2004 ABSTRACTS

No. 2004-1

Calibrating significant p-values.

Stephan Morgenthaler and Robert G. Staudte

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

Abstract

The p-value is a frequently used measure of evidence against a null hypothesis even though its interpretation has been a matter of long-standing controversy. One reason for such controversy is that, when considered as a random variable, it has very different distributions under the null and alternative hypotheses. The purpose here is to provide a calibration scale for the p-value and other measures of evidence against the null hypothesis when, in fact, the alternative hypothesis is true. The authors also consider the related problem of combining p-values from different experiments.

No. 2004-2

Conditionally Optimal Weights of Evidence.

Stephan Morgenthaler, Swiss Federal Institute of Technology, EPFL-DMA, 1015 Lausanne, Stephan.Morgenthaler@epfl.ch

and

Robert G. Staudte, Statistics Department, La Trobe University, Melbourne, 3086, Australia, r.staudte@latrobe.edu.au

Abstract

A weight of evidence is a calibrated statistic whose values in $[0, 1]$ indicate the degree of agreement between the data and either of two hypothesis, one being treated as the null (H_0) and the other as the alternative (H_1). A value of zero means perfect agreement with the null, whereas a value of one means perfect agreement with the alternative. The optimality we consider is minimal mean squared error (MSE) under the alternative while keeping the MSE under the null below a fixed bound. This paper studies such statistics from a conditional point of view, in particular for location and scale models.

No. 2004-3

A seasonal decomposition of the estimated size of a penguin population at Phillip Island, Australia

Hsin-Chou Yang, Lynda Chambers, Richard M. Huggins
Institute of Biomedical Sciences, Academia Sinica,
Taipei, NanGang 105, Taiwan
Bureau of Meteorology Research Centre, GPO 1289k, Melbourne
Victoria 3001 Australia
Department of Statistical Science, La Trobe University, Bundoora,
Victoria 3086, Australia

Abstract

Mark recapture data from July 1990 to June 1995 were first used to estimate the population size of Little Penguins at Summerland beach, Phillip Island, Australia using nonparametric methods. Nonparametric estimates revealed seasonal variation. Seasonal time series models and autoregressive integrated moving average processes were then applied to the estimated population size to help explain the seasonal variation and long-term trends.

No. 2004-4

An examination of the effect of heterogeneity on the estimation of population size using capture-recapture data.

Wen-Han Hwang
Department of Statistics, Feng Chia University, Taichung, Taiwan, and
Richard Huggins
Department of Statistical Science, La Trobe University,
Victoria 3086, Australia

Abstract

Part of the folklore of capture-recapture experiments is that ignoring heterogeneity of capture probabilities results in a negative bias. This has been based on experience and simulation studies but is often interpreted as being due to individuals with lower capture probabilities. Here estimating equation arguments are used to show that the effect on Horvitz-Thompson type estimators of ignoring heterogeneity in capture-recapture experiments is to introduce a negative bias. The negative bias holds even if the small number of heterogeneous individuals have capture probabilities larger than the homogeneous majority and this is confirmed by simulations. The arguments are extended to continuous time experiments and to an influence function constructed to determine the effect of a small number of individuals with heterogeneous capture probabilities in an otherwise homogeneous population and the influence function is also shown to be negative.

No. 2004-5

ON THE COVERAGE PROBABILITY OF CONFIDENCE INTERVALS IN REGRESSION AFTER VARIABLE SELECTION.

PAUL KABAILA

Department of Statistics
La Trobe University

Abstract

We consider a linear regression model with regression parameter vector β. The parameter of interest is $\theta = a^T\beta$ where a is specified. We suppose that, as a first step, a data-based variable selection (e.g. minimum AIC) is used to select a model. It is common statistical practice to then carry out inference about θ, using the same data, based on the (false) assumption that the selected model had been given to us *a priori*. We consider a confidence interval for θ with nominal coverage $1 - \alpha$ constructed on this (false) assumption. We call this the naive $1 - \alpha$ confidence interval. The minimum coverage probability of this confidence interval can be calculated for simple variable selection procedures involving only a single variable (Kabaila (1998)). However, the kinds of variable selection procedures used in practice are typically much more complicated. For the real-life data presented in this paper, there are 20 variables each of which is to be either included or not, leading to 2^{20} different models. The coverage probability at any given value of the parameters will, of course, provide an upper bound on the minimum coverage probability of the naive confidence interval. In this paper we derive a new Monte Carlo simulation estimator of the coverage probability that uses conditioning for variance reduction. For this real-life data, the gain in efficiency of this Monte Carlo simulation due to conditioning ranged from 2 to 6. We also present a simple 1-dimensional search strategy for parameter values at which the coverage probability is relatively small. For this real-life data, this search leads to parameter values for which the coverage probability of the naive 0.95 confidence interval is 0.79 for variable selection using AIC and 0.70 for variable selection using BIC, showing that these confidence intervals are completely inadequate.

Key words: AIC; BIC; Mallows' criterion; adjusted R^2-statistic; "best subset" regression; forward selection; backward elimination; coverage probability.

No. 2004-6

IMPROVED SIMULATED ANNEALING FOR THE DECOMPOSITION OF NORMAL MIXTURES

Mark J. Fielding

Department of Statistics
La Trobe University

Abstract

We introduce an improved simulated annealing algorithm for the decomposition of mixtures of normal distributions, and gain superior results to a more complicated algorithm as reported in the literature. Mixtures of up to 4 distributions are considered, allowing unequal variances, where our goal is to obtain estimates close to those obtained if it were known from which distribution each observation came.

No. 2004-7

Global Optimization and Buehler Confidence Limits

PAUL KABAILA and MICHAEL JERIE
Department of Statistical Science, La Trobe University, Victoria 3086, Australia
(e–mail: *P.Kabaila@latrobe.edu.au*, *M.Jerie@latrobe.edu.au*).

Abstract

Buehler confidence limits have a wide range of statistical applications including reliability and biostatistics. The computation of these confidence limits requires the solution of a global optimization problem that has some very special characteristics. We use these special characteristics to design an elegant new algorithm for its solution. This algorithm is illustrated with numerical results for a Buehler upper confidence limit for the probability of failure of a two-component parallel system based on binomial failure data for each component.

Key words: Global optimization, Buehler confidence limit.

No. 2004-8

Strong Limit Theorems on Model Selection in Generalized Linear Regression with Binomial Responses

Guoqi Qian
Department of Statistical Science, La Trobe University
Melbourne VIC 3086, Australia; email: g.qian@latrobe.edu.au
and Yuehua Wu
Department of Mathematics and Statistics, York University
Toronto, ON M3J 1P3, Canada

Abstract

We prove a law of iterated logarithm for the maximum likelihood estimator of the parameters in a generalized linear regression model with binomial response. This result is then used to derive an asymptotic bound for the difference between the maximum log-likelihood function and the true log-likelihood. It is further used to establish the strong consistency of some penalized likelihood based model selection criteria. We have shown that, under some general conditions, a model selection criterion will select the simplest correct model almost surely if the penalty term is an increasing function of the model dimension and has an order between $O(\log\log n)$ and $O(n)$. Cases involving the commonly used link functions are discussed for illustration of the results.

Keywords: law of the iterated logarithm, generalized linear models, maximum likelihood estimator, model selection, strong consistency.

No. 2004-9

Some Strong Convergence Results for Log-linear Regression Model Selection

Guoqi Qian and Qing Xia
Department of Statistical Science
La Trobe University
Melbourne, VIC 3086, Australia

Abstract

In this paper we first derive a law of iterated logarithm for the maximum likelihood estimator of the parameters in a log-linear regression model. We then use this result to establish the strong consistency of a class of model selection criteria in log-linear regression model selection. We show that under some general conditions, a model selection criterion, which consists of a minus maximum log-likelihood and a penalty term, will select the simplest correct model almost surely if the penalty term increases with model dimension and has an order in between $O(\log\log n)$ and $O(n)$.

Keywords: Law of iterated logarithm; Log-linear regression; Maximum likelihood estimator; Model selection; Strong consistency.

No. 2004-10

A general framework for comparing locations of heterogeneous populations, with interval estimates of weighted effect sizes in the one-way heteroscedastic ANOVA.

E. Kulinskaya and Robert G. Staudte
Department of Statistical Science
La Trobe University
Melbourne, VIC 3086, Australia

Abstract

A general framework for comparing population locations in the presence of heteroscedasticity is provided. A single number called the weighted effect size summarizes the differences in population locations after weighting each according to the difficulty of estimating population location, whether the difficulty is due to unknown population variances, unequal sample sizes or the presence of outliers. For an ANOVA weighted for unequal variances, we find interval estimates for the weighted effect size. In addition, the weighted effect size is shown to be a monotone function of a suitably defined weighted coefficient of determination, which means that interval estimates of the former are readily transformed to the latter. Extensive simulations demonstrate the accuracy of the nominal 95% coverage of these intervals for a wide range of parameters.

Keywords: coefficient of determination, effect size, noncentrality parameter, two-sample problem, weighted intraclass correlation coefficient.

No. 2004-11

Calibrating significant p-values.

Stephan Morgenthaler and Robert G. Staudte

Department of Statistical Science

La Trobe University

Melbourne, VIC 3086, Australia

Abstract

The p-value is a frequently used measure of evidence against a null hypothesis even though its interpretation has been a matter of long-standing controversy. One reason for such controversy is that, when considered as a random variable, it has very different distributions under the null and alternative hypotheses. The purpose here is to provide a calibration scale for the p-value evidence against the null hypothesis when, in fact, the alternative hypothesis is true; this scale is already the basis for the Liptak-Stouffer method in meta-analysis, and we explain how to extend their method of combining p-values when a random effect is present. A Bayesian calibration scale due to Selke, Bayarri and Berger (2001) which provides a lower bound on the posterior probability of the null hypothesis is shown to uniformly underestimate the p-value evidence on the scale advocated here.

Keywords: Bayes factor, conditional frequentist test, meta-analysis, standardized effect size, variance stabilizing distribution, Z-score.

No. 2004-12

Combining and interpreting the evidence in experiments.

Robert G. Staudte, Stephan Morgenthaler and Elena Kulinskaya

Department of Statistical Science

La Trobe University

Melbourne, VIC 3086, Australia

Abstract

Morgenthaler and Staudte (2004) proposed defining the experimental evidence against a null hypothesis of zero effect as a variance stabilizing transformation of the p-value. For one-sided alternatives this often turns out to be the observed standardized effect. They further explained how to combine and interpret such evidence from different one-sample experiments, assuming fixed and random effect models. Here we consider a more general situation where the fixed effects, as well as the population standard deviations, from different experiments can differ. We propose a new definition of evidence for two-sided alternatives in terms of evidence for one-sided alternatives. The evidence against the hypothesis of zero effect is found for a random standardized effects model. We extend these results to the two-sample case and illustrate them with data taken from the medical review literature.

Keywords: meta-analysis, p-values, random effects, standardized effect, statistical evidence, variance stabilizing transformation.

No. 2004-13

Combining the evidence in matched pair or case-control studies.

Robert G. Staudte, Stephan Morgenthaler and Elena Kulinskaya

Department of Statistical Science

La Trobe University

Melbourne, VIC 3086, Australia

Abstract

For matched pair experiments with dichotomous responses in which a binomial model is appropriate, the evidence against the null hypothesis of no effect is defined on a natural calibration scale which allows for interpretation under alternative hypotheses. A simple method for combining such evidence in similar studies is proposed and examined. The methodology is extended to case-control studies in which the evidence against the null hypothesis of equal risks is defined; such evidence is the same whether risk difference, relative risk, or odds ratios are of interest. The same method for combining evidence in the matched pairs case can be employed in these two-sample studies. Applications are given to reviews in the meta-analytic literature.

Keywords: meta-analysis, p-values, relative risk, risk difference, odds ratio, standardized effect, statistical evidence, variance stabilizing transformation.