studies reported an analysis in terms of prevalence and description of pathological scarring after burns and the impact of certain risk factors is still poorly understood. The present work is aimed at determining those factors associated with an increased risk of the development of hypertrophic burn scars. A retrospective analysis of clinical records at the Burn Outpatient Clinic in Turin between January 1994 and May 15, 2006 was carried out. A Bayesian Network (BN) was introduced to model relationships among risk factors and contemporary to allow for the presence of multiple burn scars per patient. Combining in a coherent way qualitative and quantitative data, a Bayesian Network was built up through a structural learning process which enabled the system to learn the dependencies among variables. Thus the BN model provided and validated a classification model taking into consideration the uncertainty associated with clustered data. It also provided a straightforward tool to enter evidence and make inferences about risk factors. The identification of the principal risk factors for pathological scarring after burns should be valuable in early risk stratification and may help in assessing outcomes adjusted for patient risk. It should facilitate the choice of optimal medical prevention, helping physicians to target follow-up timing and aggressive or experimental therapies to those likely to high risk.

Probabilistic Inductive Classes Of Graphs

Natasa Kejzar
University of Ljubljana

Graphs are skeletons of networks. Many models for network evolution model only this skeleton. Inductive definitions of graphs enables one to describe the evolution of a graph in some prescribed manners. The transitions (transformations by rules) can be viewed as implicit time steps. We extend the standard notion of inductive classes of graphs (ICGs) by imposing a probability space in the choice of (1) graphs from basis, (2) rules and (3) their initial elements and call it probabilistic ICGs (PICGs).

Many of the existing models of growing networks can be cast as PICGs. We will look at some graph/network generations and generations of processes on graphs according to specific PICG rules which serve as models for specific real-world networks (i.e. spread of rumors among people, changing the structure of links in a fixed social network in time etc.). Results regarding expected size and order are derived and the method for determining asymptotic degree distribution in such graphs is presented.

Use of ROC Curves in Diagnostics of Glaucoma

Marek Sedlacík

University of Defence, Brno (CZ)

The receiver operating characteristic (ROC) curve is widely applied in measuring discriminatory ability of diagnostic tests. This makes ROC curves one of the most actively used research techniques in various branches. In this contribution the attention is concentrated on methods of ROC curve estimation. Several approaches are mentioned: the classical estimator based on the empirical CDF, the weighted regression estimator, the estimator based on the best unbiased CDF estimator, the estimator based on the latent random variable or GLM estimators. These methods were programmed in the computing software MATLAB and the procedures are applied to the ophthalmological real data processing to detect the glaucoma.

Key words: ROC curve; Estimation; Medical diagnostic testing.

Problems and Possibilities of Cluster Analysis: Application on Geochemical Data

Matthias Templ, Peter Filzmoser (both Vienna University of Technology),
Clemens Reimann (Geological Survey of Norway)

Cluster analysis is a method for finding groups in multivariate data without providing any information about group membership (unsupervised classification). Many different clustering algorithms have been proposed in the literature, and many methods are implemented in R. Unfortunately, for real data sets without obvious grouping structure, different cluster algorithms will in general give slightly different results, sometimes even completely different results. For the user it would thus be important to know which clustering methods are “ideally suited” for analyzing the data at hand.

The results of cluster analysis applied on complex data depends strongly on different parameters and assumptions. Firstly, one have to consider an appropriate transformation, especially when the data are of compositional nature, for example. For most of real data it is also useful to standardise it in respect to the different scale of variables. Furthermore, the user has to consider an appropriate distance measure, since most of the existing clustering algorithms will use the distance matrix to start. In addition to that, the selected number of clusters plays an important role in finding good clusters. When using bootstrap samples of the data, one can also consider the stability of results arising from different clustering algorithms and different distance measures with the RAND index. Most of the additionally available cluster validity measures are quite useful regarding to (artificial) data with a strong clustering structure but they provide more or less poor results when applied on real complex data. We want to give deeper insights into these problems of cluster analysis by using real complex data sets from geochemistry. A package (clustTool) for clustering data is developed and downloadable on CRAN.

Especially for the clustering of geochemical data, there is a graphical user interface implemented in package clustTool. Since cluster analysis and diagnostic tools for cluster results on complex geochemical data were in hand of specialists we want to provide methods and some diagnostic tools to the users with this GUI.

Keywords: Cluster Analysis, Multivariate Methods, Stability, Geochemical Data

Statistical Inference on Negative Binomial Parameters Estimation

Lucie Doudová

University of Defence, Brno (CZ)

The negative binomial distribution (NBD) is often used to modeling the biological data. This distribution depends on two parameters μ - expectation and κ - dispersion parameter. In the paper the likelihood function of (NBD) can be written in form

$$\ln p(y; \mu, \kappa) = \ln \Gamma(y + \kappa) - \ln \Gamma(y + 1) - \ln \Gamma(\kappa) + \kappa \ln \frac{\kappa}{\kappa + \mu} + y \ln \frac{\mu}{\kappa + \mu}$$

There are several different method of parameters estimation. The classical estimators are obtained by the method of moments (MM) or by the maximum likelihood methods (MLE). Then less frequent are methods of estimation based on extended quasi-likelihood (EQL), double extended quasi-likelihood (DEQL) and the method based on bias correction (BC). These methods will be studied in the paper, computer algorithms will be implemented in MATLAB and compared.

Further the model where both parameters depend on covariates will be studied. At first in the situation when parameter κ is known and fixed the generalized linear models (GLM) techniques will be used. The test statistics (Wald and likelihood ratio) will be considered and the power of tests based on these statistics will be compared for different parameter estimation. Similarly when both parameters are unknown the general likelihood approach will be used. Because in many statistical articles the Γ -function in likelihood expression is substituted by the Stirling's approximation, to avoid to work with digamma and trigamma functions, the influence of the substitution on parameter estimation will be numerically studied. Presented methods and results will be applied and compared using real medical data.

Key words : negative binomial distribution, method of moments, maximum likelihood methods, extended quasi-likelihood, double extended quasi-likelihood, bias correction

Bayesian Approach to Estimation of Time-Variable Parameters

Jaromír Tonner

Masaryk university, Faculty of Economics and Administration,
Dept. of Applied Mathematics and Computer Science,
Lipová 41a, Brno-Pisárky, 602 00, Czech Republic
Fax: +420-549-49-1720; Tel: : +420-549-49-1710
E-mail: jtonner@tiscali.cz

The aim of this paper is to identify essential changes in monetary policy by time – ariable parameters estimation of Gal'ı - Monacelli model of real economy via modified Extended Bootstrap Filter Smoother (EBSFS) (see details of toolbox in [7]). First of all it was necessary to transform original concept into the DSGE (Dynamic Stochastic eneral Equilibrium) structure containing the linear rational expectations (LRE model)(see [4]). The solution of the LRE problem can be carried out by well-known numerical Generalized Schur Decomposition (GSD) of appropriate matrices (see [5]) but only in case of time - constant parameters model. In our case it is necessary to solve the GSD in each step of Bootstrap filter running to obtain a solved DSGE model which is prepared to estimate parameters in each period t (Conditional Probability Density Functions of the states and parameters are time variant and not unimodal due to structure changes). For prior parameters estimation (all important procedures are implemented in DYNARE (see [3])) was also chosen Bayesian approach in combination with maximum likelihood method (see [2]) where resulting estimated standard deviations have to be appropriately adjusted for Bootstrap filter (EBSFS) running. Especially we would like to identify changes in DSGE model structure related implicitly with changes of monetary regimes (e.g. transition of fixed exchange rate to floating in the Czech economy in 1997). We expect it must be represented by essential change of DSGE structure which is given by parameters values.

Keywords

symbolically solved DSGE model, Extended Bootstrap Filter Smoother, sequential Monte Carlo method, time - variable parameters estimation, Gal'ı - Monacelli model

JEL: C11, E27, E37, E52

References

- [1] Gal'ı J., Monacelli T. Monetary Policy and Exchange Rate Volatility in a Small Open Economy. Review of Economic Studies (2005) 707-734.
- [2] Hamilton, James D. Time Series Analysis. Princeton: Princeton University Press, 1994.

[3] Juillard, Michel. Dynare manual.
www.cepremap.cnrs.fr/dynare/download/manual

This paper has been worked as a part of research activities at the grant project of GA CR No. 402/05/2172 and with support of MSMT project Research centres 1M0524.

Repeated significance tests controlling the False Discovery Rate

Sonja Zehetmayer

Co-Authors: Martin Posch, Peter Bauer

Section of Medical Statistics, Medical University of Vienna, Spitalgasse 23,
1090 Vienna, Austria

When testing a single hypothesis repeatedly at several interim analyses, adjusted significance levels have to be applied at each interim look to control the overall Type I Error rate. There is a rich literature on such group sequential trials investigating the choice and computation of adjusted critical values. Surprisingly, if a large number of hypotheses are tested controlling the False Discovery Rate (a frequently used error criterion for large scale multiple testing problems), we can show that under quite general conditions no adjustment of the critical value for multiple interim looks is necessary. This holds asymptotically (for a large number of hypotheses) under all scenarios but the global null hypothesis where all null hypotheses are true. Similar results are given for a procedure controlling the per-comparison error rate.

We also investigate the properties of the sequential procedure for the case of finite number of null hypotheses. Based on simulation results and theoretical approximations we show that the False Discovery Rate is well controlled unless the expected number of rejected null hypotheses at the interim analysis where the trial stops is very small.

This finding permits the researcher to perform repeated interim analyses and to stop the trial whenever it seems appropriate without compromising the rate of false discoveries. This might be useful in microarray experiments, where an a priori sample size calculation is often difficult as the power of the experiment depends on many unknown parameters.

Session 3: Methods in Meta analysis

Evidence Synthesis and Improving Efficiency in How We Evaluate Healthcare

Alex Sutton

Department of Health Sciences, University of Leicester, UK

Background: Systematic review and meta-analysis currently underpin much of evidence-based medicine. While such methodologies bring order to previous research, the way future research is planned is still not as coherent or efficient as it could be.

Aims: To outline a cyclic framework for healthcare evaluation, incorporating emerging methodologies, aimed at i) designing future research more efficiently using information contained within the existing evidence-base; and ii) making more efficient use of existing research results reducing the number of future research studies required. Methodologies for updating systematic reviews will also be considered.

Results: The framework presented insists that an up-to-date meta-analysis of existing evidence should always be considered before future trials are conducted. Such a meta-analysis should inform design issues such as sample size determination. Consideration should also be given to how the newly planned study would contribute to the totality of evidence through its incorporation into an updated meta-analysis. We illustrate how new studies can have very low power to change inferences of an existing meta-analysis, particularly when between study heterogeneity is taken into consideration. The contexts in which the use of individual patient data meta-analysis and mixed treatment comparisons modelling may be beneficial before further studies are conducted is considered.

Conclusions: While we have come a long way in making clinical practice more evidence-based, we are lagging some way behind when it comes to future study design and analysis which needs to be made more evidence-based also. The framework presented is a proposal for how this situation can be improved.

Meta-analysis: Perspective from clinical epidemiology
Mathias Egger

Why meta-analyses should be based on individual patient data

Marc Buyse, ScD

IDDI, Louvain-la-Neuve, Belgium
and Center for Statistics, Hasselt University, Belgium

“If a problem is very important and needs an answer, statisticians should not just make do by calculating a weighted estimate based on published data. Instead, they should point out the difficulties of oversimplified approaches and try to get funding for a rigorous scientific investigation.” (Sauerbrei, Blettner and Royston, *J Clin Epidemiol* 2001; **54**: 537-8).

Meta-analysis of clinical trials can either be based on summary data extracted from the literature, or on individual patient data collected from all trials, whether published or not (IPDMAs). IPDMAs have several major advantages:

- (a) **Absence of bias** : by including data from all trials, IPDMAs avoid the well-known problem of publication bias, whereby trials with better results are more likely to be published than those with negative results (Rothstein, Sutton and Borenstein (Editors), *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, Wiley, 2005).
- (b) **Maximization of power** : analyses of individual survival times are more powerful than those based on summary statistics, and they do not require arbitrary assumptions about censoring (Buyse and Ryan, *Statist in Med* 1987; **6**: 565-76; Tudur *et al*, *JRSS A* 2001; **164**: 357-70).
- (c) **Quality assurance** : individual patient data can be scrutinized and checked statistically; in contrast, published reports contain too little information for even rudimentary quality checks to be performed (Clarke *et al*, In: *The Cochrane Library*, Issue 4, 2002, Update Software).
- (d) **Investigations of heterogeneity** : besides estimation of an overall treatment effect, one of the main goals of meta-analysis is to investigate heterogeneity (Berlin, *Am J Epidemiol* 1995; **142**: 383-7). Detailed investigations based on individual prognostic characteristics are only possible in IPDMAs. Meta-regressions based on summary statistics are potentially subject to ecological and/or baseline risk bias (Sharp *et al*, *Br Med J* 1996; **313**: 735-8; Buyse and Piedbois, *Statist in Med* 1996; **15**: 2797-812).
- (e) **Further analyses** : IPDMAs usually include baseline and secondary endpoint data that allow further analyses to be performed, including prognostic factor analyses, subset analyses, indirect treatment comparisons, and validation of surrogate endpoints (Buyse *et al*, *Biostatistics* 2000; **1**: 49-68; Buyse *et al*, *J Clin Oncol* 2003; **21**: 1682-7).

Assessing Heterogeneity in Meta-Analyses

Martina Mittlböck and Harald Heinzl

Core Unit for Medical Statistics and Informatics, Medical University of Vienna
Spitalgasse 23, A-1090 Vienna

A meta-analysis is the combination of clinical trial results in order to arrive at summary conclusions about a medical research question. A major problem is to decide whether the trials under review are homogeneous or heterogeneous. This is an important step which determines the statistical methods to be used and the interpretation of the results. Tests of heterogeneity are not always reliable because of low power for sparse data and the detection of irrelevantly small amount of heterogeneity when many studies are involved. Thus measures of heterogeneity are better suited to determine the amount of between-study variance and its impact on meta-analysis results.

Measures for heterogeneity with different scaling are reviewed. Simulation study results are used to assess various properties of the heterogeneity test and heterogeneity measures.

Heterogeneity test and heterogeneity measures are not directly related to the absolute amount of between-study variance but to the relative increase of variance due to heterogeneity. A heterogeneity measure scaled to a fixed interval needs reference values for proper interpretation. A measure defined by the relation of between- to within-study variance has a more natural interpretation but no upper limit. Investigated heterogeneity measures depend on the variance of the individual effects and thus on the number of patients in the studies.

Meta-analysis of clinical trials with heterogeneous results provides an opportunity to learn about variations in treatment effectiveness. Knowledge about the sources of heterogeneity can ultimately lead to better clinical understanding of the optimal way to treat patients.

References:

- (1) Higgins J, Thompson S. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002;21:1539-1558.
- (2) Mittlböck M, Heinzl H. A simulation study comparing properties of the heterogeneity measures in meta-analyses. *Statistics in Medicine* 2006;25:4321-4333.

Session 4a: Free Papers (II)

Sample Size Distribution - an Overlooked Source of Spurious Bias in Funnel Plots

G Rücker, G Schwarzer, JR Carpenter

Institute of Medical Biometry and Medical Informatics,
University Medical Centre Freiburg,
Germany (ruecker@imbi.uni-freiburg.de)

Background

Comparing the recently proposed arcsine test with established tests for publication bias in meta-analyses with binary outcomes (1), we found that power and size depended on the variability of the sample sizes of the included trials. With little variation in sample size, and no small ($n < 20$) trials, established tests (2; 3) showed greatly inflated type 1 error, which was corrected by recent proposals (4; 5; 6). By contrast, when the included trials had a range of sample sizes from 6 to 1000 or more, established tests performed not worse.

This talk explains these observations, and outlines the advantages of our newly proposed arcsine test.

Methods

Tests on publication bias are based on a funnel plot, which plots standard error versus treatment effect. If the outcome is binary, treatment is often measured by the log odds ratio (logOR). For this measure, the standard error depends on the estimated treatment effect (6). This induces a relationship between the performance of tests for publication bias and the sizes of the trials in the meta-analysis. We discuss the effects and consequences of this.

We further propose using the variance stabilising property of the arcsine transform, thus replacing logOR by the arcsine risk difference in tests for publication bias. We show how

this simple proposal avoids many problems with existing tests.

Results

For fixed sample size and population treatment effects, the log odds ratio is approximately related to its standard error through the hyperbolic cosine. Thus, if a meta-analysis contains a range of study sizes, the final look of the funnel plot arises from a mix of these curves.

We found that excluding small sample sizes, regardless of the treatment effect, systematically removes some of these curves from the mixture and creates an artificial appearance of publication bias. Further, if trial sizes are small, we may

get trials with no events in one or possibly both arms. Excluding these also artificially creates the appearance of publication bias and additionally leads to an overestimate of the treatment effect.

Conclusion

Using an arcsine risk difference alleviates the key problems with the log odds ratio and log risk ratio. Specifically, it

- reduces error inflation in tests for publication bias,
- reduces dependence of tests on sample size distribution,
- does not require to exclude trials with no events in one or both groups,
- is not affected by reversion of the outcome direction (in contrast to the risk ratio) ,
- has a geometric interpretation as an arc length corresponding to the risk difference,
- is easy to calculate.

References

- [1] Rücker G, Schwarzer G, Carpenter JR. Arcsine test for publication bias in meta-analyses with binary outcomes 2007. Provisionally accepted.
- [2] Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50:1088–1101.
- [3] Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997; 315:629–634.
- [4] Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 2006; 25(20):3443–3457.
- [5] Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association* 2006; 295:676–680.
- [6] Schwarzer G, Antes G, Schumacher M. A test for publication bias in meta-analysis with sparse binary data. *Statistics in Medicine* 2007; 26(4):721–733.

Modeling and Simulation using metadata: Away from inventing the wheel again

Klaas Prins
Pharsight Corporation

Today's main issue pharma is facing is the exponential increase in cost associated with drug development and a stagnating number of drugs that make it to the market. Costs rise because of intensified attention to adverse events and more stringent rules for establishing efficacy claims. Traditionally, pharma would go after information on drug treatment vs. placebo or vs. competitors themselves by executing an impressive sequence of trials.

As many companies invest in similar drug therapies, similar information is gathered multiple times (inventing the wheel again). Increasingly, this type of information is available through public sources, including scientific literature, regulatory agency summary bases of approval, meeting proceedings, product inserts, and news releases from academic centers, research institutes, and drug companies, since they are encouraged today to publish all (negative and positive) clinical trial data.

The information is mainly available as summary clinical trial data, as reported in tables and graphs of scientific publications, but also subject-level data might be available. Specific methods exist to extract information from the mentioned sources and compile it in a database (called 'metadata base'). The data bases include information on in- and exclusion criteria, population characteristics (demographics etc...), dropout, and anything else available and considered useful.

The extraction process is preceded by scientific review and discussions involving clinical experts with therapeutic area knowledge, modeling scientists, and database specialists. This metadata base becomes a critical foundation upon which simulation models may be built.

Pharmaceutical scientists increasingly seek the assistance of modeling and simulation (M&S) methods to enhance experiential understanding of go/no-go decisions and improving efficiency in drug development. This technology quantifies knowledge gained from the clinical behavior of comparators and may as such provide invaluable insight regarding managing development of an NCE by turning point estimates into (probabilistic) expectations with uncertainty. M&S techniques can also be utilized to quantify knowledge from summary statistics and appropriate statistical tools are available for making joint inferences from models based on individual data as well as summary statistic data.

As such, M&S of metadata broadens the horizon for drug development. It opens up quantitative information on placebo and competitor drugs across a number of covariates that would not be accessible through execution of clinical trials. It cannot replace all clinical trials, but it may reduce the number of futile trials and improve the information density obtained from future trials.

We will discuss specific methods for data extraction, database design, and recommendations for information review and selection, best practices, leveraging into M&S analyses and inference making, and a case study illustrates a recent application.

Adaptive Truncated Sequential Tests and the Bonferroni Procedure

Martin Posch

Medical University of Vienna

A simple construction principle for truncated sequential tests is proposed. An application of this principle gives a uniform improvement of the Bonferroni test. The stopping boundaries are defined via conditional expectations of a positive test statistics. The testing procedure stops with a rejection of the null hypothesis as soon as the conditional expectation of this test statistics, given the data observed so far, is larger or equal to one. This gives a computationally efficient procedure to compute stopping boundaries without the need for multiple integration. Additionally, we construct truncated sequential tests including futility boundaries. Finally it is discussed how adaptations (as a modification of the maximal sample size) can be implemented.

Use of Stochastic Differential Equations for Pharmacokinetic Modeling

Ronald Gieschke, Paul Jordan

Biomathematics, Pharma Development, F. Hoffmann – La Roche, Basel

Pharmacokinetic modeling is often the empirical approach to fit models of adequate complexity to blood concentration data. These models are generally derived from pharmacokinetic processes determining the fate of the drug in the body. Structural model misspecification is generally subsumed in the residual error. Stochastic differential equations allow a decomposition of the total error into process error (Wiener noise) and measurement error. Moreover, they allow tracking the time course of model parameters which are traditionally handled as time independent values.

The software NONMEM can be regarded as the golden standard for nonlinear mixed effects pharmacokinetic modeling, allowing parameter estimation for user-defined differential equation systems. Estimation of parameters in stochastic differential equations became possible with the current NONMEM version 6 which allows to access and change integration variables during numerical integration. Consequently, is it possible to implement the extended Kalman filter, an optimal filtering technique, in NONMEM.

We developed an R suite of programs that simulated, by means of the Euler-Maruyama method, single subject data according to a given stochastic differential equation and fitted the results with NONMEM 6. The following questions were explored: How well can a process error be detected for different signal/noise ratios, and how well can a structural model misspecification be determined. We found that the detection of process noise requires dense sampling and a high Wiener noise/measurement noise ratio. Similarly, structural model misspecification, as assessed by the difference in objective function values between the misspecified (ODE) and SDE model, could only be shown for large signal/noise ratios.

Further investigations in multiple subject data (population pharmacokinetics) will show whether these results can be generalized.

References:

Tornøe CW, Overgaard RV, Agerso H, Nielsen HA, Madsen H: Stochastic Differential Equations in NONMEM: Implementation, application, and comparison with ordinary differential equations. *Pharm Res* 22(8): 1247-1258 (2005)

Higham DJ: An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM REVIEW* Vol. 43 (3) 525–546 (2001)

A mixed approach for proving non-inferiority with respect to binary endpoints

Valentin Rousson, Burkhardt Seifert

University of Zurich

When a new treatment is compared to an established one in a randomized study, it is standard practice to statistically test for equivalence (or for non-inferiority) rather than for significance. When the endpoint is binary, one usually compares two treatments using either an odds-ratio or a difference of proportions. In this talk, we propose a mixed approach which uses an odds-ratio to define "practical equivalence" and which uses then a difference of proportions to show non-inferiority. The mixed approach is shown to be more powerful than the conventional odds-ratio approach when the efficiency of the established treatment is known (with good precision) and when it is high (with more than 50% of success), as is often the case. The gain of power achieved may lead in turn to a substantial reduction in the sample size needed to prove non-inferiority. The method can be generalized to ordinal endpoints.

Session 4b: Young Statisticians (II)

One-Sided Confidence Regions for a Multivariate Location Parameter

Michael Vock,
IMSV, University of Bern

We investigate confidence regions obtained by the inversion of a one-sided test for a multivariate location parameter. These confidence regions may have an unpleasant shape. In particular, if the null and alternative hypothesis are both composite and complementary, the confidence region usually does not resemble the alternative parameter region in shape, but rather a reflected version of the null parameter region. Such a confidence region may be difficult to interpret, unnecessarily large, and therefore not very useful.

We illustrate this effect and show one possibility of obtaining confidence regions for the location parameter that are smaller and have a more suitable shape for the type of problems investigated. This method is based on the closed testing principle applied to a family of nested hypotheses.

Surveillance of Infectious Disease Data using Cumulative Sum Methods

Michaela Paul[†], Michael Höhle , Leonhard Held[†]

[†] Institute for Social- and Preventive Medicine, University of Zurich
Hirschengraben 84
8001 Zurich, Switzerland
Department of Statistics, University of Munich, Germany
michaela.paul@ifspm.uzh.ch

The ongoing systematic collection and analysis of public health data is essential to prevent and control the spread of infectious diseases. In this talk, we will present methods for the prospective surveillance of time series of counts and proportions. The methods are based on cumulative sum (CUSUM) schemes developed for statistical process control (SPC) applications. Standard CUSUM schemes assume that the in-control parameter is constant, however, infectious diseases often show seasonal variations in the mean. We therefore examined two approaches that allow for a time-varying in-control parameter. In the approach of Rogerson and Yamada (2004) the standard CUSUM method itself is modified and applied to the original observations. A different and also more common choice is to use a model for the in-control parameter, compute some choice of approximate normal residuals based on this model and then monitor these residuals with a standard CUSUM for normal variates.

By using direct Monte Carlo simulation we showed better performance in terms of the run length of Rogerson and Yamada's¹⁾ approach for the original data. In specific situations the performance of the method using normal residuals as proposed e.g. by Rossi et al.²⁾ proved to be very unstable and lead to substantially more false alarms.

1) Rogerson, P. A. and I. Yamada (2004). *Approaches to syndromic surveillance when data consist of small regional counts. Morbidity and Mortality Weekly Report 53/Supplement, 79–85.*

2) Rossi, G., L. Lampugnani, and M. Marchi (1999). *An approximate CUSUM procedure for surveillance of health events. Statistics in Medicine 18, 2111–2122.*

Bayesian Variable Selection for Detecting Adaptive Genomic Differences among Populations

Andrea Riebler†, Leonhard Held† and Wolfgang Stephan*

† Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich,

8001 Zurich, Switzerland

* Section of Evolutionary Biology, Department of Biology II, University of Munich, 82152 Planegg-Martinsried, Germany

An important goal of population genetics is to detect DNA regions (loci) affected by natural selection. Differential adaptation can lead to large between-population differences in allele frequency counts. This genetic differentiation can be quantified using the F_{st} -coefficient. In this talk, we will present a F_{st} -based Bayesian hierarchical model implemented via Markov chain Monte Carlo (MCMC), for the detection of loci that might be subject to selection (Beaumont and Balding, 2004). This approach divides the F_{st} -influencing factors into locus-specific effects, population-specific effects and effects that are specific for the locus in combination with the population. We will extend this model by introducing Bayesian variable selection to automatically select non-neutral locus effects. Using gene frequency datasets mainly simulated from a Wright-Fisher model with migration we compare the statistical power of the different methods to detect non-neutrally behaving loci.

References

Beaumont, M. A. and Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans, *Molecular Ecology* 13: 969–980.

Area under the curve (AUC) adjusted for covariates

Antonia Zapf and Edgar Brunner

University of Göttingen
Department of Medical Statistics

Essential elements of medical statistics are studies, comparing two diagnostic or therapy methods. A widely accepted measure to assess the overall accuracy of a diagnostic test is the area under the receiver operating characteristic curve (AUC).

Frequently covariates have a strong effect on a diagnostic result or the success of a therapy. In randomized clinical trials, but in case of baseline imbalances or in epidemiological studies the distributions of the covariates may be very different in the two groups. In this case, biased results may be obtained when the statistics are not adjusted for covariates.

The aim of this talk is to present some ideas how to adjust the AUC for covariates. The proposed ranking procedures are able to handle metric as well as ordered categorical response variables and covariates.

In this talk, I will present rank procedures for the analysis of such designs. Also estimators and confidence intervals adjusted for covariates will be discussed.

Gene Expression Data, Survival Analysis and Power¹

Nicolas Fournier

EPFL – SB IMA STAP, Station 8
CH – 1015 Lausanne

We link gene expression data to survival analysis by constructing a genetic risk factor that predicts reappearance of metastasis after breast cancer surgery. The expression of nearly 25'000 different genes in 295 patients was provided by a former study from the Netherlands².

At first we identify the genes with significant differential expression between two groups of patients (patient free of metastasis for at least 5 years after their surgical intervention, and others). To isolate such genes, we use a modification of the Student's t-test for unpaired observations, the SAM method³ (Significance Analysis of Microarrays).

We then use the expressions of the selected genes and construct the risk factor. We use a correlation measure between a patient's genetic profile and an average good profile which is determined with healthy patients, and create three different risk groups. Non-parametric (Kaplan-Meier survival curves estimates and log-rank test) and semi-parametric analysis (Cox's proportional hazards model) are then performed to attest the quality of the classification.

Finally, we try to evaluate the effects of cross-validation on the power of our methods. To achieve this, we generate a population composed of two groups facing different risks, and covariates, that would help determining which group the patient belongs to. We then perform a semi-parametric analysis (Cox's model), using a cross-validation procedure to detect the different risks faced. An estimation of the power of this method is finally obtained by repeating the whole procedure several times (Monte Carlo sampling).

¹ Nicolas Fournier, Expression génétique, analyse de survie et puissance sous cross-validation, Master's project, EPFL, 2007.

² Marc van de Vijver et al., A gene expression signature as a predictor of survival in breast cancer, New England Journal of Medicine, vol 347, no 25, 2002, pp. 1999-2009.

³ Virginia G. Tusher et al., Significance analysis of microarrays applied to the ionizing radiation response, Proceedings of the National Academy of Sciences, vol 98, no 9, 2001, pp. 5116-5121

Nonparametric Analysis of Clustered-Data in Diagnostic Trials

Frank Konietzschke and Edgar Brunner

University of Göttingen
Department of Medical Statistics

Diagnostic trials are important clinical trials in medical research. Therefore it is important to derive statistical methods and models for such trials, in particular for diagnostic trials with multiple readers and methods. Such designs are requested by Guidelines or “Notes for guidance” from regulatory authorities.

In diagnostic trials, clustered data are obtained when several parts of the same patient are observed at a time where no, several, or all parts may be diseased or non-diseased. This means that rather complicated correlation structures are to be taken care of.

The PhD thesis “Nonparametric Analysis of Diagnostic Trials, Göttingen 2006” by Carola Werner discusses nonparametric methods for clustered data in multiple reader studies where a new approach is developed to evaluate such complex designs. The disadvantage of the proposed procedures is, that the test statistic (a quadratic form) can become negative if certain conditions on sample sizes are not fulfilled.

In this presentation I will propose a new method to overcome this disadvantage. In a simulation study, the results are compared with those obtained by Carola Werner’s method.

On a new approach to regression with ordinal explanatory variables

We consider the general regression problem of modeling $y \in \mathbb{R}$ based on $w \in \mathbb{R}^p$. Therefore, we are given a training sample (y_i, w_i) for $i = 1, \dots, n$ and aim to maximize a concave real-valued criterion function $L(\beta) = L(\beta, y, W)$ over some finite-dimensional parameter $\beta \in \mathbb{R}^p$, where $W = (w_i^\top)_{i=1}^n \in \mathbb{R}^{n \times p}$. Examples we are aiming at are linear, logistic, and Cox regression.

Assume that the first f columns $w_{\cdot j} = (w_{ij})_{i=1}^n, j = 1, \dots, f$ of W are ordinal variables, each with k_j groups. Examples for such type of data are measurements of disease stages in the analysis of oncology data.

To estimate β we introduce the new data matrix $X \in \mathbb{R}^{n \times d}$ by replacing each column $w_{\cdot j}$ in W by $k_j - 1$ columns consisting of dummy variables for each but the first group of the ordinal variable. In order to respect the ordinal character of the $w_{\cdot j}$'s, $j = 1, \dots, f$, we constrain maximization of the updated functional $L(\beta) = L(\beta, y, X)$ to the following space of parameters:

$$\mathcal{B}(f, k_1, \dots, k_f) = \{\beta \in \mathbb{R}^d : \beta_{j, l+1} - \beta_{j, l} \geq 0, j = 1, \dots, f, l = 2, \dots, k_j - 1\}$$

where $\beta_{j, l}$ is the coefficient of the dummy variable corresponding to the level l of the j -th ordinal variable. This construction ensures that the estimated parameter corresponding to a “higher” group of an ordinal variable is at least as big as those of “lower” groups.

Some very specific cases allow maximization of L via the pool-adjacent-violaters algorithm. In general, $\hat{\beta} := \arg \max_{\beta \in \mathcal{B}} L(\beta)$ can be computed in finitely many steps and very efficiently via an active set algorithm. Characterizations of $\hat{\beta}$ connecting the constrained and the unconstrained maximizers are given. The fast algorithms allow bootstrap inference in these models. Simulations comparing these new to existing approaches will be presented. Finally, an application to real data from oncology is indicated.

Session 5: Statistical dose finding strategies

Bayesian Decision Theoretic Dose Finding: a Principled Decision Theoretic Approach to Dose Finding, Including Simulation Based Sequential Design.

Peter Müller

Department of Biostatistics
The University of Texas M. D. Anderson Cancer Center
Houston, USA

We describe the use of a successful combination of Bayesian inference and decision theory in clinical trial design. The discussion follows two specific examples. Both applications implement decision theoretic dose-finding.

One application is for two-agent dose finding. The trial includes a pragmatic, but ad-hoc initial exploration of the bivariate dose-response surface followed by a second stage that is based on a formal decision theoretic argument.

The second trial uses a decision theoretic design throughout. The trial involves three important decisions, adaptive dose allocation, optimal stopping of the trial, and the optimal terminal decision upon stopping. For all three decisions we use a formal Bayesian decision-theoretic approach.

The two applications demonstrate how Bayesian posterior inference and decision-theoretic approaches combine to provide a coherent solution in a complex application.

The main challenges are the need for a flexible probability model for the unknown dose-response curve, a delayed response, the sequential nature of the stopping decision, and the complex considerations involved in the terminal decision.

References:

Thall, P., Millikan, R., Mueller, P., and Lee, S-J. (2003). "Dose-Finding with two agents in phase I oncology trials." *Biometrics*, 59 (3), 487.

Mueller, P., Berry, D.A., Grieve, A.P., and Krams, M. (2006). "A Bayesian Decision-Theoretic Dose Finding Trial" *Decision Analysis*, 3, 197-207.

Optimal Designs for Dose Finding Studies

Holger Dette¹⁾, Frank Bretz²⁾, Jose Pinheiro²⁾, Andrey Pepelyshev³⁾,

¹⁾ Ruhr-Universität Bochum, ²⁾ Novartis Pharmaceuticals, ³⁾ St Petersburg State University, Russia

Identifying the "right" dose is a very critical step in the clinical development process of any medicinal drug. Selecting too high a dose can result in unacceptable toxicity and associated safety problems, while choosing too low a dose leads to smaller chances of showing sufficient effectiveness in confirmatory trials. In this paper we consider the problem of obtaining dose finding designs, efficiently learning about the minimum effective dose (MED), defined as the smallest dose which shows a clinically relevant and a statistically significant effect. The sensitivity of these designs with respect to misspecification in the true parameter and in the underlying dose-response model is investigated. Robust optimal designs are constructed taking into account, simultaneously, a set of potential dose response profiles within classes of models commonly used.

Dose finding supported by trial simulation: An example from the practice

Ulrich Beyer

F. Hoffmann-La Roche, Basel

In the development of a new compound, the selection of an adequate dose and schedule is one of the crucial steps and involves usually pk, efficacy and safety data analyses as well as modeling and simulation approaches. This selection process is discussed on the basis of an example of a long acting erythropoietic agent recently developed. The dose selection for this compound was not straightforward as the required doses are individualized and differ from patient to patient. The presentation will describe the phase II clinical studies to determine an appropriate starting dose for phase III and the use of modeling and simulation at end of phase II, to support a successful labeling support.

Adaptive Dose Ranging Studies

Frank Bretz

Novartis Pharma AG, Basel
(frank.bretz@novartis.com)

Despite revolutionary improvements in basic biomedical science, the number of new drug applications submitted to the FDA (U.S. Food and Drug Administration) has shown a declining trend over the past several years. Both the FDA and PhRMA (Pharmaceutical Research and Manufacturers of America) have started initiatives to identify and address the main drivers leading to the pharmaceutical industry's current problems. One well-known such driver is poor dose selection resulting from incorrect or incomplete knowledge of the dose response relationship for both efficacy and safety.

This talk will discuss innovative adaptive dose ranging studies aimed at striking a balance between additional dose response information and increased costs/timelines. In these studies, the number of doses and/or the allocation of patients to doses are allowed to change during the study, as increasing efficacy and safety information becomes available. PhRMA has formed a working group on adaptive dose ranging studies, whose work will be presented at this talk. Different types of adaptive dose ranging designs and methods focusing on Phase II trials will be described together with the results of a comprehensive simulation study evaluating the performance of the methods under a variety of trial scenarios. Recommendations on the practical use and potential gains associated with adaptive dose ranging studies will be discussed.

Session 6: Statistical methods in Genetics and Proteomics

Statistical Classification for Reliable High-volume Genetic Measurements

Paul H.C. Eilers

Methodology and Statistics
Faculty of Social and Behavioural Sciences
Utrecht University, The Netherlands
e-mail: p.h.c.eilers@uu.nl

Single nucleotide polymorphisms (SNPs, pronounced as "snips") are mutations in which only one of the bases (A, C, G or T) that make up our DNA has changed. SNPs occur very frequently (one million or more on the whole genome) and they can tell a lot about changes in DNA, between people or between tumors and normal tissue. Modern technology allows the measurement of (hundreds of) thousands of SNPs at the same time. Unfortunately, depending of the method used and the quality of the biological samples, the measurements are not perfect. Advanced statistical classification methods are very useful to improve the determination of genotypes and to quantify reliability. I will describe the application of mixtures of regression models in this context.

When using SNP results to compare tumors and normal tissue, new challenges show up. We do have spatial information, the location of SNPs on chromosomes. Because genomic changes tend to show spatial correlations, the task is to classify changed segments.

Key words:

Em-Algorithm, Parametric Mixtures, Single Nucleotide, Polymorphism (SNP)

Gene Expression Signatures in Human Disease: An Illustration with Genomic Data on Breast Cancer

Darlene Goldstein
University of Lausanne

Gene expression profiling is gaining increasing prominence for subtype identification, diagnosis and prognosis in human disease, particularly for cancers. However, genes constituting the signatures vary across studies, with little overlap.

In this talk, I will introduce the concept of biologically based coexpression modules - sets of genes with highly correlated expression - and provide a unification of proposed signatures using quantitative measures of module activity. The methods will be illustrated with a comprehensive analysis of publicly available expression data on 2833 breast cancer tumors.

This is joint work with Pratyaksha Wirapati and Mauro Delorenzi of the Swiss Institute of Bioinformatics.

Two-stage designs for proteomic and gene expression studies applying methods differing in costs

Alexander Goll
Medical University of Vienna

In gene expression and proteomic studies we generally deal with large numbers of hypotheses, where only for a small fraction of the hypotheses noticeable effects exist. Due to limited resources, the number of observations per hypotheses in a conventional single-stage design is low which limits the power. It has been shown that two-stage pilot and integrated designs are a good option to improve the power. In these sequential designs, the first stage is used to screen for the promising hypotheses, which are further investigated in the second stage. In the following we more thoroughly investigate this type of two-stage designs where the costs per measurement and effect sizes differ between the first and second stage. To compare different designs we assume that the total costs of the experiment are fixed. Both integrated and pilot designs are based on procedures either controlling the family wise type 1 error rate (FWE) or the false discovery rate (FDR). Two scenarios are considered: In the first scenario the experimenter from the beginning may have the choice between two methods that differ in costs and effect sizes (a low-cost standard method or a high-cost improved method). In the second scenario different costs per measurement may arise if the same method is applied at both stages but specific experimental devices have to be produced at higher costs per measurement for the selected markers at the second stage. For the first scenario we show that depending on the cost and the effect size ratios between the methods it is preferable either to apply the low-cost or the high-cost method at both stages. For the second scenario we will show for which cost ratios between stages it is worthwhile to use (optimal) two-stage designs as compared to the single stage design. Finally we also look how design misspecifications in the planning phase would change the power of two-stage designs as compared to the single-stage design.

False Discovery Rate Procedures in Practice: a Critical Appraisal

Michael G. Schimek¹⁾ and Tomas Pavlık²⁾

¹⁾Medical University of Graz, Institute for Medical Informatics, Statistics and Documentation, Auenbruggerplatz 2, 8036 Graz, Austria,
michael.schimek@meduni-graz.at

²⁾Masaryk University, Department of Applied Mathematics and Institute of Biostatistics and Analyses, 66295 Brno, Czech Republic

The microarray technology developed in recent years allows for measuring expression levels of thousands of genes simultaneously. In most microarray experiments comparative measurements are taken from subjects belonging to pre-specified groups. Statistical procedures to identify differentially expressed genes involve a serious multiple comparison problem as we have to carry out as many hypothesis testings as the number of candidate genes in the experiment. If we apply the usual type I error rate α in each testing, then the probability for rejecting any truly null hypothesis will greatly exceed the intended overall α level. The false discovery rate (abbr. FDR) is a key tool for type I error control.

Although overall FDR control is not always sufficient to obtain reliable test results (see e.g. Ploner et al., 2006, for a local approach), up to now the majority of differential gene expression techniques is based on global parameters. The well-known problems due to a sometimes substantial variability of measured gene expression, specifically in the presence of low variance genes, is usually tackled by 'correcting' constants in the estimation of the pooled variance.

In this presentation our main interest is the popular SAM approach ('Significance analysis of Microarrays', Chu et al., 2005), the most frequently applied software in molecular biology and medicine. It comprises a FDR estimation concept (Tusher, Tibshirani and Chu, 2001). Moreover we study some alternative procedures (Schwender, Krause and Ickstadt, 2003; Grant, Liu and Stoeckert, 2005).

Via simulations we compare these procedures for the ordinary as well as the modified t-test statistic under various choices of the 'correcting' constant ('fudge factor' in SAM jargon). Special emphasis is given to the influence of correlated measurements due to co-expressed genes. We demonstrate that the comparative findings obtained for the above FDR procedures depend much on the adopted type of correlation model. As a matter of fact, biologist's knowledge of structural features of genetic pathways is still very limited. We biostatisticians can only formally compare these procedures assuming various, hopefully realistic, scenarios.

Finally we draw conclusions for the practical use of SAM and other FDR techniques available, also with respect to the choice of 'correcting' constants. In addition we consider the perspectives of 'automatic' empirical Bayes procedures avoiding FDR estimation at all.

References:

Chu, G. et al. (2005). SAM Significance Analysis of Microarrays -User guide and technical document. Technical Report, Stanford University.

Grant, G.R., Liu, J. and Stoeckert Jr., C.J. (2005). A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics*, 21, 2684–2690.

Ploner, A. et al. (2006). Multidimensional local false discovery rate for microarray studies. *Bioinformatics*, 22, 556–565

Schwender, H., Krause, A. and Ickstadt, K. (2003). Comparison of empirical Bayes and the Significance Analysis of Microarrays. Technical Report, SFB 475, University of Dortmund.

Tusher, V., Tibshirani, R. and Chu, C. (2001). Significance analysis of microarrays applied to transcriptional responses to ionizing radiation. *Proceedings of the National Academy of Sciences*, 98, 5116–5121.