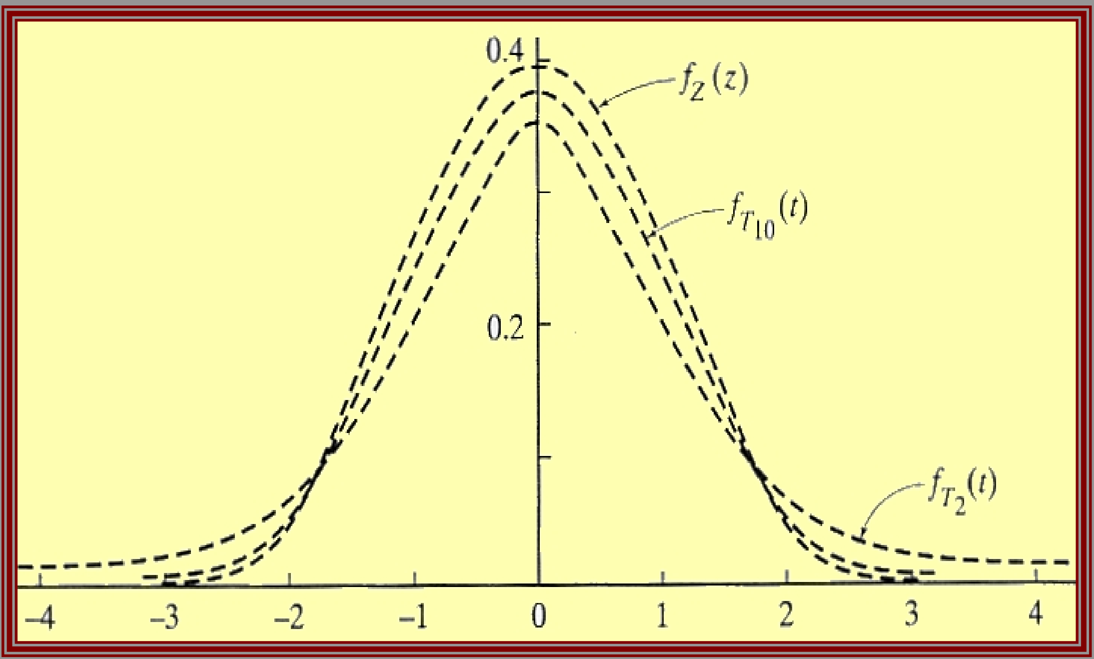
ISSN 1726-3328

JPSS

A comprehensive journal of probability and statistics
for theorists, methodologists, practitioners, teachers, and others



JOURNAL OF PROBABILITY AND STATISTICAL SCIENCE

Volume 7 Number 1 February 2009

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 7 Number 1

February 2009

JPSS Editorial Board

機統學誌 **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.

魏蘇珊文教事業機構發行

總公司: 中華民國臺灣新竹市建美路 2 巷 26 號。

JPSS

Journal of Probability and Statistical Science

Volume 7 Number 1 February 2009

Table of Contents

Journal of Probability and Statistical Science
7(1), 1-9, Feb. 2009

Transient Analysis of Multistage Degraded Systems with *L* Exponential Failure Modes and Partial Repairs of General Distribution Modeled by Hypoexponential Distribution

Magdi S. Moustafa
The American University in Cairo

ABSTRACT The paper presents a model of multistage degraded system subject to *L* random failure modes and partial repairs. A transient analysis is performed and transient probabilities are calculated to find the availability, the means of life time and operational life time. In the paper, constant state dependent transition rates for the degradation process as well as failure process are considered. On the other hand the partial repairs follow general distributions with square of coefficient of variation less than or equal to one and state dependent which can be modeled by hypoexponential or Erlang distributions. This paper extends previous systems that can be considered as particular cases of this one. Numerical examples are provided to illustrate applicability of the expressions that are obtained throughout the paper.

Keywords Degraded system; Exponential failures; Partial repairs of general distribution; Transient measurements.

1. Introduction

Failures can be classified into degraded and critical (see [1]). According to Rausand and Oien [2], critical or random failure which is sudden and ceases one or more fundamental functions of the system requires repair to return the system to a satisfactory condition. Degraded failure which is gradual, partial, or both may not cease the fundamental function and there can be multiple stages of degradation, and the system may fail after a certain number of stages. In each degradation stage, the system can fail critically (randomly), and a repair to return the system back to its degradation stage before the failure is considered as a minimal repair or partial repair.

□Received October 2007, revised December 2007 and April 2008, in final form July 2008.
□Magdi S. Moustafa is Professor and Chairman, Department of Mathematics & Actuarial Science, The American University in Cairo, Cairo, Egypt; email: mmostafa@aucegypt.edu.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 11-18, Feb. 2009

Estimation of $P(X \le Y)$ for the Uniform Distribution in the Presence of Outliers

Ulhas Jayram Dixit and K. D. Phal
University of Mumbai

ABSTRACT The maximum likelihood and the uniformly minimum variance unbiased estimator (UMVUE) of $P(X \le Y)$ are derived, where both X and Y have uniform distribution and outliers are generated from Generalized Uniform Distribution (GUD). It is shown that UMVUE is better than MLE when one parameter of GUD is known. When both parameters of the GUD are unknown, $P(X \le Y)$ is estimated by using mixture estimate. It is shown that estimator of $P(X \le Y)$ is consistent.

Keywords Outliers; GUD; UMVUE; MLE; MSE.

1. Introduction

The estimation of $R = P(X \le Y)$ plays an important role in reliability analysis. For example, when X is stress (or demand) and Y is strength (or supply), R is taken as the measure of performance of the system.

Ooms and Moore [2] had shown that as a plant develops into the reproductive phase of growth, a mat of smaller roots grows near the surface to a depth of (1/6)th of maximum depth achieved. Normally the mass of a root has a uniform distribution. Dixit *et al*. [1] assumes that a set of random variables $(X_1, X_2, \cdots, X_n)$ represents the masses of roots, where some of these roots have different masses. Therefore, those masses have different uniform distributions with unknown parameters. Hence, we can assume that the some (say k) different observations out of n random variables are present and that these k observations are distributed with p.d.f. $g(x, \theta, \alpha)$ where

---

□Received June 2007, revised October 2008, in final form November 2008.
□Authors of this paper are affiliated with the Department of Statistics at the University of Mumbai, Vidyanagari, Santacruz(E), Mumbai-400098, MS, India; email address of Ulhas Jayram Dixit: udixit@gmail.com.
□AMS Classification: 62F10.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 19-27, Feb. 2009

On the Reliability Function of a Coherent Structure of Components Sharing a Common Environment

G. Chaudhuri
Indian Institute of Management Kozhikode

ABSTRACT The reliability function of a coherent system of components sharing a common random environment is studied and compared with the reliability function obtained under the assumption of independence of component failures. It is shown that the assumption of independence underestimates the reliability of the system composed of IFRA (increasing failure rate average) Weibull components sharing a common random environment whereas in case of exponential components the latter crosses the former reliability function exactly once from below. Thus, for an initial period of operation, the assumption of independence overestimates the system reliability. This generalized the results of Currit and Singpurwalla [4] and Chaudhuri [2] to a more general coherent structure. That the reliability function of the system operating in a random environment does not belong to any of the aging classes of life distributions considered in the literature is also established. Illustrative examples are given.

Keywords System reliability; Random environment; Weibull distribution; Coherent system; Minimal path set; the CHA representation; HNBUE class of life distributions.

1. Introduction

In assessing the reliability of a system, the assumption of independence of component life-lengths is made in practice. But this is seldom realistic when the system is operating in a common random environment. The net effect of the random environment is to induce dependence among the life-lengths of the components. For instance, applied voltage is always a stress to components of a system receiving power from a power pack. When the components draw voltage from a power pack, we have a situation where the components share a common random environment, randomness being due to manufacturing variability. Thus, a harsh environment should cause all the components to fail earlier than otherwise.

□Received June 2007, revised October 2008, in final form November 2008.
□G. Chaudhuri is affiliated with the Indian Institute of Management Kozhikode, IIMK campus P. O., Calicut 673570, India; email: chaudhuri@iimk.ac.in.

ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 29-41, Feb. 2009

Significance Testing for Cronbach's Alpha

Mezbahue Rahman   Larry M. Pearson   Kumiko Suzuki
Minnesota State University

ABSTRACT Cronbach's Alpha is used in checking for reliability and/or consistency of a data set having multiple responses for each subject. Here, we explore the statistical inferences of the Cronbach Alpha coefficient. Several parametric exact and asymptotic testing methods are presented along with non-parametric bootstrap and permutation estimates of the standard errors. Comparison among the methods is done using simulation. Applications are shown for two different sets of real life data. Among the asymptotic tests, the one using logarithmic transformation performed better.

Keywords Maximum likelihood estimate; Monte Carlo simulation.

1. Introduction

Let us consider $X_1, X_2, \cdots, X_p$. To be p different measurements for an attribute on n different subjects. Cronbach [3] proposed a multiple correlation coefficient α in measuring the consistency among sets of measurements. Cronbach's Alpha is widely used in psychometric studies. Variations of Cronbach's Alpha and their explanations prior to Cronbach [3] can be found in Cronbach [2], Kuder and Richardson [12], and the references therein. Following Cronbach [3], different authors studied the properties of the coefficient, such as, Novic and Lewis [16], Feldt *et al*. [7], Miller [14], Bonett [1], Koning and Franses [11], Duhacheck *et al*. [5], and the references therein. Cronbach's Alpha is given by

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p} VAR(X_i)}{\sum_{i=1}^{p}\sum_{i'=1}^{p} COV(X_i, X_{i'})}\right) = \frac{p}{1-p}\left(1 - \frac{\mathrm{tr}\Phi}{j'\Phi j}\right) \tag{1}$$

where VAR stands for variance, COV stands for covariance, tr stands for trace of a matrix, Φ is the variance-covariance matrix, and j is a column vector of ones.

□Received May 2008, revised August 2008, in final form September 2008.
□Authors of this article are affiliated with the Department of Mathematics and Statistics, Minnesota State University, Mankato, MN 56002, USA; emails: mezbahur.rahman@mnsu.edu, larry.pearson@mnsu.edu, and suzukikumicom@hotmail.com.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 43-56, Feb. 2009

Estimation after Selection in the Modified Power Series Distributions under Stein's Loss Function

Riyadh R. Al-Mosawi
Thiqar University

Ashok Shanubhogue
Sardar Patel University

ABSTRACT Let $\Pi_1, \Pi_2, \cdots, \Pi_p$ be p ($p \geq 2$) independent left-truncated modified power series populations with functions of parameters $h(\theta_1), h(\theta_2), \cdots, h(\theta_p)$, respectively. For each i = 1, 2, …, p, assume Z_i denotes the sum of n independent observations from the population Π_i. For selecting the best population, the one associated with the largest parameter θ_M (smallest parameter θ_N), we consider the natural selection rule which selects the population Π_i if and only if $Z_i \geq Z_j$, $\forall j \neq i$. Since our problem is in the discrete case, there is a high probability of existing ties in the values of Z_i. So, we break these ties by ordering the tied Z_i from the smallest to the largest index. In this paper, we consider the problem of estimating the function $h(\theta_M)(h(\theta_N))$ of the selected population under Stein's loss function and construct two different classes of improved estimators dominating the natural estimator by solving certain difference in- equalities. In particular, improved estimators for the selected left-truncated generalized Poisson and left-truncated generalized negative binomial population are also presented.

Keywords Estimation after selection; Stein's loss function; Difference inequalities; Truncated modified power series distributions.

1. Introduction

Let $\Pi_1, \Pi_2, \cdots, \Pi_p$ be p ($p \geq 2$) independent left-truncated modified power series populations with densities (with respect to counting measure)

$$P(X = x) = P_X(\theta_i) = \frac{a(x)(h(\theta_i))^x}{f(\theta_i)},\ x = t,\ t+1, \cdots;\ t = 1,\ 2, \cdots$$

□ Received August 2008, revised December 2008, in final form January 2009.
□ Riyadh R. Al-Mosawi is affiliated with the Department of Mathematics at Thiqar University, Iraq; email: riyadhrm@gmail.com. Ashok Shanubhogue is affiliated with the Department of Statistics at Sardar Patel University, Gujarat, India; email: a_shanubhogue@yahoo.com.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 57-69, Feb. 2009

Conditional Cumulative Distribution Estimation and Its Applications

Ali Laksaci
Université de Sidi Bel Abbès

Fouzia Maref
Center Universitaire Moulay Taher

ABSTRACT In this paper we prove the uniform almost complete convergence (with rate) of kernel estimator of the conditional cumulative distribution function when the explanatory variable is curve. As an application, we use this result to derive same asymptotic propriety for the conditional quantiles estimated by kernel method.

Keywords Kernel estimation; Conditional quantiles; Conditional cumulative distribution; Functional random variables; Semi-metric space; Small balls probability.

1. Introduction

The estimation of the conditional distribution function is an important subject in statistics. Indeed, several prediction tools in nonparametric statistics, such as the conditional mode, the conditional median or the conditional quantiles, are based on the preliminary estimate of this functional parameter. The main goal of this paper is to study this non-parametric model when the explanatory variable is functional.

The literature on estimation of the nonparametric models related to the conditional distribution is quite important when the data are real (see for instances, Roussas [15] and Stute [16] for previous results and Gannoun *et al*. [12], Deheuvels and Mason [4] for recent advances and references). In the functional case, the first results were obtained by Ferraty *et al*. [7]. They established the almost complete consistency in the i.i.d. case of the kernel estimators of the conditional distribution function and of the conditional density. An application to the conditional mode and conditional quantiles is also presented in their work. The

□Received May 2008, revised June 2008, in final form July 2008.
□Ali Laksaci is affiliated with Laboratoire de Mathématiques, Université de Sidi Bel Abbès, BP 89 Sidi Bel Abbès 2200, Algeria; email: alilak@yahoo.fr. Fouzia Maref is affiliated with Départment de Mathématiques, Center Universitaire Moulay Taher; email: fouziamaref@yahoo.fr.
□Mathematics subject classification: 62G05, 62G99, 62G20.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 71-80, Feb. 2009

Bayesian Estimation in Kim and Warde's (2005) Mixed Randomized Response Model Using Mixed Prior Distribution

Zawar Hussain and Javid Shabbir
Quaid-i-Azam *University*

ABSTRACT In this study, we have considered the Bayesian estimation of the population proportion of a sensitive characteristic when the data are obtained through Kim and Warde [7] mixed randomized response model (RRM). To do Bayesian estimation, in particular, we assumed the availability of multiple prior informations and used a mixture of them as a prior distribution. For comparison purposes, the mean squared errors of the Bayes estimator as well as of the usual maximum likelihood estimator (MLE) are used. It has been observed than Bayes estimator outperforms the MLE over a wider range of population proportion and other design parameters of the Kim and Warde [7] RRM.

Keywords Bayesian estimation; Randomized response model; Mean squared error.

1. Introduction

Warner [19] introduced an ingenious method to decrease the biasness in the parameters and to increase the response rate. Warner's model consists of two complimentary questions A and A^c to be answered on probability basis. Thus far, many developments of Warner's Randomized Response model (RRM) have been proposed by many researchers. Greenberg *et al*. [4], Mangat [8], Singh and Horn [15], Christofides [2], Kim and Warde [6], and Kim and Warde [7] are some of the many to be cited. The interested readers may be referred to Chaudhuri and Mukerjee [1]. In some practical situations, prior information regarding parameter to be estimated is available and can be used along with the sample information for

□Received June 2008, revised November 2008, in final form December 2008.
□Authors of this paper are affiliated to the Department of Statistics at Quaid-i-Azam University 45320, Islamabad 44000, Pakistan; email of Zawar Hussain (corresponding author): zhlangah@yahoo.com.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 81-90, Feb. 2009

Choosing between Ratio and Product Estimators for the Same Data-Set

M. C. Agrawal
University of Delhi

A. B. Sthapit
Tribhuvan University

ABSTRACT We have shown in this paper that, in many practical situations, the ratio or the product estimator that exploits the auxiliary variable *x* could be rendered more efficiently if we make use of the available *x*-variable in the form of its reciprocal coupled with an interchange of estimators. We have used exact and approximate mean square errors of the customary ratio and product estimators and derived the relevant necessary and sufficient conditions for superior performance of one relative to the other using the same data-set. We have furnished illustrations to underscore the fact that considerable gains are attainable if we invert the available auxiliary information (which is positively correlated with the main variable) and employ the usual product estimator instead of the customary ratio estimator.

Keywords Ratio estimator; Product estimator; Reciprocal of the auxiliary information; Exact bias; Exact mean square error.

1. Introduction

Consider a finite population of size *N* arbitrarily labeled 1, 2, …, *N*. Let y_i and x_i (i = 1, 2, …, *N*) be the measurements on the i^{th} unit in respect of the main and the auxiliary variables *y* and *x* respectively. A sample *s* of size n is drawn from the population employing simple random sampling without replacement design. Let $\bar{y}_s$ and $\bar{x}_s$ be the sample means of *y* and *x* variables respectively and $\bar{Y}$ and $\bar{X}$ be the corresponding population means. Further let $\tilde{x}_s$ and $\tilde{X}$ be, respectively, the sample and the population harmonic means of *x*-variable defined by

$$\tilde{x}_s = \frac{n}{\sum_{i=1}^{n}(1/x_i)} \quad \text{and} \quad \tilde{X} = \frac{N}{\sum_{i=1}^{N}(1/x_i)}.$$

□Received November 2006, revised September 2008, in final form November 2008.
□M. C. Agrawal is affiliated to the Department of Statistics at University of Delhi, Delhi 110 007, India; email: mc_agrawal@yahoo.com. A. B. Sthapit is affiliated to the Department of Statistics at Tribhuvan University, Kathmandu, Nepal; email: azaya_sthapit@enet.com.np.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 91-100, Feb. 2009

Section of Teaching and Applications

The Moments of a Compound Poisson Process with Exponential or Centered Normal Jumps

Shai Covo
Bar Ilan University

ABSTRACT We present simple explicit formulas for the moments of a compound Poisson process in the cases of exponential and centered normal jump distributions. We also review the general case of, loosely speaking, arbitrary jump distribution. As a lemma, we obtain the ascending factorial moments of the Poisson distribution, a simple result which seems to be unrecognized in the literature.

Keywords Compound Poisson process; Moments; Cumulants; Bell polynomials; Generating function; Faà di Bruno's formula; Poisson distribution; Ascending factorial moments.

1. Introduction

Let $N=\{N_t,\ t\geq 0\}$ be a Poisson process with rate λ, and $X_1, X_2, \ldots$ a sequence of i.i.d. random variables with common distribution *F*, independent of *N*. Then the process $S=\{S_t,\ t\geq 0\}$ defined by $S_t=\sum_{i=1}^{N_t} X_i$ is a compound Poisson process with rate λ and jump distributuion *F*. In this paper, we present (see Propositions 1 and 2) explicit formulas for the moments of S_t in the cases $F = \text{Exponential}(\theta)$ (i.e., $F(\mathrm{d}x)=\theta^{-1}\mathrm{e}^{-x/\theta}\mathrm{d}x, x>0$) and $F=\mathrm{N}(0,\sigma^2)$. Somewhat surprisingly, considering their obvious importance (the exponential jump case is particularly popular), elegance, and simplicity, our results are apparently not directly available in the literature. Actually, both can be regarded as simple observations. In the exponential jump case, the proposition follows immediately from some *generating function* example, upon identification with the moment generating function (MGF) of S_t, as

□ Received July 2008, revised November 2008, in final form December 2008.
□ Shai Covo is affiliated with the Department of Mathematics at Bar Ilan University, 52900 Ramat-Gan, Israel; email: green355@netvision.net.il.
□ This paper is part of the author's Ph.D thesis, prepared at Bar Ilan University under the supervision of Professor E. Merzbach. Work supported by the Doctoral Fellowship of Excellence, Bar Ilan University.

 ISSN 1726-3328

Journal of Probability and Statistical Science
7(1), 101-103, Feb. 2009

The Problem of Extremely Large Sample Sizes in Hypothesis Testing

David L. Farnsworth
Rochester Institute of Technology

ABSTRACT The seldom-addressed problem of extremely large sample sizes causing inordinate numbers of rejections of hypotheses is discussed. A suggestion is made for a simple and realistic procedure to neutralize the effect of those sample sizes. The strategy involves widening the set of null values and taking a subsample.

Keywords Anticonservative test; Difference that makes a difference; Extremely large sample size; Subsample.

1. Introduction

Statisticians know that large sample sizes can lead to the rejection of almost any null hypothesis. This is an important problem because extremely large sample sizes are more commonplace in this day of automatic data collection with digital devices. Sets of hundreds of thousands and, even, millions of values are not unusual. For continuous random variables, no null hypothesis is exactly true. If the samples are large enough, those hypotheses will almost certainly be rejected. The phenomenon of large samples producing inappropriate rejections of null hypotheses may be both one of the most widely known problems and one of the most ignored in statistical practice. For instance, the *z*-score for the test for a population mean is

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}.$$

A sufficiently large sample size, *n*, can produce a very large *z*-score and, hence, rejection of the hypothesized mean μ_0 for any sample mean that is different from the hypothesized mean. For example, the *z*-score is 104.9 for a two-sided, α-level test of the hypothesized mean $\mu_0 = 654$ with $n = 275{,}000$, $\bar{x} = 651$, and $s = 15$.

□Received September 2008, revised December 2008, in final form January 2009.
□David L. Farnsworth is affiliated with the School of Mathematical Sciences at Rochester Institute of Technology, Rochester, New York 14623, USA; email: dlfsma@rit.edu.

 ISSN 1726-3328