ISSN 1726-3328

JPSS

Journal of Probability and Statistical Science

A Comprehensive Journal of Probability and Statistics
for Theorists, Methodologists, Practitioners, Teachers, and Others

Volume 3 Number 1

February 2005

JPSS Editorial Board

JPSS Editorial Board

(continued)

JPSS

ISSN 1726-3328

Journal of Probability and Statistical Science

Published by: Susan Rivers' Cultural Institute, Hsinchu, Taiwan, ROC

(Cosponsored by: The International Chinese Association of Quantitative Management, Taiwan, ROC)

Aims and Scope The ***Journal of Probability and Statistical Science*** (***JPSS***, ISSN 1726-3328) is a modified version of the ***Journal of Propagations in Probability and Statistics*** (***JPPS***, ISSN 1607-7083). ***JPSS***, like its predecessor ***JPPS***, is a multipurpose and comprehensive journal of probability and statistics that publishes papers of interest to a broad audience of theorists, methodologists, practitioners, teachers, and any other users of probability and/or statistics. The scope of ***JPSS*** is intended to be quite broad, including all the major areas of probability and statistics. Research papers involving probability and/or statistics, either theoretical or applied in nature, will be seriously considered for publication. Additionally, papers involving innovative techniques or methods in teaching probability and/or statistics will also be considered. Papers submitted for publication consideration will be peer reviewed. Initially, we will publish semiannually, one issue each in February and August. Hopefully, as time accrues, we will be able to publish quarterly. It is the goal of ***JPSS*** to publish a wide range of works involving probability and/or statistics (theoretical and/or applied in nature) and provide widespread availability of such to a broad audience of people interested in probability and/or statistics.

Submission and Review Policies

1. Three hard copies of the manuscript written in English should be mailed to the **Editor-in-Chief, an Editor, a Coordinating Editor,** or to the **Managing Editor** at the address provided in the ***JPSS*** Editorial Board. Alternatively, submission of manuscript by email attachment is also acceptable.
2. A manuscript is accepted only with the understanding that the text has not appeared in publication, and that it is not being simultaneously reviewed by any other journal.
3. Submitted manuscripts are refereed by a double-blind process, meaning that the reviewers will not know the names of the authors and vice versa.
4. If an article is accepted for publication, the author(s) will be required to provide an electronic copy of the paper, in **Micro-soft Word** or **PCTEX** format, on a floppy disk or through an email attachment. The authors will also be requested to transfer their copyright on certain conditions to the publisher.

Publisher Harold C. H. Chen, Head, Susan Rivers' Cultural Institute. Address: 26, Lane 2, Chien Mei Road, Hsinchu, Taiwan, ROC. Phone: (03)5716594, Fax: (03)5712524.

JPSS

Journal of Probability and Statistical Science

Volume 3 Number 1 February 2005

Table of Contents

Theory and Methods

Teaching and Applications

Appendix

Journal of Probability and Statistical Science
3(1), 1-16, Feb. 2005

Interval Estimation of the Common Lognormal Mean of Several Populations

Ayman Baklizi and Mohammed AL-Haj Ebrahem
Yarmouk University

ABSTRACT We consider interval estimation of the common mean of m lognormal populations having possibly unequal variances. We studied and developed several types of large sample and bootstrap intervals. Their performances are investigated using simulation techniques and compared in terms of attainment of the nominal confidence level, symmetry of lower and upper error rates, and expected length. Recommendations concerning their use are given.

Keywords Bootstrap; Common mean; Interval estimation; Lognormal distribution.

1. Introduction

The lognormal distribution has many special characteristics and features together with its relation with the normal distribution that allowed it to be used as a model in various real life applications. In particular it is used in analyzing biological data (Koch [16]), and for analyzing data in workplace exposure to contaminants (Lyles *et al*. [19]). It is also of importance in modeling lifetimes of products and individuals (Lawless [18]). Various other motivations and applications of the lognormal distribution can be found in Johnson *et al*. [14].

Let the random variable Y be normally distributed with mean μ and variance σ^2. Let also $X = e^Y$, then X is said to have a lognormal distribution. The probability density function of X can be shown to be

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), \quad 0 < x < \infty . \tag{1}$$

☐ Received November 2003, revised March/June 2004, in final form August 2004.
☐ Ayman Baklizi and Mohammed AL-Haj Ebrahem are Assistant Professors in the Department of Statistics at Yarmouk University, Irbid, Jordan; email: baklizi1@hotmail.com and m_hassanb@hotmail.com.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 17-29, Feb. 2005

Improved Simultaneous Estimation of Small Domain Sizes through Model Assistance

Arijit Chaudhuri Mausumi Bose J. K. Ghosh
Indian Statistical Institute

ABSTRACT We consider the problem of simultaneous estimation of totals of several variables in a survey population. Using a stratified two-stage survey sampling scheme with varying probability sampling at both stages, we examine the efficacies of alternative versions of generalized regression (greg) estimators. These estimators are model-assisted but design-based and hence robust. They are contrasted against empirical Bayes estimators which are model-dependent. An illustration with live data motivates our recommendation for synthetic greg estimators based on a regression model through the origin, with a common slope for the first stage units. This recommendation may be adopted in comparable situations which are plentiful in real life survey practices.

Keywords Domain; Economic census; Empirical Bayes; Generalized regression; Simulation; Synthetic estimation; Two-stage stratified sampling; Varying probabilities.

1. Introduction

We consider a survey situation where the objective is to simultaneously estimate the totals of several variables in a population on the basis of a single sample. This problem is motivated by a survey problem in India where the totals of earners from 10 different vocations in district are to be separately and simultaneously estimated from a single sample drawn from the district. We study the problem in a general framework and propose a technique with some alternative estimators. Finally, we apply this approach to live Indian data in order to compare the efficacies of these alternative estimators.

Given the population, we initially choose an efficient sampling design to draw the sample. Let the population be such that, stratified two-stage sampling with unequal selection

☐ Received December 2003, revised March/September 2004, in final form October 2004.
☐ Authors of this paper are all affiliated to Indian Statistical Institute, Kolkata 700108, India; email address of Arijit Chaudhuri and Mausumi Bose: achau@isical.ac.in and mausumi@isical.ac.in.
Research of Arijit Chaudhuri is partially supported by CSIR grant no. 21(0539)/02 EMR-II.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 31-41, Feb. 2005

Unequally Spaced Variance-Change Model

Tze-San Lee
Western Illinois University

Nanjiang Hou
Northwestern University

ABSTRACT Little was done in studying the issue of changes in variance for irregularly spaced time-related data. Here is presented an unequally spaced variance-change model which can be used to detect a change in the variance of the irregularly spaced environmental exposure data. Unobservable errors contaminated in the data are modeled by the Ornstein-Uhlenbeck process. By employing the likelihood ratio test, a test statistics is proposed to locate a change-point in time at which the variance of the underlying observations has been changed. The workers' exposure data collected from an asbestos-removal team is used to illustrate how to use the variance-change model to find the change-point at which the variance of the workers' asbestos exposure data had changed.

Keywords Asbestos exposure data; Brownian motion; Change-point; Likelihood ratio test; Ornstein-Uhlenbeck process.

1. Introduction

The detection of possible changes in the variance of time-related series has many real-world applications. For example, in the area of statistical quality control, control of variability in the output of industrial mass production has become the main criterion in the experimental design of the production system since G. Taguchi promotes the idea of off-line quality control (Kackar [11]). In the area of finance, Fielitz [7] showed that among the financial data being examined the variance in the rate of return has changed at some time point in more than 85% of them. Clark [5] used a subordinated stochastic model to show that the variance itself of the price change in a speculative market is a random variable. Hsu [8] used the weekly

☐ Received November 2003, revised March/June 2004, in final form July 2004.
☐ Tze-San Lee is Mathematical Statistician, Biometry Activity, CDC/NCEH/EHHE, Mail Stop E70, Atlanta, GA 30333, USA, email: tjl3@cdc.gov. Dr. Lee was a Professor in the Department of Mathematics at Western Illinois University, Macomb, IL 61455, USA. Nanjiang Hou is Statistical Analyst/Programmer II, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA, email: n-hou@northwestern.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 43-58, Feb. 2005

Skew Weibull Distributions on the Real Line II: Estimation and Applications

Višnja Jurić
University of California

Tomasz J. Kozubowski
University of Nevada

ABSTRACT We discuss maximum likelihood estimation connected with skew Weibull distributions introduced by Jurić and Kozubowski in [8], and present an application from finance.

Keywords Asymmetric Laplace law; Double exponential distribution; Double Weibull distribution; Foreign currency exchange rate; Maximum likelihood estimation; Skew normal distribution.

1. Introduction

Using the approach of Fernandez and Steel [5] and the fact that a symmetric Weibull variable can be represented as a power of a classical Laplace variable, Jurić and Kozubowski [8] defined two classes of asymmetric Weibull distributions and studied their basic properties. Introducing skewness into the symmetric Weibull distribution (see Balakrishnan and Kocherlakota [2]) via two inverse scale factors (one on the positive, and one on the negative half-axis) leads to asymmetric double Weibull distributions of type I, denoted by $ADW_\alpha(\sigma,\kappa)$. They are given by the probability density function (p.d.f.)

$$g(x)=\frac{1}{\sigma^\alpha}\frac{\alpha\kappa}{1+\kappa^2}\left[(x^+\kappa)^{\alpha-1}\exp\{-(\frac{\kappa}{\sigma}x^+)^\alpha\}+(\frac{x^-}{\kappa})^{\alpha-1}\exp\{-(\frac{x^-}{\kappa\sigma})^\alpha\}\right],\ x\neq 0 \qquad (1)$$

where $\alpha>0,\sigma>0,\kappa>0$ and

☐ Received February 2004, revised April 2004, in final form May 2004.
☐ Višnja Jurić is Graduate Student in the Department of Statistics and Applied Probability at the University of California at Santa Barbara, Santa Barbara, CA 93106, USA. Tomasz J. Kozubowski is an Associate Professor in the Department of Mathematics and Statistics at the University of Nevada at Reno, Reno, NV 89557, USA; email: tkozubow @unr.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 59-96, Feb. 2005

On the Estimation of a Linear Time Trend Regression with a One-way Error Component Model in the Presence of Serially Correlated Errors: Part II

Chihwa Kao
Syracuse University

Jamie Emerson
Clarkson University

ABSTRACT This paper continues the theoretical investigation of Kao and Emerson [19]. We show that a naïve GLS (NGLS) could be less efficient than the FE or FD when there is no intercept in the model and the error term is $I(1)$. The NGLS is as efficient as the GLS both for $I(0)$ and $I(1)$ error terms when there is an intercept in the model. Finally, we present Monte Carlo results to evaluate the finite sample properties of the estimators.

Keywords Random effects; Fixed effects; Nonstationarity; Panel time series.

1. Introduction

This paper is the sequel to Kao and Emerson [19]. In part I, the limiting distributions are established for ordinary least squares (OLS), fixed effects (FE), first difference (FD), and generalized least squares (GLS) estimators in a linear time trend regression with a one-way error component model in the presence of serially correlated errors.

We show that when the error term is $I(0)$, the FE estimator is asymptotically equivalent to the GLS estimator and the OLS estimator is less efficient than the GLS estimator. However, when the error term is $I(1)$, the FD estimator is asymptotically equivalent to the GLS estimator. The FE estimator is not as efficient as either the FD estimator or the GLS estimator. However, when the error term is $I(1)$, the naïve GLS (NGLS) estimator could be less efficient than the FE or FD when there is no intercept in the model. The NGLS estimator is as efficient as the GLS estimator both for $I(0)$ and $I(1)$ error terms when there is an

□ Received October 2003, revised April 2004, in final form December 2004.
□ Chihwa Kao is a Professor of Economics in the Center for Policy Research at Syracuse University, Syracuse, NY 13244-1020, USA; email: cdkao@maxwell.syr.edu. Jamie Emerson is an Assistant Professor in the School of Business at Clarkson University, Potsdam, NY 13676, USA; email: jemerson@clarkson.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 97-104, Feb. 2005

A Class of Extended Generalized Negative Binomial Distributions

Anwar Hassan
University of Kashmir

ABSTRACT A class of extended generalized negative binomial distributions (EGNBD) which contains GNBD as its subclass has been introduced. Several properties of EGNBD are discussed and recurrence relations for factorial moments are obtained.

Keywords Generalized negative binomial distributions; Generalized logarithmic series distributions.

1. Introduction

Jain and Consul [12] defined generalized negative binomial distribution (GNBD) as

$$\Pr[X = x] = \frac{m}{m+\beta x}\binom{m+\beta x}{x}\alpha^x(1-\alpha)^{m+\beta x-x}$$

$$= 0, \text{ elsewhere,} \qquad (1.1)$$

where $x = 0, 1, 2, \cdots$, $0 < \alpha < 1$, $m > 0$, and $1 \le \beta \le 1/\alpha$.

It was subsequently obtained by Consul and Shenton [6] as a particular family of Lagrangian probability distributions. The probability model (1.1) reduces to the binomial distribution when $\beta = 0$ and to negative binomial when $\beta = 1$, and m is an integer. Consul and Gupta [5] modified the parametric space of the distribution and they also studied some of its interesting properties. Jain and Consul [12] obtained the first four non-central moments by using a recurrence relation. Shoukri [14] obtained a recurrence relation among the central moments. The GNBD model (1.1) has many important applications in queuing theory and branching processes. Famoye and Consul [8] considered a stochastic urn model for the GNBD

☐ Received April 2004, revised October 2004, in final form December 2004.
☐ Anwar Hassan is a Reader in Statistics at the University of Kashmir, Srinagar, India; email: anwar_husan@yahoo.com.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 105-120, Feb. 2005

A Study of Nonparametric Models with the *p*-th Autoregressive Time Series Errors

Tsair-chuan Lin
National Taipei University

ABSTRACT The estimation of the autoregressive coefficients is studied under non-parametric and semiparametric models with autoregressive time series errors. We estimate the relationship via regression surface function nonparametrically and then use the estimated residuals to estimate the second-order characteristics of the unknown noise process. It is shown that under some assumptions these estimators are asymptotically equivalent to the estimators based on the autoregressive error process. For real data analysis, we find that the stock data has nonlinear characteristic and the nonparametric prediction model performs better than the classical multivariate time series model.

Keywords Autoregressive model; Time series model; Nonlinear and nonstationary model; Nonparametric regression.

1. Introduction

Many time series models in practice are best considered as a function of its past values or further considered as a function of its past values plus a function of another time series values. The classical transfer model surpasses the univariate time series, and is designed to exploit the relationship between models when one serves as a leading indicator for the other. Among the world's stock markets, the Dow Jones Index acts as a key role and hence its change always becomes a deep impact to the others. In this study, we are concerned with a nonparametric and nonlinear setup that is capable of accommodating mean nonstationarity and nonlinearity simultaneously.

To ease understanding, we introduce some basic setups. In univariate time series setup, let $\Phi = (\phi_1, \phi_2, \cdots, \phi_p)'$ be an unknown vector such that the polynomial $\phi(z) = 1 + \phi_1 z + \cdots + \phi_p z^p$ has no root inside the closed unit circle. Suppose that $\{Z_t\}$ is a zero-mean,

☐ Received February 2004, revised April and July 2004, in final form Jan 2005.
☐ Tsair-chuan Lin is an Associate Professor in the Department of Statistics at National Taipei University, San Shia, Taipei County 237, Taiwan, ROC; email: tsair@mail.ntpu.edu.tw.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 121-133, Feb. 2005

Interval Estimation for the Difference between Exponential Guarantee Time Parameters

Weiwen Miao
Macalester College

Paul C. Chiou
Lamar University

ABSTRACT This paper studies the interval estimation for the difference between location parameters of exponential distributions. Because the two scale parameters are nuisance parameters in the construction of a confidence interval, the interval depends on estimators of the two scale parameters. If the two scale parameters are identical, then a pooled estimator is used in the construction of the interval for the difference. We consider three confidence intervals, namely, the never-pool, always-pool, and sometimes-pool confidence intervals. The sometimes-pool confidence interval depends on the significance level of a preliminary test (pre-test). Different values of significance level result in different confidence intervals. In this article, the AIC (Akaike's Information Criterion [1]) significance level is used for the pre-test.

Keywords Akaike's Information Criterion; Exponential location parameter; Hazard function; Pooled estimator; Pre-test; Type II censored sample.

1. Introduction

The two-parameter exponential distribution plays an important role in the field of life testing and reliability theory since it is the only continuous distribution with a constant hazard function (failure rate). The reciprocal of the scale parameter is the failure rate. The location (threshold) parameter can translate the distribution along the time axis, so it is also known as the minimum life or guarantee time parameter. The guarantee time parameter can be used to model warranty periods for some products.

Epstein and Sobel [8] obtained the minimum variance unbiased estimator (MVUE) for its scale parameter and location parameter, respectively. The shrinkage estimators for the scale

☐ Received February 2004, revised July/September 2004, in final form November 2004.
☐ Weiwen Miao is an Assistant Professor in the Department of Mathematics and Computer Science at Macalester College, Saint Paul, MN 55105, USA; email: miao@macalester.edu. Paul C. Chiou is a Professor in the Department of Mathematics at Lamar University, Beaumont, TX 77710, USA; email: chiou@math.lamar.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 135-138, Feb. 2005

An Alternative Proof for a Continuity Property of Positive Definite Matrices

André I. Khuri
University of Florida

ABSTRACT This note concerns an article by Hui-Kuang Hsieh, which appeared in the August 2004 issue of this journal. The note provides an alternative method for proving the main proposition in that article without relying on a property of continuous functions used by the author. The proposed method is also easy to use from the computational point of view. In addition, an error is pointed out in the statement of Hsieh's proposition.

Keywords Hsieh's proposition; Quadratic form; Spectral decomposition theorem.

1. Hsieh's Proposition

Hsieh [3] stated and proved the following proposition:

Proposition Let $\mathbf{\Sigma}$ be a $p \times p$ positive definite and symmetric matrix, and $\mathbf{A}$ be any $p \times p$ matrix. Then there exists a positive number t_0 such that $\mathbf{M}(t) = \mathbf{\Sigma} - t\mathbf{A}$ is positive definite for all t with $|t| < t_0$.

In proving this proposition, the author invoked a certain result stated in Rao ([5], page 36) giving the necessary and sufficient condition for $\mathbf{M}(t)$ to be positive definite. This result, however, requires that $\mathbf{M}(t)$ be symmetric (see also Harville [2], Theorem 14.9.5, page 247, and Graybill [1], Theorem 12.2.2, page 396). Consequently, $\mathbf{A}$ must also be symmetric. The proposition, however, states that $\mathbf{A}$ is "any $p \times p$ matrix". Hence, Hsieh's proof is not

☐ Received November 2004, revised December 2004, in final form January 2005.
☐ André I. Khuri is a Professor in the Department of Statistics at the University of Florida, Gainesville, FL 32611-8545, USA; email: ufakhuri@stat.ufl.edu. Professor Khuri is Fellow of the American Statistical Association and an elected member of the International Statistical Institute. He served as an associate editor for the journal *Technometrics* from 1983 to 1992. According to a 1990 survey by the National Science Foundation, Professor Khuri was ranked 27th among the top 100 most prolific statistics journal authors in the world.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 139-144, Feb. 2005

On the Confusion about the (n – 1)-Divisor of the Standard Deviation

Mohammad Fraiwan Al-Saleh
Yarmouk University

ABSTRACT The divisor $n-1$ in the defining formula of the sample standard deviation (S) seems to be slightly odd, and puzzles many students in all levels. There are some attempts to justify the choice of this divisor, but none of which is completely satisfactory. In this paper, we review and discuss different arguments that are found in many introductory textbooks to justify the use of $n-1$ rather than the more natural n divisor. A strategy to eliminate such confusion is suggested.

Keywords Standard deviation; Variance; Unbiased estimator; Degrees of freedom; Infinite 0/0.

1. Introduction

Measures of central tendency such as the mean, median or mode provide an incomplete description of a data set. A sufficient description of a data set should also include measures of variation or spread of the data points. These measures tell us how close are the values clustered around their mean. The standard deviation (S) is the most important and widely used measure of variability of a data set. Roughly speaking, S measures the variation by determining how far on the average are the data points from their average value. Thus, it is essentially defined to be the average distance of the values from their mean. Mathematically, if $X_1, X_2, \ldots, X_n$ is a random sample, then the variance is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

and standard deviation is the positive square root of S^2.

☐ Received May 2004, revised July/September 2004, in final form October 2004.
☐ Mohammad Fraiwan Al-Saleh is a Professor in the Department of Statistics at Yarmouk University, Irbid, Jordan; email: m-saleh@yu.edu.jo.

 ISSN 1726-3328

Journal of Probability and Statistical Science
3(1), 145-159, Feb. 2005

Robust Annual Forecasting of Inpatient Hospital Bed-Stays Using Short Daily Time Series

Alexander Basilevsky — *The University of Winnipeg*
Scott Nowicki, J. Trumble Waddell — *Winnipeg Regional Health Authority*

ABSTRACT Decisions by health authorities for hospital infrastructure require annual forecasts of occupancy rates for general planning purposes and capital investment. The Winnipeg Regional Health Authority (WRHA) has collected a time series of daily bed-stays for the six Winnipeg hospitals over a period of 5 years from 1995-2000, which provides the inpatient count analyzed in this paper. The series exhibits several outlier time periods due to major flooding of the Red River as well as hospital labor disputes, but no systematic movement is observed except for a mild downward trend. The outliers together with the shortness of the series present a difficulty for modelling purposes, which must precede any attempt at forecasting. It is found that a preliminary discrete Karhunen-Loeve analysis of the time series, in the form of a Principal Components decomposition of a lagged correlation matrix, provides a good aid to model specification and hence 1-year ahead forecasts. The Principal Components Analysis reveals the presence of fairly regular quarterly movements hidden in the daily series but periodic monthly or weekly variation is absent (Figure 1). Secondly, we find that due to the presence of outliers in the series the Minimum Absolute Deviations (MAD) regression model provides better 1-year ahead forecasts than do four other forecasting procedures.

Keywords Robust forecasting; Time series; MAD; OLS; ARIMA; Holt-Winters; Exponential smoothing models; Health care planning; Hospital bed-stays.

☐ Received June 2004, revised September 2004, in final form November 2004.

☐ Alexander Basilevsky is a Professor in the Department of Mathematics and Statistics at the University of Winnipeg, Winnipeg, Manitoba, Canada R3B 2E9; email: math.stats@uwinnipeg.ca. Professor Basilevsky was the former president-at-large of the Statistical Association of Manitoba. He is the author of the two books: *Statistical Factor Analysis and Related Methods* and *Applied Matrix Algebra in Statistical Sciences***.** Scott Nowicki is a Senior Statistician and J. Trumble Waddell is the Regional Director of the Population Health and Health Systems Analysis Unit at the Winnipeg Regional Health Authority, Winnipeg, Manitoba, Canada R3C 4Y1; emails: snowicki@wrha.mb.ca and JWaddell@wrha.mb.ca.

 ISSN 1726-3328

JPSS

Journal of Probability and Statistical Science

Volume 3 Number 1 February 2005

Table of Contents

Theory and Methods

Teaching and Applications