

**Wiener Biometrische Sektion
der Internationalen Biometrischen Gesellschaft
Region Österreich – Schweiz**
<http://www.meduniwien.ac.at/wbs/>

Einladung zum

BIOMETRISCHEN KOLLOQUIUM

Am Donnerstag, 5. Dezember 2013 um 16:30 Uhr (s.t.)

im Seminarraum (Ebene 3, Raum 88.03.513) des
Zentrums für Medizinische Statistik, Informatik und Intelligente Systeme (CeMSIIS)
der Medizinischen Universität Wien, Spitalgasse 23, 1090 Wien
(Plan siehe <http://www.muw.ac.at/cemsiis/allgemeines/anschrift/>)

Vortragender:

ALEXANDRA POSEKANY

Institute for Statistics and Mathematics, Wirtschaftsuniversität Wien

**IMPROVING ROBUSTNESS AND RELIABILITY OF
BIOINFORMATICAL INFERENCE**

Wir freuen uns auf zahlreichen Besuch.

Gerhard Svolba
Präsident

Franz König
Sekretär

Improving robustness and reliability of bioinformatical inference

Alexandra Posekany

During the past decades high-throughput technologies have been established in biological and medical research. These methods led to an increase in computational approaches dealing with such data. In addition to computational hindrances, bioinformatical data poses several challenges for data analysis: Assumptions made for general approaches in statistical modelling are often not fulfilled, sample sizes are very small compared to the number of variables to estimate, while the observations are overdispersed and the noise is behaving far from Gaussian. These challenges were met in the PhD thesis, [1], in which two hierarchical Bayesian models were designed, including the implementation of MCMC algorithms for inference and performing sanity checks with artificial data.

As can be seen in [1], our focus lay in developing and applying different approaches towards robust inference in a fully Bayesian setting. Starting from Bayesian robustness considerations, which have their origin in frequently occurring criticism of the subjective choice of prior distributions, every part of the Bayesian model, prior, likelihood and loss function, has become the focus of robustness considerations. Our goal was improving bioinformatical inference based on well-founded statistical theory, which we achieved by designing Bayesian models tailored for the above-mentioned specific types of challenges in the field of microarray analysis. On the one hand, we developed an approach towards likelihood robustness which can be linked with the idea of model selection. On the other hand, we designed a scheme for robust mixture models allowing the identification of systematically outlying values as well as conclusions on the behaviour on a gene and array level.

Defining hierarchical Bayesian models has the advantage of providing an additional degree of robustness regarding the choice of priors and model parameters. In order to approach the challenge posed by overdispersion, student's t distributions and mixtures of student's t and normal distributions are considered to gain robustness with respect to the likelihood. Including student's t distributions in the inference only makes sense, if differing between Gaussian and student's t distributions carries valuable information, as an additional computational burden is implied by making the model more complex than a Gaussian mixture model with enough components to approximate any given likelihood. Thus, in addition to noise estimates on microarray, a measure for "non-Gaussianity" was introduced to estimate the "distance" from the normal distribution regarding the 'tail weights' of the distribution which is hard to calculate for mixtures of Gaussians. Based on the quantile-based measure of 'peakedness', we propose a quality control of microarrays by assessing the distribution of noisy genes between the arrays.

As published in [2], the first model led to the revelation that microarrays were far more heavy-tailed than previously assumed and that discrepancies in the results occur if not turning to robust error models. Our mixture model building on these findings provides an approach for simultaneously performing inference and checking the quality of the incoming data, a feature also useful for constructing bioinformatical analysis pipelines.

Literatur

- [1] A. Posekany **Bayesian Foundations for improving robustness and reliability of computational biological inference**; PhD thesis, TU Wien; 2013
- [2] A. Posekany, K. Felsenstein, P. Sykacek **Biological Assessment of robust noise models in microarray data analysis**; *Bioinformatics* 27.6; 2011; pp. 807-814