2005 ABSTRACTS

No. 2005-1

THE RELATIVE EFFICIENCY OF PREDICTION INTERVALS.

PAUL KABAILA

Department of Statistical Science, La Trobe University,
Victoria 3086, Australia

Abstract

Consider two prediction intervals with the same minimum coverage probabilities. Suppose that we wish to compare the efficiencies of these prediction intervals. At first sight, it seems reasonable to measure the relative efficiencies of these prediction intervals by the ratio of their expected lengths. We argue that this is a flawed measure, particularly when the sample size is large. In its place, we propose to measure this relative efficiency by the ratio of sample sizes required for the expected lengths of the prediction intervals to be equal. We illustrate the application of this measure in the following situation. Suppose that $X_1, \ldots, X_n, X_{n+1}$ are independent and identically distributed. The data is $X_1, \ldots, X_n$ and our aim is to find a prediction interval for X_{n+1}. Suppose that we use a nonparametric equi-tailed prediction interval based on sample quantiles. The question we ask is what is the efficiency of this prediction interval by comparison with the standard equi-tailed parametric prediction interval when X_1 has an $N(\mu, \sigma^2)$ distribution.

No. 2005-2

A Latent Variable Model for Estimating Disease Transmission Rate from Data on Household Outbreaks.

Ning Li, Guoqi Qian and Richard Huggins

Department of Statistical Science, La Trobe University,
Victoria 3086, Australia

Abstract

A Bayesian latent variable model is proposed for studying household epidemics of infectious diseases in this paper. This model is more general and flexible than the commonly used chain binomial epidemic model. In particular, the model allows for the heterogeneity of the infection transmission rates in related to the sizes and generations of the infectives. Moreover, the model assumes the availability of only the household outbreak sizes, which is more reasonable than assuming the availability of the hardly observed infection chains. The Tanner-Wong's IP algorithm is employed for effective simulations and inferences of this model. Finally, this model was applied to analyzing a real data set on F-4 Asian influenza.

Key Words: Disease transmission rate, infection chain, outbreak size, data augmentation, Tanner-Wong algorithm, Bayesian statistics.

No. 2005-3

Valid Confidence Intervals in Regression after Variable Selection II.

Paul Kabaila and Jarrod Tuck

Department of Statistical Science, La Trobe University,
Victoria 3086, Australia

Abstract

We consider a linear regression model with regression parameters (β_1, β_2) and error variance parameter σ^2. The parameter of interest is β_1. We suppose that, as a first step, we test the null hypothesis $\beta_2 = 0$ against the alternative hypothesis that $\beta_2 \neq 0$. It is common practice to use a data-based model selection followed by inference, using the same data, based on the (false) assumption that the selected model had been given to us *a priori*. Consider a confidence interval for β_1 with nominal coverage $1 - \alpha$, constructed from the same data as for this hypothesis test, based on this (false) assumption. We call this the naive $1 - \alpha$ confidence interval for β_1. This confidence interval has a minimum coverage probability that is less than $1 - \alpha$ i.e. it is not a valid confidence interval. We may find a valid confidence interval for β_1 as follows. We adjust the length of the naive confidence interval when the null hypothesis that $\beta_2 = 0$ is accepted, so that the adjusted interval has minimum coverage probability $1 - \alpha$. We call this the 'singly-adjusted' confidence interval for β_1. Suppose that we have some prior belief (though not certainty) that $|\beta_2|/\sigma$ is small. Kabaila (1998) shows that the 'singly-adjusted' confidence interval will, in some cases, be preferable to the standard confidence interval for β_1 when the sample size is small. He also shows that the 'singly-adjusted' confidence interval cannot be better than this standard confidence interval in large samples. Is this disappointing feature of the 'singly-adjusted' interval due to some fundamental limitation or is it merely due to the particular form of adjustment used? We show that this disappointing feature is due to the the particular form of adjustment used. We do this by adjusting the length of the naive confidence interval *both* when the null hypothesis that $\beta_2 = 0$ is accepted and when it is rejected, so that it has minimum coverage probability $1 - \alpha$. We call this the 'doubly-adjusted' confidence interval. We show that in large samples this confidence interval will, in some cases, be preferable to the standard confidence interval for β_1.

No. 2005-4

On Time Series Model Selection Involving Many Candidate ARMA Models

Guoqi Qian and Xindong Zhao
Department of Statistical Science, La Trobe University,
Melbourne, VIC 3086, Australia

Abstract

We study how to perform model selection for time series data where millions of candidate ARMA models may be eligible for selection. We propose a feasible non-Bayesian computing method based on the Gibbs sampler. By this method model selection is performed through a random sample generation algorithm, and given a model of fixed dimension the parameter estimation is done through the maximum likelihood method. Our method takes into account several computing difficulties encountered in estimating ARMA models. The method is found to have probability of 1 in the limit in selecting the best candidate model under some regularity conditions. We then propose several empirical rules to implement our computing method for applications. Finally a simulation study and an example on modelling China's Consumer Price Index (CPI) data are presented for purpose of illustration and verification.

Keywords: Autoregressive-moving average (ARMA) models, Gibbs sampler and time series model selection.

No. 2005-5

Confidence intervals utilizing prior information in the Behrens-Fisher problem

Paul Kabaila[a,*], Jarrod Tuck[a]

[a] *Department of Statistical Science, La Trobe University, Victoria 3086, Australia*

Abstract

Suppose that we have two independent random samples of size $f + 1$. One of these samples is from an $N(\mu_1, \sigma_1^2)$ distribution and the other is from an $N(\mu_2, \sigma_2^2)$ distribution, where $\sigma_1^2/\sigma_2^2 \in (0, \infty)$. Suppose that we have prior information (which is not a certainty) that σ_1^2/σ_2^2 differs negligibly from 1. We consider a broad class $\mathcal{C}$ of confidence intervals for $\mu_1 - \mu_2$ with minimum coverage probability $1-\alpha$. The adjusted Welch $1-\alpha$ confidence interval, described by Kabaila (2005), is a member of this class. We use this interval as a standard against which other members of $\mathcal{C}$ will be judged. A confidence interval J, belonging to $\mathcal{C}$, makes profitable use of this prior information if the ratio (expected length of J) / (expected length of adjusted Welch $1-\alpha$ confidence interval) is (a) significantly less than 1 for $\sigma_1^2/\sigma_2^2 = 1$ and (b) not too much larger than 1 for all other values of σ_1^2/σ_2^2. Does there exist a confidence interval J, belonging to $\mathcal{C}$, that satisfies these conditions? We focus on the question of whether condition (a) can be satisfied. To help answer this question, we extend the general compromise decision-theoretic methodology of Kempthorne (1983, 1988). It is by no means obvious that the computations required for the application of this general methodology are feasible. A very careful analysis shows, however, how to make these computations feasible. Our conclusion from these computations is that, for $1-\alpha = 0.95$, the answer to this question is no for all $f \in \{2, \ldots, 100\}$.

Keywords: Behrens-Fisher; Confidence Interval; Prior information; Decision Theory.

MSC: 62F25

Short title: Behrens-Fisher problem

* Corresponding author. Tel.: +61 3 9479 2594, fax: +61 3 9479 2466.
E-mail address: P.Kabaila@latrobe.edu.au (Paul Kabaila).