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 4 Number 1

February 2006

JPSS Editorial Board

JPSS Editorial Board

(continued)

JPSS

Journal of Probability and Statistical Science

Volume 4 Number 1 February 2006

Table of Contents

Journal of Probability and Statistical Science
4(1), 1-20, Feb. 2006

Location-Scale Models for Multilevel Ordinal Data: Between- and Within-Subjects Variance Modeling

Donald Hedeker Michael Berbaum Robin Mermelstein
University of Illinois at Chicago

ABSTRACT Mixed-effects logistic regression models are described for analysis of two-level ordinal outcomes, where observations are observed clustered within subjects. Random effects are included in the model to account for the correlation of the clustered observations. This correlation can be the same for all subjects or allowed to vary by groups of subjects. Additionally, whereas the usual logistic model assumes that the covariate effects are the same across the cumulative logits, *i.e.*, proportional odds assumption, we describe two extensions to relax this assumption. The first permits separate covariate effects to be estimated for each of the $C - 1$ cumulative logits, where C = number of ordered categories. The second extension instead allows covariates to influence the scale of the ordinal response, in addition to their usual influence on the location. This latter extension can be more parsimonious and can be used to partition the degree of within- and between-subjects variance. An analysis is presented of a dataset from an adolescent smoking study, highlighting and comparing these extensions of the proportional odds mixed model.

Keywords Categorical data; Multilevel data; Proportional odds assumption; Logistic regression; Clustering; Repeated observations; Complex variance; Mixed-effects models.

1. Introduction

The ordinal logistic regression model, described as the proportional odds model by McCullagh [36], is a popular model for analyzing ordinal outcomes. For multilevel data, where observations are nested within clusters (*e.g.*, classes, schools, clinics) or are repeatedly

☐ Received July 2005, revised December 2005, in final form January 2006.
☐ Donald Hedeker and Michael Berbaum are affiliated with the Division of Epidemiology and Biostatistics, and the Institute for Health Research and Policy, in the School of Public Health; Robin Mermelstein is affiliated with the Institute for Health Research and Policy, and the Department of Psychology; University of Illinois at Chicago, Chicago, IL 60612, USA; website: www.uic.edu/~hedeker/ml.html.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 21-30, Feb. 2006

Empirical Likelihood Estimation from Incomplete Data

Nancy L. Glenn
University of South Carolina

ABSTRACT The expectation-maximization (EM) algorithm is a general iterative algorithm that employs parametric procedures for maximum likelihood estimation for missing data analysis. Since the EM algorithm uses parametric procedures, strong assumptions are made about the underlying distribution of the data. In this paper we alter the EM algorithm by using a nonparametric procedure, empirical likelihood (EL), for determining the maximum likelihood estimator and missing values. The EL algorithm has an advantage since it makes few assumptions about the underlying distribution of the data. Using simulated data, a fiddler crab salinity data set secured from a salinity gradient in the North Inlet Estuary of South Carolina (Brodie *et al*. [1]), and the well-studied seed data found in Snedecor [10], we show that the EL algorithm provides convergence rates and estimation values that are comparable to those of EM algorithm. We investigate the EL algorithm's robustness to initial estimates of starting values. We assume data are missing at random.

Keywords Empirical likelihood; Expectation-maximization algorithm; Missing data.

1. Introduction

The expectation-maximization (EM) algorithm, a commonly used missing data technique, was introduced by Hartley [3]. Missing data analysis covers a variety of problems that are often seen in practical applications (Little and Rubin [4]). We consider a nonparametric likelihood technique for the analysis of data with missing values, and assume that data are missing at random. A discussion of likelihood based missing data methods can be found in Little and Yau [5], Yau and Little [12], and Peng *et al*. [9]. These methods typically employ the

☐ Received December 2004, revised May 2005, in final form August 2005.
☐ Nancy L. Glenn is an Assistant Professor in the Department of Statistics at the University of South Carolina, Columbia, South Carolina 29208, USA; email: nglenn@stat.sc.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 31-39, Feb. 2006

A General Functional Relation between a Random Variable and Its Length Biased Counterpart

Tomasz J. Kozubowski
University of Nevada at Reno

ABSTRACT A pair of general functional relations between a positive random variable X with density f and its length biased counterpart with density $g(x) = xf(x)/E[X]$ is derived. These relations unify previously considered cases, including the inverse Gaussian and the lognormal distributions studied by Sansgiry and Akman [10]. A generalization to weighted distributions with densities of the form $g(x) = w(x)f(x)/E[w(X)]$ is obtained and illustrated on an exponential tilt of X (the Esscher transform of X), where $w(x) = e^{\theta x}$ with $E[e^{\theta X}] < \infty$. The functional relations are shown to be related to the probability integral transformation.

Keywords Actuarial science; Esscher transform; Exponential tilt; Inverse Gaussian distribution; Lognormal distribution; Probability integral transformation; Reciprocal property; Size-biased distribution; Weighted distribution.

1. Introduction

Weighted distribution appeared in the statistical literatures following the fundamental work of Rao [8] as natural models for various sampling situations where the probability of recording a given realization of a stochastic experiment is proportional to some weight function. Following Patil *et al*. [6], suppose X is a nonnegative random variable with p.d.f. $f(x)$. Suppose further that an observation x of X is recorded with probability proportional to a nonnegative weight function $w(x)$ so that $P(recording \mid X = x) = w(x)$. Then the recorded x is an observation from a random variable Y with the p.d.f.

$$f_w(y) = \frac{w(y)f(y)}{\int w(y)f(y)dy}. \tag{1}$$

☐ Received March 2005, revised July 2005, in final form August 2005.
☐ 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
4(1), 41-49, Feb. 2006

Asymptotic Bounds for Coverage Probabilities for a Class of Confidence Intervals for Ratio of Means in a Bivariate Normal Distribution

Jan Hannig
Colorado State University

ABSTRACT Finding confidence intervals for ratios, particularly when the numerator and denominator estimates are distributed as Gaussian random variables, is a problem that has attracted the attention of numerous researchers. Such problems arise frequently in the field of metrology since many quantities of interest are calculated by first measuring more basic quantities and then using the ratio of the basic quantities. In this paper we develop a simple asymptotic formula that allows us to estimate the true coverage of this confidence interval. In particular, the formula allows one to compute a value fo k which will result in an associated confidence coefficient very nearly equal to the desired value.

Keywords Confidence intervals; Ratios; Asymptotic evaluations; Gaussian random variables.

1. Introduction

Calculating confidence intervals for ratios, particularly when the numerator and denominator estimates are distributed as Gaussian random variables, is a problem that has attracted the attention of numerous researchers (e.g., Fieller [3], Buonaccorsi and Gatsonis [2]). Such problems arise frequently in the field of metrology since many quantities of interest are calculated by first measuring more basic quantities and then using the ratio of the basic quantities. For instance, Ohm's law tells us that $I = V/R$ where I is current, V is voltage and R is resistance. Thus, current in a conductor may be "measured" by first measuring voltage

---

☐ Received December 2004, revised September 2005, in final form October 2005.
☐ Jan Hannig is an Assistant Professor in the Department of Statistics at Colorado State University, Fort Collins, CO 80523-1877, USA; email: hannig@stat.colostate.edu. This research is supported in part by the National Science Foundation under grant DMS-0504737.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 51-64, Feb. 2006

Properties of Generalized Log-Logistic Families of Lifetime Distributions

James U. Gleaton
University of North Florida

James D. Lynch
University of South Carolina

ABSTRACT It is established that the generalized log-logistic transformation of lifetime distributions is an equivalence relation, partitioning the family of distributions, with any member of an equivalence class being considered the class generator. It is shown that for all distributions in a class, the log-odds rate has the same functional form, whereas the hazard rate classification may differ within the same class. Conditions for the existence of moments and moment generating functions of transformed distributions are discussed. It is shown that the Kullback-Leibler information relating two members of a class is a strictly increasing function of the transformation parameter ratio alone.

Keywords Generalized log-logistic transformation, Equivalence relation, Log-odds rate, Kullback-Leibler information, Maximum entropy principle.

1. Introduction

A previous paper [2] examined the strength distribution of an inhomogeneous bundle of brittle elastic fibers under equal load sharing when the bundle is subjected to a tensile load. In this system, "inhomogeneous" means that the fibers, although all having the same Young's modulus of elasticity, differ in cross-sectional area. "Equal load sharing" means that when a fiber breaks, its share of the load is divided equally among the remaining surviving fibers. The strain ε induced in a fiber subjected to a tensile load is the proportional increase in length of the fiber. It is assumed that Hooke's law holds, i.e., the induced strain is proportional to the applied load, up to the breaking strain of the fiber.

☐ Received June 2005, revised November 2005, in final form December 2005.
☐ James U. Gleaton is a Visiting Assistant Professor in the Department of Mathematics and Statistics at the University of North Florida, Jacksonville, FL 32224, USA; email: jgleaton@unf.edu. James D. Lynch is the Director of the Center for Reliability and Quality Sciences at the University of South Carolina, Columbia, SC 29208, USA; email: lynch@stat.sc.edu. The work of the second author was supported in part by NSF Grants DMS 9877107 and NSF DMS 0243594.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 65-72, Feb. 2006

Characterizations of General Classes of Doubly Truncated Distributions Based on Record Values

Khalaf S. Sultan
King Saud University

G. A. Abd El-Mougod
South Valley University

ABSTRACT In this paper, we characterize general classes of doubly truncated absolutely continuous distributions by considering the conditional expectation of functions of record values. The specific distributions considered as particular case of the general cases are the doubly truncated distribution of each of the Weibull, inverse Weibull, Pareto, power function, beta of the first kind, Rayleigh and Burr distributions.

Keywords Record values; Conditional expectation; Truncated distributions and characterizations.

1. Introduction

Record values have been extensively studied in literature. For excellent reviews, we refer to Ahsanullah ([3], [4], [7]), Arnold *et al*. ([12], [13]), Balakrishnan [14], Franco and Ruiz ([19], [20]), Galambos [21] and Nagaraja [28]. Some specific distributions have been characterized by using the relationships of the expected values of record values, see Ahsanullah ([3], [5], [6]), Balakrishnan and Ahsanullah [15], Gupta [22], Kamps ([23], [24]), Lin ([25], [26]), Nagaraja [27] and AL-Hussaini and Ahmad [10]. Some recurrence relations for single and product moments of record values for specific continuous distributions have been established by Balakrishnan and Ahsanullah ([15], [16]), Balakrishnan *et al*. ([17], [18]) and Sultan and Balakrishnan [29]. Ahmad and Fawzy [2] characterized some doubly truncated distributions based on the concept of generalized order statistics. AL-Hussaini and Ahmad [9] have constructed Bayesian interval prediction of generalized order statistics and

☐ Received December 2004, revised September/November 2005, in final form December 2005.
☐ Khalaf S. Sultan (corresponding author) is affiliated to the Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; email: ksultan@ksu.edu.sa. G. A. Abd El-Mougod is affiliated to the Department of Mathematics, Faculty of Science, South Valley University, Sohag, Egypt.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 73-84, Feb. 2006

Inferences on Exponentiated Power Law Process

K. Muralidharan
M. S. University of Baroda

Rupal Shah
Sardar Patel University

ABSTRACT Non-homogeneous Poisson Processes (NHPP) provide models for a variety of physical phenomena. For an NHPP, the probability of a failure in a small time interval depends only on the time since the initial startup of the system and not on the previous pattern of failure. We propose exponentiated power law process (EPLP) as a particular case of NHPP, where system improvement is affected by failure and repair. Other than studying the inferences of the model parameters we also provide the current reliability estimation and future reliability of the system based on predictive likelihood. The theoretical findings are supplemented by numerical computations.

Keywords Current intensity**;** Intensity function; Repairable systems; Future reliability; Predictive likelihood; Power law process; Renewal process.

1. Introduction

A power law process (PLP) is a special case of non-homogeneous Poisson process (NHPP) and is considered as a model for reliability growth. A PLP is characterized by the rate of occurrence or intensity function

$$\lambda(t) = (\beta/\theta)(t/\theta)^{\beta-1}, \quad \theta > 0, \beta > 0, t > 0 \tag{1.1}$$

and the mean value function or cumulative intensity function, $\lambda(t) = (t/\theta)^{\beta}$, linear in time *t*. Here θ and β are respectively called the scale and shape parameter of the process. For $\beta = 1$ the process reduces to a homogeneous Poisson Process (HPP). Otherwise, a PLP provides a model for a system whose reliability changes as it ages. If $\beta > 1$, it models a deteriorating system and when $\beta < 1$, it provides a model for reliability growth.

☐ Received January 2005, revised June 2005, in final form December 2005.
☐ K. Muralidharan is a Professor in the Department of Statistics at M. S. University of Baroda, Baroda, 390002, India; email: lmv_murali@yahoo.com. Rupal Shah is affiliated to the Department of Statistics, Sardar Patel University, V. V. Nagar, Anand-388 120, India.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 85-98, Feb. 2006

Using the Pareto Distribution in Queueing Modeling

John Shortle
Donald Gross
George Mason University

Martin Fischer
Denise Masi
Mitretek Systems

ABSTRACT The purpose of this paper is to show that using different forms of the Pareto distribution can result in drastically different congestion measures in a queueing system. We consider a 1-parameter form and a 2-parameter form of the Pareto. We show that matching the first two moments is not sufficient to guarantee similar performance measures. In fact, different forms of the Pareto can yield *qualitatively* different behavior in a queueing system, with respect to analogous Markovian models. We specifically consider P/M/1 and M/P/1 queueing systems.

Keywords Pareto distribution; Queueing modeling.

1. Introduction

The Pareto distribution was named after Italian economist Vilfredo Pareto (1848 – 1923). In*"Cours d'economie politique professe a l'universite de Lausanne"* (three volumes, 1896, 1897), Pareto argued that the distribution of income in a country can be described by the formula (called Pareto's Law)

$$\log(N) = \log(A) + m\log(x),$$

☐ Received February 2005, revised September 2005, in final form November 2005.

☐ John F. Shortle is Assistant Professor and Donald Gross is Research Professor, Department of Systems Engineering and Operations Research, George Mason University, Fairfax, Virginia 22030, USA; emails: jshortle@gmu.edu and dgross1@gmu.edu. Professor Shortle was a recipient of the INFORMS Daniel H. Wagner Prize for excellence in Operations Research Practice. Professor Gross is the co-author of the well-known book, *Fundamentals of Queueing Theory*, and is past president of INFORMS. He was Director, Operations Research and Production Systems at the National Science Foundation, 1988-1990; 1996. He has received the INFORMS Kimball Medal for Service to the Operations Research Profession. Martin J. Fischer is Senior Fellow and Denise M. B. Masi is Principal Engineer, Mitretek's Center for Telecommunications and Advanced Technology, 3150 Fairfax Park Drive South, Falls Church, Virginia 22042, USA; emails: mfischer @mitretek.org and dmasi@mitretek.org.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 99-105, Feb. 2006

Binomial Tails Domination for Random Graphs via Bell Polynomials

Andrzej Korzeniowski
University of Texas at Arlington

ABSTRACT We obtain an inequality between tails of binomial distributions by establishing a combinatorial identity for a sum involving binomial and multinomial coefficients associated with the Bell polynomials. As an application a connection to random graphs is presented.

Keywords Bell polynomials; Di Bruno formula; Random graphs.

1. Introduction

As remarked in [5] the task of reducing combinatorial sums to a single closed form expression is an art in itself. Indeed, despite various algorithms and techniques developed over the years, quite often one encounters a situation in which the known methodology does not apply and requires a novel approach. This paper is an example of such a case in which a combinatorial identity emerged as a by-product of studying polynomials associated with binomial random graphs. Namely, let $Q(x)$ be a binomial-type polynomial defined by

$$Q(x)=\sum_{i=0}^{m}\binom{n}{i}(x^k-1)^i, \quad 0\le m\le n, \quad k\ge 1 \tag{1}$$

and consider

$$Q^j(1)=\left.\frac{d^jQ(x)}{dx^j}\right|_{x=1}, \ 0\le j\le m. \tag{2}$$

Evaluating $Q^j(1)$ when $k=1$ is trivial however for $k\ge 2$ the matter is far less obvious.

☐ Received June 2005, revised October/November 2005, in final form December 2005.
☐ Andrzej Korzeniowski is an Associate Professor in the Department of Mathematics at the University of Texas at Arlington, Arlington, TX 76019, USA; email: korzeniowski@uta.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
4(1), 107-121, Feb. 2006

Bayesian and Random Effects Approach in Meta-Analysis of Relative Risk from Observational Studies

Yiqing Song
Harvard Medical School

Ka He Borko Jovanovic
Northwestern University

ABSTRACT Meta-analysis is applied in epidemiological research to summarize findings from a set of available studies, to maximize statistical power of testing and precision in estimation, typically that of an existing effect across studies. Heterogeneous nature of observational studies makes the homogeneity assumption inherent in the fixed-effect approach unrealistic. A random-effects model is more appropriate in such study settings because it can allow for both within and between-study heterogeneity. To further incorporate uncertainty introduced by residual variation from between-study heterogeneity, a fully Bayesian approach has been recommended. In this article we review the random effects, empirical Bayesian and two versions of fully Bayesian methods and then applied them to two real data sets. We find that all of the models yield comparable results in the absence of substantial between-study heterogeneity. However, when such heterogeneity is present, a fully Bayesian approach does seem to provide an advantage. An additional appeal of a Bayesian approach for synthesizing evidence from observational studies lies in the specification of prior distributions for unknown parameters in a hierarchical model to explicitly account for uncertainty from between-study variation as well as the effects of study-level covariates.

Keywords Meta-analysis; Heterogeneity; Uncertainty; DerSimonian and Laird method; Empirical Bayes method; Bayesian approach.

☐ Received May 2005, revised November 2005, in final form December 2005.

☐ Yiqing Song is a Research Fellow, Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA; email: ysong@hsph.harvard.edu. Ka He is an Assistant Professor and Borko Jovanovic is a Research Associate Professor, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, 60611, USA; emails: kahe@northwestern.edu and borko@northwestern.edu. The work of the first author was supported in part by the Carson's Family Fellowship from Harvard School of Public Health.

 ISSN 1726-3328