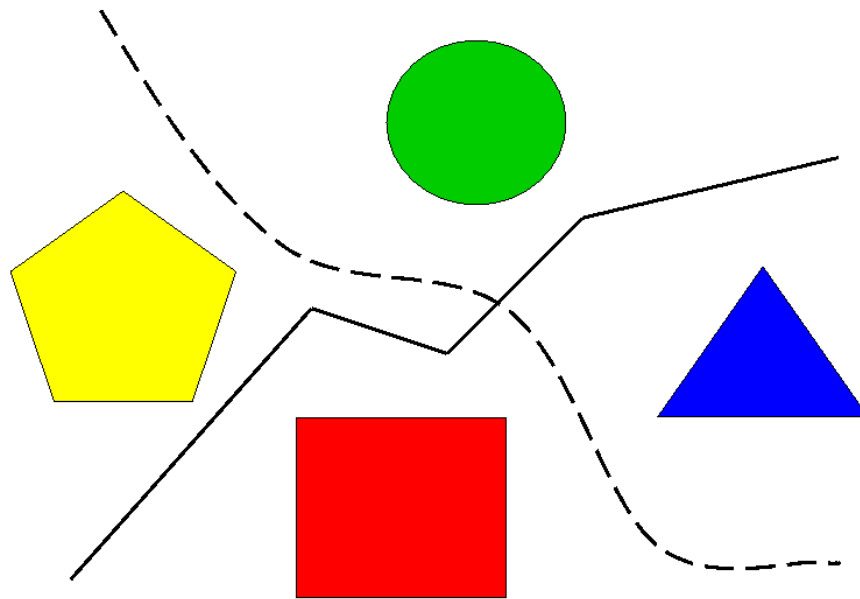


3. Herbstkolloquium des Graduiertenkollegs "Statistische Modellbildung"



Statistische Modellbildung

Zu diesem Kolloquium wird eingeladen

am Freitag / Samstag, 24./25. November 2006,

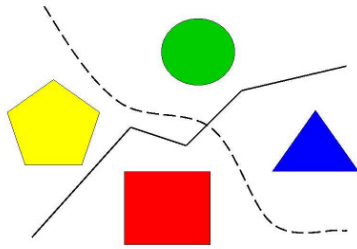
UNIVERSITÄTSKOLLEG BOMMERHOLZ

- Lehr- und Weiterbildungsstätte der Universität Dortmund -

Bommerholzer Straße 60, 58456 Witten.

(Tel.: ++49 (0)2302 / 39 60, Fax: ++49 (0)2302 / 39 63 20)





Statistische Modellbildung

3. Herbstkolloquium des Graduiertenkollegs "Statistische Modellbildung"

Freitag, 24. November 2006

Abfahrt nach Witten: ab Dortmund gegen 14.00 Uhr

Vortragsprogramm I

- 15.30 h Begrüßung**
Prof. Dr. Joachim Hartung
- 15.45 h Göran Kauermann**
Fakultät für Wirtschaftswissenschaften
Universität Bielefeld **Penalisierte Spline Schätzung**
- 16.30 h Gernot A. Fink**
Fachbereich Informatik / Institut für
Roboterforschung, Universität Dortmund **Grundlagen der automatischen
Spracherkennung**
- 17.15 h Vorstellung der Mitglieder des Kollegs**
Einführung in die Posterausstellung
- 19.00 h Posterausstellung:**
Präsentation der Dissertations- und Postdoc-Projekte
Diskussionen in Arbeitsgruppen

Samstag, 25. November 2006

Vortragsprogramm II

9.00 h **Shalabh**
Department of Mathematics and Statistics,
Indian Institute of Technology, Kanpur, India
/ Fakultät für Mathematik, Informatik und
Statistik, Ludwig-Maximilians-Universität
München

**Parametric estimation
in measurement error models**

9.45 h **Ingileif B. Hallgrímsson**
Department of Statistics
University of Oxford, United Kingdom

**Detecting gene interactions in case-
control association studies**

10.30 h **Pause**

Vortragsprogramm III

11.00 h **Michael Hauptmann**
Netherlands Cancer Institute
Amsterdam, The Netherlands

**Statistical methods
for assessing latency
in epidemiologic studies**

11.45 h **Werner Brannath**
Institut für Medizinische Statistik
Medizinische Universität Wien, Österreich

**Adaptive Treatment Selection
and Efficacy Testing in Clinical Trials**

12.30 h **Pause**

Vortragsprogramm IV

14.00 h **Sonja Kuhnt**
Department of Mathematics and Computer
Science, Eindhoven Technical University,
The Netherlands

**Graphical Models:
Outliers and Model Choice**

14.45 h **Uwe Ligges**
Fachbereich Statistik, Universität Dortmund

**Simulationen in R:
Effiziente Programme,
Wahl von Betriebssystem
und CPU optimierter BLAS**

15.30 h **Pause**

Werner Brannath:

Adaptive Treatment Selection and Efficacy Testing in Clinical Trials

Modern clinical trial designs with a selection of treatments at an interim analysis based on cumulated interim data allows investigators to combine different clinical trial phases in the development process of drugs. For this reason the underlying statistical methods have attracted much interest in academic institutions, pharmaceutical companies and regulatory bodies like the FDA in the USA or the BfArM in Germany. Of particular interest are the flexible two stage closed tests proposed by Bauer and Kieser (1999) which are currently under an intensive discussion. The underlying principle of flexible two stage tests allows investigators to control the multiple type I error rate while selecting treatments mid-trial based on all the interim data and any other information from in- and outside the trial. The main advantage of this method is that there is no need to specify a selection rule in advance in order to control the nominal level. Hence, treatment selection can follow complex decision rules which are based on several measurements simultaneously, including safety, primary and secondary efficacy measurements. The way these measurements influence the selection process may even depend on external events in a complex manner.

The talk will give an introduction to the principles of flexible two stage closed tests, including a discussion on limitations and some of the current developments.

Literature:

- W. Brannath, M. Posch, and P. Bauer (2002) Recursive combination tests. *Journal of the American Statistical Association*, 97:236-244.
- Bauer, P. and Kieser, M. (1999). Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine*, 18:1833-1848.
- Hommel, G. (2001). Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal*, 43(5):581-589.
- Bretz, F., Schmidli, H., Koenig, F., Racine, A., and Maurer, W. (2006). Confirmatory seamless phase II/III clinical trials with hypotheses selection at interim: General concepts. *Biometrical Journal*, 48(4):635-643
- Franz König, Peter Bauer, Werner Brannath (2006). An adaptive hierarchical test procedure for selecting safe and efficient treatments. *Biometrical Journal* 2006, 48:663-78.
- Martin Posch, Franz Koenig, Michael Branson, Werner Brannath, Cornelia Dunger-Baldauf, Peter Bauer (2005). Testing and estimation in flexible group sequential designs with adaptive treatment selection. *Statistics in Medicine*, 30(24):3697-3714.

Gernot A. Fink:

Grundlagen der automatischen Spracherkennung

Ziel der automatischen Spracherkennung ist die möglichst exakte Rekonstruktion einer textuellen Darstellung des Gesprochenen allein auf der Basis des akustischen Signals. Dies gelingt am besten bei der Erkennung einiger weniger Wörter, wie z.B. bei einfachen menügeführten Telephonauskunftssystemen. Immer schwieriger wird

Abstracts

die Aufgabe, je mehr Variabilität durch umfangreiche Wortschätze, wechselnde Sprecher und Aufnahmebedingungen, zunehmend spontanen Sprechstil sowie die Beeinflussung durch Hintergrundgeräusche entsteht. Daher lassen sich die bekannten Diktieranwendungen mit Wortschätzen von bis zu 100.000 Vollformen derzeit nur unter geeigneten Restriktionen realisieren.

Im Laufe der Entwicklung von Spracherkennungssystemen hat sich die Verwendung einiger Verfahren als Standard herausgebildet. Als Basis dient die Parametrisierung kurzer Sprachsignalabschnitte in Form von Merkmalsvektoren, die für Sprachlaute wesentliche spektrale Eigenschaften repräsentieren. Darauf aufbauend erfolgt die Modellierung akustischer Ereignisse mit Hilfe von statistischen Modellen, und auch syntaktisch-semantische Restriktionen werden überwiegend statistisch und selten deklarativ eingebracht.

Der Vortrag gibt eine Einführung in die Modellierungstechniken, die in aktuellen Spracherkennungssystemen zum Einsatz kommen. Dabei werden sowohl die Verfahren zum automatischen Parametertraining als auch zur effizienten Anwendung der Modelle im Erkennungsprozess vorgestellt. Außerdem wird auf die trotz vorwiegend statistischer Methoden notwendigen deklarativen Modellierungsaspekte eingegangen.

Ingileif B. Hallgrímsdóttir:

Detecting gene interactions in case-control association studies

Disease-marker association studies are used to search for correlation between disease and genetic marker alleles by contrasting cases with healthy controls. Several large case-control genome-wide association studies are underway for various common diseases, such as type 1 and type 2 diabetes, hypertension, arthritis, schizophrenia and others. It is known that gene x gene and gene x environment interactions play an important role in the etiology of these diseases, yet the statistical analysis is usually based on testing for main effects, rather than interactions. I will present a new complete description and classification of two-locus (two-gene) disease models based on a geometric approach. For each of the possible 69 classes of models we study the power of the standard statistical tests, both single-locus tests (main effects) and two-locus tests (interactions). I will also discuss work in progress on a novel two-locus test.

Joint work with Debbie Yuster.

Michael Hauptmann:

Statistical methods for assessing latency in epidemiologic studies

Different approaches have been proposed to investigate latency in epidemiologic studies where detailed exposure histories are available. We demonstrate the application of a flexible, yet parsimonious, spline function model to investigate latency patterns for radon progeny exposure and lung cancer in the Colorado Plateau uranium miners cohort. The model extends a previously proposed bilinear model. The excess relative risk (ERR) reached a maximum of 0.6 per 100 working level months, for exposures received 14 years previously. The ERR then declined, and was estimated to approach zero for exposures received 35 years and more in

Abstracts

the past. The point-wise 95% confidence intervals supported ERRs > 0 for the period 9-32 years before the event. The estimated latency curve was homogeneous across categories of attained age, duration of exposure, rate of exposure, and smoking. In conclusion, the proposed spline model is a flexible tool for latency analyses, and extends previously used methods.

Göran Kauermann:

Penalisierte Spline Schätzung

Der Vortrag führt in das Stoffgebiet der Penalisierten Spline Schätzung als Glättungsmethode ein. Dabei werden die Grundideen vorgestellt und mit Datenbeispielen ergänzt. Insbesondere wird der Zusammenhang zwischen Penalisierter Spline Schätzung und linearen gemischten Modellen herausgearbeitet. Dies erlaubt es die Wahl des Glättungsparameters nach einem Maximum Likelihood Ansatz zu bestimmen. Die Idee wird skizziert und mit Datenbeispielen untermauert.

Sonja Kuhnt:

Graphical Models: Outliers and Model Choice

Graphical models combine multivariate statistical models with a representation of conditional independence properties by a mathematical graph. The main research question usually concerns the association structure between the considered random variables, such that model building methods play a major role. Observations which deviate strongly from the main part of the data can severely influence the model selection process (Kuhnt and Becker, 2003). We start with a formal definition of outliers with respect to the conditional Gaussian distribution, which captures the perception of outliers as surprising observations. Conditional Gaussian distributions (Lauritzen and Wermuth, 1983) cover discrete as well as continuous variables. One-step outlier identification procedures are introduced based on the classical maximum likelihood estimator and a new modified maximum likelihood estimator (Kuhnt, 2006). We combine well-known model building strategies (Edwards, 2000, Chapt. 6) with the identification of outliers and show examples as well as simulation results.

Literature:

- Edwards D. Introduction to Graphical Modelling. New York: Springer, 2000.
 - Lauritzen S.L., Wermuth N. Graphical Models for Associations between Variables, Some of Which are Qualitative and Some Quantitative. *Annals of Statistics* 17, 31-57, 1989.
 - Kuhnt S. Robust Graphical Modelling for Mixed Variables. Habilitationsschrift, Department of Statistics, University of Dortmund, Germany, 2006.
 - Kuhnt S, Becker, C. Sensitivity of Graphical Modeling against Contamination. In: Schader M, Gaul W, Vichi M, editors. *Between Data Science and Applied Data Analysis*. Heidelberg: Springer, 279-287, 2003.
-

Abstracts

Uwe Ligges:

Simulationen in R: Effiziente Programme, Wahl von Betriebssystem und CPU optimierter BLAS

Häufig haben sich am Fachbereich Statistik bei Promovierenden Fragen zur Behandlung langer Simulationen in R ergeben. Einige der häufig gestellten Fragen werden in diesem Vortrag aufgegriffen. So werden typische Verbesserungen langsamer Programme auf dem Weg zu effizienten Programmen gezeigt. Außerdem wird auf die Wahl eines geeigneten Betriebssystems und auf die Wahl geeigneter CPU optimierter BLAS Routinen eingegangen. Abschließend wird kurz der Einsatz parallelen Rechnens diskutiert.

Shalabh:

Parametric estimation in measurement error models

The talk will introduce an idea about the linear measurement error models. Different forms like as functional form, structural form, and ultrastructural form will be considered. Some issues related to the estimation of parameters and non-normal distribution of measurement errors will be addressed.

Rückblick auf das 2. Herbstkolloquium, 25./26. November 2005



Retrospection on the 2nd Autumn Symposium, November 25/26, 2005

