MATHEMATIK-ARBEITSPAPIERE

A: MATHEMATISCHE FORSCHUNGSPAPIERE

# Asymptotic Normality Of Goodness-of-Fit Statistics For Sparse Poisson Data

Ursula Müller
Gerhard Osius

Preprint No. 51 January 1998

FACHBEREICH MATHEMATIK UND INFORMATIK
UNIVERSITÄT BREMEN

Bibliothekstraße
D-28359 Bremen
Germany

# Asymptotic Normality of Goodness–of–Fit Statistics for Sparse Poisson Data

Ursula Müller and Gerhard Osius
Department of Mathematics and Computer Sciences
University of Bremen

**Abstract**

Goodness–of–fit tests for discrete data and models with parameters to be estimated are usually based on Pearson's $\chi^2$ or the Likelihood Ratio Statistic, which both are included in the family of Power–Divergence Statistics $SD_\lambda$. It is known that $SD_\lambda$ is asymptotically $\chi^2$ distributed for the common sampling schemes, which yield contingency tables being Poisson or conditional Poisson, e.g. product–multinomial, and for an asymptotic approach with the number of cells being fixed. Here a limiting normal distribution of $SD_\lambda$ for Poisson distributed $J \times K$ tables is presented considering an increasing cells approach, i.e. beside the total size the number of covariable groups $J$ increases, whereas the number of categories $K$ and the number of model parameters remains fixed. In contrast to the "fixed cells" asymptotics an increase of all expected values is not required — the expectations of the cells may be large but need not be, which allows an application of the deduced tests to sparse data. The peculiarity of the here considered approach is that the underlying class of models to test does not specify the marginal distributions of the (covariable) groups and categories — only the associations, i. e. the odds ratios, are modelled with a finite number of parameters. One thus has to deal with an asymptotically infinite number of nuisance parameters.

Key words: goodness–of–fit, Power Divergence Statistics, contingency tables, Poisson data, sparse data, odds ratios.