Deutsches Diabetes-Zentrum

Leibniz Gemeinschaft

Heinrich Heine Universität Düsseldorf

Programm

1. SAW Workshop

„Multiplizität, Modellvalidierung und Reproduzierbarkeit in hochdimensionalen Microarray-Daten“

am 17.06.2010 im DDZ, Düsseldorf

10:50 bis 11:00	Begrüßung
11:00 bis 11:30	Julia Benditkis und Arnold Janssen: *Zur Kontrolle der FDR*
11:30 bis 12:00	Veronika Gontscharuk: *Weak dependence and error rate control in multiple hypotheses testing*
12:00 bis 13:00	Mittagspause
13:00 bis 13:30	Stefan Lehr: *DDZ Proteomics-Plattform: Methoden zur quantitativen Proteinpatternanalyse*
13:35 bis 14:10	Sandra Landwehr und Daniela Dossing: *Zur Schätzung des Erwartungswertes in einem linearen Modell*
14:10 bis 14:40	Marsel Scheer: *Simultaneous control of false discovery rate and expected number of false rejections*
14:40 bis 15:00	Kaffeepause
15:00 bis 15:30	Vladimir Spokoiny: *Bivariate penalty choice in model selection*
15:30 bis 16:00	Elmar Diederichs: *Dimension Reduction by Saddle Point NonGaussian Component Analysis*
16:00 bis 16:45	Diskussion zukünftiger Aktivitäten

Wo: Vorstandszimmer, Ebene 7 (bis zur Mittagspause)

Neuer Seminarraum, Ebene 0 (nach der Mittagspause)

Abstracts

Julia Benditkis und Arnold Janssen (Mathematisches Institut der Heinrich-Heine-Universität)

In dem Vortrag wird gezeigt, wie aus Schätzern für die lokale FDR ein neues multiples Testverfahren entwickelt werden kann. Insbesondere wird der Zusammenhang zur bekannten asymptotisch optimalen Ablehnkurve diskutiert. Das Verfahren kontrolliert als erweitertes Step-Down-Verfahren die FDR zu jedem Stichprobenumfang.

Veronika Gontscharuk (Deutsches Diabetes Zentrum, Institut für Biometrie und Epidemiologie)

Dependence in multiple hypotheses testing problems is a serious issue regardless of the underlying error rate criteria. We restrict our attention to situations where some kind of weak dependence appears. More precisely, we consider models where the empirical cumulative distribution function of p-values under true null hypotheses converges in the sense of the extended Glivenko-Cantelli theorem. Block-dependence of test statistics and pair wise comparisons are important examples of weak dependence. If the proportion of true hypotheses tends to one we are often faced with the so-called null-problem, that is, the number of rejected hypotheses tends to zero and it is typically not clear whether we have asymptotic control of the underlying error rate criterion.

Daniela Dossing und Sandra Landwehr (Deutsches Diabetes Zentrum, Institut für Biometrie und Epidemiologie)

In dem Vortrag werden gängige Methoden zur Schätzung des Erwartungswertes in einem linearen Modell, wie beispielsweise die Maximum Likelihood Methode und das Vorgehen von James und Stein, mit einem neuen Zugang über Permutationen der Fehlerterme verglichen. Es werden anhand von Simulationen die Güte des jeweiligen Schätzers untersucht sowie dabei auftauchende Probleme diskutiert.

Marsel Scheer (Deutsches Diabetes Zentrum, Institut für Biometrie und Epidemiologie)

Much research has been done concerning control of the false discovery rate (FDR) in multiple hypothesis testing problems. Control of the FDR means to control the expected proportion of rejected true hypotheses with respect to all rejected hypotheses. Methodology and application of FDR controlling procedures are still very active research fields, see for example [Benjamini and Hochberg (1995)], [Benjamini and Yekutieli(2001)], [Finner and Roters(2002)], [Storey et al.(2004)], [Finner and Dickhaus and Roters(2007)], [Finner and Dickhaus and Roters(2009)]. A further interesting characteristic of multiple testing procedures is the expected type 1 error rate (EER). The EER is defined as the expected number of false rejections divided by the number of all hypotheses. In this talk we investigate methods which allow simultaneous control of the FDR and the EER. Thereby we discuss relationships between FDR, EER and rejection curves defining multiple test procedures. It will be shown that the asymptotically optimal rejection curve (AORC) introduced in [Finner and Dickhaus and Roters(2009)] plays a crucial role for powerful procedures controlling FDR and EER simultaneously.

Vladimir Spokoiny (WIAS, Berlin)

Within the penalized model selection approach the selected model is defined by minimization of the penalized empirical risk. A number of oracle risk bounds for such methods are available. However, the choice of the penalty function is critical, while a general strategy in this question is lacking. Hence an alternative method of model selection based on saddle point optimization is presented. The basic observation behind this new method is, that the empirical risk minimizer is a saddle point of a bivariate function built as the difference between empirical risks of two models. An extension of this idea is to define the model selector via a saddle point of a penalized difference. The penalty is bivariate also and has to be selected by the condition that the oracle model can be rejected against a more complex model only with very small probability.

Elmar Diederichs (WIAS, Berlin)

Sparse NonGaussian Component Analysis is an unsupervised, linear iterative and structure adaptive extraction method for semi-parametric high dimensional data analysis based on estimating low-dimensional non-Gaussian components of high-dimensional data. We report the current developmental state of this approach, that is based on semi-definite programming. It can be demonstrated, that the statistical sensitivity of this method to a broad variety of deviation from normality is high while its computational effort is decreased. We present new numerical experiments with high-dimensional data obtained from MD-simulations of proteins, that illustrates the breakthrough in this field.