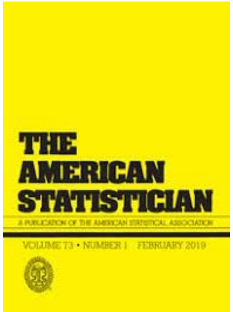

The American Statistician, ISSN 0003-1305
Volume 77, number 1 (february 2023)

Bayesian Testing of Linear Versus Nonlinear Effects Using Gaussian Process Priors P. 1-11

Joris Mulder

Abstract

A Bayes factor is proposed for testing whether the effect of a key predictor variable on a dependent variable is linear or nonlinear, possibly while controlling for certain covariates. The test can be used (i) in substantive research for assessing the nature of the relationship between certain variables based on scientific expectations, and (ii) for statistical model building to infer whether a (transformed) variable should be added as a linear or nonlinear predictor in a regression model. Under the nonlinear model, a Gaussian process prior is employed using a parameterization similar to Zellner's *g* prior resulting in a scale-invariant test. Unlike existing *p*-values, the proposed Bayes factor can be used for quantifying the relative evidence in the data in favor of linearity. Furthermore the Bayes factor does not overestimate the evidence against the linear null model resulting in more parsimonious models. An extension is proposed for Bayesian one-sided testing of whether a nonlinear effect is consistently positive, consistently negative, or neither. Applications are provided from various fields including social network research and education.

A Study on the Power Parameter in Power Prior Bayesian Analysis P. 12-19

Zifei Han, Keying Ye & Min Wang

Abstract

The *power prior* and its variations have been proven to be a useful class of informative priors in Bayesian inference due to their flexibility in incorporating the historical information by raising the likelihood of the historical data to a fractional power δ. The derivation of the marginal likelihood based on the original power prior, and its variation, the normalized power prior, introduces a scaling factor $C(\delta)$ in the form of a prior predictive distribution with powered likelihood. In this article, we show that the scaling factor might be infinite for some positive δ with conventionally used initial priors, which would change the admissible set of the power parameter. This result seems to have been almost completely ignored in the literature. We then illustrate that such a phenomenon may jeopardize the posterior inference under the power priors when the initial prior of the model parameters is improper. The main findings of this article suggest that special attention should be paid when the suggested level of borrowing is close to 0, while the actual optimum might be below the suggested value. We use a normal linear model as an example for illustrative purposes.

Coherent Tests for Interval Null Hypotheses P. 20-28

Spencer Hansen & Ken Rice

Abstract

In a celebrated 1996 article, Schervish showed that, for testing interval null hypotheses, tests typically viewed as optimal can be logically incoherent. Specifically, one may fail to reject a specific interval null, but nevertheless—testing at the same level with the same data—reject a larger null, in which the original one is nested. This result has been used to argue against the widespread practice of viewing *p*-values as measures of evidence. In the current work we approach tests of interval nulls using simple Bayesian decision theory, and establish straightforward conditions that ensure coherence in Schervish's sense. From these, we go on to establish novel frequentist criteria—different to Type I error rate—that, when controlled at fixed levels, give tests that are coherent in Schervish's sense. The results suggest that exploring frequentist properties beyond the familiar Neyman–Pearson framework may ameliorate some of statistical testing's well-known problems.

Abstract

We survey a range of popular generalized count distributions, investigating which (if any) can be arbitrarily underdispersed, that is, its variance can be arbitrarily small compared to its mean. A philosophical implication is that some models failing this simple criterion should not be considered as "statistical models" according to McCullagh's extendibility criterion. Four practical implications are also discussed: (i) functional independence of parameters, (ii) double generalized linear models, (iii) simulation of underdispersed counts, and (iv) severely underdispersed count regression. We suggest that all future generalizations of the Poisson distribution be tested against this key property.

Abstract

In the paired data setting, the sign test is often described in statistical textbooks as a test for comparing differences between the medians of two marginal distributions. There is an implicit assumption that the median of the differences is equivalent to the difference of the medians when employing the sign test in this fashion. We demonstrate however that given asymmetry in the bivariate distribution of the paired data, there are often scenarios where the median of the differences is not equal to the difference of the medians. Further, we show that these scenarios will lead to a false interpretation of the sign test for its intended use in the paired data setting. We illustrate the false-interpretation concept via theory, a simulation study, and through a real-world example based on breast cancer RNA sequencing data obtained from the Cancer Genome Atlas (TCGA).

Abstract

Shared frailty models, that is, hazard regression models for censored data including random effects acting multiplicatively on the hazard, are commonly used to analyze time-to-event data possessing a hierarchical structure. When the random effects are assumed to be normally distributed, the cluster-specific marginal likelihood has no closed-form expression. A powerful method for approximating such integrals is the adaptive Gauss-Hermite quadrature (AGHQ). However, this method requires the estimation of the mode of the integrand in the expression defining the cluster-specific marginal likelihood: it is generally obtained through a nested optimization at the cluster level for each evaluation of the likelihood function. In this work, we show that in the case of a parametric shared frailty model including a normal random intercept, the cluster-specific modes can be determined analytically by using the principal branch of the Lambert function, $W0$. Besides removing the need for the nested optimization procedure, it provides closed-form formulas for the gradient and Hessian of the approximated likelihood making its maximization by Newton-type algorithms convenient and efficient. The Lambert-based AGHQ (LAGHQ) might be applied to other problems involving similar integrals, such as the normally distributed random intercept Poisson model and the computation of probabilities from a Poisson lognormal distribution.

Analytical Problem Solving Based on Causal, Correlational and Deductive Models P. 51-61

Jeroen de Mast, Stefan H. Steiner, Wim P. M. Nuijten & Daniel Kapitan

Abstract

Many approaches for solving problems in business and industry are based on analytics and statistical modeling. Analytical problem solving is driven by the modeling of relationships between dependent (Y) and independent (X) variables, and we discuss three frameworks for modeling such relationships: cause-and-effect modeling, popular in applied statistics and beyond, correlational predictive modeling, popular in machine learning, and deductive (first-principles) modeling, popular in business analytics and operations research. We aim to explain the differences between these types of models, and flesh out the implications of these differences for study design, for discovering potential X/Y relationships, and for the types of solution patterns that each type of modeling could support. We use our account to clarify the popular descriptive-diagnostic-predictive-prescriptive analytics framework, but extend it to offer a more complete model of the process of analytical problem solving, reflecting the essential differences between causal, correlational, and deductive models.

A Statistical Basis for Reporting Strength of Evidence as Pool Reduction P. 62-71

Dan J. Spitzner

Abstract

This article establishes a statistical basis for an evidence-reporting strategy that interprets strength of evidence in terms of a reduction in the size of a pool of relevant conceptual objects. The strategy is motivated by debates in forensic science, wherein the pool would consist of sources of forensic material. An advantage of using the pool-reduction strategy is that it highlights uncertainty that cannot be resolved by empirical considerations. It is shown mathematically to reflect a nonstandard formulation of a Bayes factor, and to extend for use in problems of general quantitative inference. A number of conventions are proposed for full effectiveness of the strategy's implementation in practice.

Assignment-Control Plots: A Visual Companion for Causal Inference Study Design P. 72-84

Rachael C. Aikens & Michael Baiocchi

Abstract

An important step for any causal inference study design is understanding the distribution of the subjects in terms of measured baseline covariates. However, not all baseline variation is equally important. We propose a set of visualizations that reduce the space of measured covariates into two components of baseline variation important to the design of an observational causal inference study: a propensity score summarizing baseline variation associated with treatment assignment and a prognostic score summarizing baseline variation associated with the untreated potential outcome. These *assignment-control plots* and variations thereof visualize study design tradeoffs and illustrate core methodological concepts in causal inference. As a practical demonstration, we apply assignment-control plots to a hypothetical study of cardiothoracic surgery. To demonstrate how these plots can be used to illustrate nuanced concepts, we use them to visualize unmeasured confounding and to consider the relationship between propensity scores and instrumental variables. While the family of visualization tools for studies of causality is relatively sparse, simple visual tools can be an asset to education, application, and methods development.

Black Box Variational Bayesian Model Averaging P. 85-96

Vojtech Kejzlar, Shrijita Bhattacharya, Mookyong Son & Tapabrata Maiti

Abstract

For many decades now, Bayesian Model Averaging (BMA) has been a popular framework to systematically account for model uncertainty that arises in situations when multiple competing models are available to describe the same or similar physical process. The implementation of this framework, however, comes with a multitude of practical challenges including posterior approximation via Markov chain Monte Carlo and numerical integration. We present a Variational Bayesian Inference approach to BMA as a viable alternative to the standard solutions which avoids many of the aforementioned pitfalls. The proposed method is "black box" in the sense that it can be readily applied to many models with little to no model-specific derivation. We illustrate the utility of our variational approach on a suite of examples and discuss all the necessary implementation details. Fully documented Python code with all the examples is provided as well.

"Two Truths and a Lie" as a Class-Participation Activity P. 97-101

Andrew Gelman

Abstract

We adapt the social game "Two truths and a lie" to a classroom setting to give an activity that introduces principles of statistical measurement, uncertainty, prediction, and calibration, while giving students an opportunity to meet each other. We discuss how this activity can be used in a range of different statistics courses.

The Probability Mass Function of the Kaplan–Meier Product–Limit Estimator P. 102-110

Yuxin Qin, Heather Sasinowska & Lawrence Leemis

Abstract

Kaplan and Meier's 1958 article developed a nonparametric estimator of the survivor function from a right-censored dataset. Determining the size of the support of the estimator as a function of the sample size provides a challenging exercise for students in an advanced course in mathematical statistics. We devise two algorithms for calculating the support size and calculate the associated probability mass function for small sample sizes and particular probability distributions for the failure and censoring times.