UNIVERSITÀ DEGLI STUDI DI PADOVA

Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze

Relazione finale

Akaike's Information Criterion in Generalized Estimating Equations

Relatore: Prof. Alessandra Salvan

Dipartimento di Scienze Statistiche

Laureando: Francesco Ignazio Re

Matricola: 1149556

Contents

Introduction

Overview

Statistical analysis is a process that can be broken into different steps. From data collection, through data analysis, up to the yielding of consistent results, statisticians are continuously asked to come down to compromises in the attempt of tackling the underlying trends of their object of study. Among these steps, the greatest controversy is probably bound to model selection: a bitter truth known to every statistician is that there is no such a thing as a best model. With that said, it is still reasonable to search - if not for the best - for a better model and, in this respect, several indexes were built for comparing different models with each other. A particularly powerful index is the Akaike's information criterion; it is based on the likelihood and asymptotic properties of the maximum likelihood estimator and allows model comparison in terms of predictability and parsimony. Despite being a powerful tool, its strict dependence on the likelihood implies the model distribution to be fully known: a requirement that cannot always be fulfilled. In this context, this work sets its aim at assessing methods to widen the AIC usage to those models for which there is no likelihood defined. We will specifically focus our attention on the Akaike's information criterion for models estimated through the generalized estimating equation (GEE) approach, very useful for working with correlated data, but based on the quasi-likelihood method, and hence, unconstrained by the assumption of a distribution.

Summary

Chapter 1

Models based on Maximum Likelihood Estimation

1.1 Likelihood

1.2 Linear Models

1.2.1

1.2.2 Title of subsection

1.2.3 Title of subsection

1.3 Generalized Linear Models

TABLE 1.1: ML fit of the Gamma regression model with log-link and Wald 0.95 confidence intervals for the parameters.

	Estimate	Estimated Standard Error	0.95 Confidence Interval
β_1	0.361	0.250	(-0.128, 0.851)
β_2	1.507	0.170	(1.174, 1.839)
β_3	1.859	0.165	(1.535, 2.183)
ϕ	0.223	0.079	(0.069, 0.377)

Chapter 2

Quasi-Likelihood Models

2.1 Quasi-likelihood inference

2.2 Quasi-likelihood function

2.2.1 Generalized Estimating Equations

2.2.2 Title of subsection

2.2.3 Title of subsection

2.3 Title of section

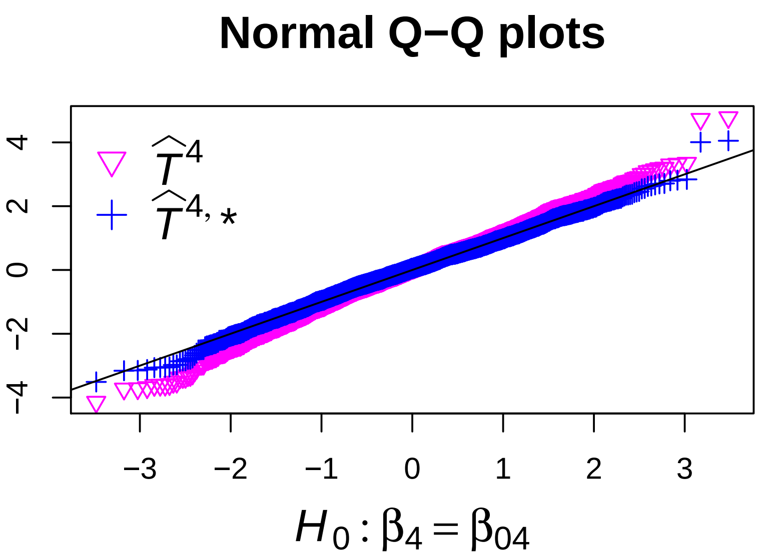

FIGURE 2.1: Normal Q-Q plots based on 2000 values of $\widehat{T}^4$ and $\widehat{T}^{4,*}$ computed under the null hypothesis $H_0: \beta_4 = \beta_{04}$ in the *clotting* example.

Chapter 3

Title of chapter

3.1 Title of section

Azzalini (2001)

3.2 Title of section

Bartlett (1953)

3.2.1 Title of subsection

Kosmidis (2016)

3.2.2 Title of subsection

Stafford (1992)

3.2.3 Title of subsection

DiCiccio and Stern (1993)

3.3 Title of section

Appendix

Bibliography

Azzalini, A. (2001) *Inferenza Statistica. Una Presentazione Basata sul Concetto di Verosimiglianza.* Milano: Springer-Verlag Italia.

Bartlett, M. S. (1953) Approximate confidence intervals. II. More than one unknown parameter. *Biometrika* **40**, 306–317.

DiCiccio, T. J. and Stern, S. E. (1993) An adjustment to profile likelihood based on observed information. Technical report, Department of Statistics, Stanford University.

Kosmidis, I. (2016) `brglm2`*: Estimation and inference for generalized linear models using explicit and implicit methods for bias reduction.* `https://github.com/ikosmidis/brglm2`.

Stafford, J. E. (1992) *Symbolic Computation and the Comparison of Traditional and Robust Test Statistics* (unpublished doctoral dissertation). University of Toronto, Canada.